Postmarket Sequential Database Surveillance of Medical Products

by

Judith C. Maro

S.M. Technology and Policy, Massachusetts Institute of Technology (2009)
B.S. Economics, The George Washington University (1999)
B.A. Political Science, The George Washington University (1999)

Submitted to the Engineering Systems Division
In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2013

Signature of Author: _____
Engineering Systems Division
November 6, 2012

Certified by: _____
Deborah J. Nightingale
Professor of the Practice of Aeronautics and Astronautics and Engineering Systems
Thesis Committee Co-Chair

Certified by: _____
Roy Welsch
Professor of Statistics and Management Science and Engineering Systems
Thesis Committee Co-Chair

Certified by: _____
Mort Webster
Assistant Professor of Engineering Systems
Thesis Committee Member

Certified by: _____
Jeffrey S. Brown
Assistant Professor of Population Medicine, Harvard Medical School
Thesis Committee Member

Accepted by: _____
Oliver L. DeWeck
Associate Professor of Aeronautics and Astronautics and Engineering Systems
Chair, Engineering Systems Division Education Committee

Postmarket Sequential Database Surveillance of Medical Products

by

Judith C. Maro

Submitted to the Engineering Systems Division
on November 6, 2012 in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Engineering Systems

ABSTRACT

This dissertation focuses on the capabilities of a novel public health data system – the Sentinel System - to supplement existing postmarket surveillance systems of the U.S. Food and Drug Administration (FDA). The Sentinel System is designed to identify and assess safety risks associated with drugs, therapeutic biologics, vaccines, and medical devices that emerge postlicensure. Per the initiating legislation, the FDA must complete *a priori* evaluations of the Sentinel System's technical capabilities to support regulatory decision-making.

This research develops qualitative and quantitative tools to aid the FDA in such evaluations, particularly with regard to the Sentinel System's novel sequential database surveillance capabilities. Sequential database surveillance is a "near real-time" sequential statistical method to evaluate pre-specified exposure-outcome pairs. A "signal" is detected when the data suggest an excess risk that is statistically significant. The qualitative tool – the Sentinel System Pre-Screening Checklist – is designed to determine whether the Sentinel System is well suited, *on its face*, to evaluate a pre-specified exposure-outcome pair. The quantitative tool - the Sequential Database Surveillance Simulator - allows the user to explore virtually whether sequential database surveillance of a particular exposure-outcome pair is likely to generate evidence to identify and assess safety risks in a timely manner to support regulatory decision-making. Particular attention is paid to accounting for uncertainties including medical product adoption and utilization, misclassification error, and the unknown true excess risk in the environment.

Using vaccine examples and the simulator to illustrate, this dissertation first demonstrates the tradeoffs associated with sample size calculations in sequential statistical analysis, particularly the tradeoff between statistical power and median sample size. Second, it demonstrates differences in performance between various surveillance configurations when using distributed database systems. Third, it demonstrates the effects of misclassification error on sequential database surveillance, and specifically how such errors may be accounted for in the design of surveillance. Fourth, it considers the complexities of modeling *new* medical product adoption, and specifically, the existence of a "dual market" phenomenon for these new medical products. This finding raises non-trivial generalizability concerns regarding evidence generated via sequential database surveillance when performed immediately post-licensure.

Thesis Co-Chairman: Deborah J. Nightingale
Title: Professor of the Practice of Aeronautics and Astronautics and Engineering Systems

Thesis Co-Chairman: Roy Welsch
Title: Professor of Statistics and Management Science and Engineering Systems

ACKNOWLEDGEMENTS

First, to my loving husband Vijay and my beautiful Clara, I could not have done this without your love and support. Thank you for always bringing a smile to my face. To the extended Maro and Vanguri families, and especially my Mom, thank you for lending a hand with extended stays that made research possible. Vijay and I are also very grateful to have relied on Ms. Bonnie Wright, who was able to care for Clara so consistently and with such love during this entire process.

To my MIT Colleagues, most notably my committee members – Professors Deborah Nightingale, Roy Welsch, and Mort Webster – thank you for your support, guidance, and reassurance. To Professors Joseph Sussman and Christopher Magee, a special thanks for their constant encouragement and wise counsel. Thanks also to Beth Milnes for always taking care of us, the students. I have been blessed to have always found a sounding board in my classmates, particularly Bryan Palmintier, Nidhi Santen, David Keith, and Jordan Peck among many others. I am also grateful to Dr. Gigi Hirsch for generously provided me working space when needed.

To my mentors at the Department of Population Medicine at Harvard Medical School - Professors Jeff Brown and Martin Kulldorff – thank you for your support and your generosity with both time and feedback. Thank you to my other colleagues in the department who kindly read and provided feedback on drafts of this document, especially Dr. Sharon Greene, Professor Meghan Baker, Professor Darren Toh, and Dr. Katherine Yih. To Dr. Tracy Lieu, Professor Grace Lee, Professor Lingling Li, and Dr. Alison Tse, thank you for inviting me to attend the vaccine safety meetings and, more generally, for inviting me into this scientific community.

I am indebted to Professor Richard Platt for always believing in me, providing research support in between fellowship opportunities, and tirelessly reading much of my work despite his incredibly demanding schedule.

I also want to thank Nicolas Beaulieu, Jim Marshall, Taliser Avery, and Lisa Trebino for their assistance with data retrieval, cleaning, and other important yet mundane research tasks. Thanks to Kim Lane, Beth Syat, and Jessica Sturtevant for being such great friends and confidantes. I'm very much looking forward to continuing my post-doctoral research in such an intellectually stimulating and genuinely warm environment.

Finally, this research would not have been possible without the mentorship, counsel, and fellowship support of Dr. Gerald Dal Pan of the U.S. Food and Drug Administration. It still amazes me that such an industrious public figure found so much time in his schedule for meetings, for reading the constant drafts of this dissertation, and for hosting my twice-yearly treks to the White Oak campus to meet with everyone at FDA. I look forward to our continued collaborations and friendship. Thanks also to Dr. Judy Staffa, Dr. David Graham, Dr. Solomon Ayasu, Dr. Marsha Reichman, Dr. Tarek Hammad, Dr. Mark Avigan, Dr. John Senior, Ms. Laura Governale and many others who provided me with valuable research insights.

This dissertation was supported by an appointment to the Research Participation Program at the Center for Drug Evaluation and Research administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

While I owe a debt to many who supported me in seeing this work through to fruition, any errors and omissions remain mine and mine alone.

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF TABLES

ABBREVIATIONS

AERS – Adverse Event Reporting System

BIC – Bayesian Information Criterion

CMaxSPRT – Conditional Maximized Sequential Probability Ratio Test

CPT – Current Procedural Terminology

CSSP – Conditional Sequential Sampling Procedure

EHR – Electronic Health Record

ESA – Exact Sequential Analysis

ETASU – Elements to Assure Safe Use

FDA – Food and Drug Administration

FDAAA – Food and Drug Administrations Amendments Act

GBS - Guillain-Barre Syndrome

GS LRT – Generalized Sequential Log-likelihood Ratio Test

HCPCS - Healthcare Common Procedure Coding System

HMORN – HMO Research Network

IOM – Institute of Medicine

IRD – Incidence Rate Difference

IRR - Incidence Rate Ratio

ITP - idiopathic thrombocytopenic purpura

MAD – Median Absolute Deviation from the Median

MaxSPRT – Maximized Sequential Probability Ratio Test

MF – Model Form

MP – Modular Program

MSE – Mean Squared Error

MSOC – Mini-Sentinel Operations Center

PDUFA – Prescription Drug User Fee Act

PMR – Postmarket Requirement

PPV- Positive Predictive Value

PRISM – Post-licensure Rapid Immunization Safety Monitoring

REMS – Risk Evaluation and Management Strategy

SDS – Sequential Database Surveillance

SPRT – Sequential Probability Ratio Test

VAERS – Vaccine Adverse Event Reporting System

VSD – Vaccine Safety Datalink

1 MOTIVATION AND OUTLINE OF THE DISSERTATION

In the mid 2000s, the U.S. Food and Drug Administration (FDA) came under heavy criticism after several high-profile regulatory failures to act on postmarket[1] drug safety risks in a timely manner that minimized public harm.[2] These failures prompted a landmark Institute of Medicine (IOM) study on the FDA's postmarket drug safety systems and authorities, including "the sum of all activities conducted by FDA and other stakeholders to monitor, evaluate, improve, and ensure drug safety."[3] This dissertation focuses on one of the IOM's recommendations that became a legislative mandate - the creation and use of a novel public health data system to supplement existing postmarket systems. Specifically, the IOM found the "[FDA's] ability to test drug safety hypotheses is limited," and consequently recommended:

> "that in order to facilitate the formulation and testing of drug safety hypotheses, [the FDA] (a) increase their intramural and extramural programs that access and study data from large automated healthcare databases and (b) include in these programs studies on drug utilization patterns and background incidence rates for adverse events of interest, and (c) develop and implement active surveillance of specific drugs and diseases as needed in a variety of settings."[4]

Congress responded by directing the FDA to establish a novel public health data system[5] to supplement existing systems to identify and assess safety risks[6] associated with drugs,

---

[1] The "postmarket" period is the period after licensure of a product by the U.S. Food and Drug Administration. Once licensed, the product is approved to be marketed to the general public, albeit perhaps with restrictions on access that will be discussed herein. Throughout this dissertation, the descriptors "postmarket," "postlicensure," and "postapproval" are used interchangeably to refer to this period, consistent with source material.

[2] See U.S. House Committee on Energy and Commerce and Subcommittee on Oversight and Investigations, *FDA's Role in Protecting the Public Health: Examining FDA's Review of Safety and Efficacy Concerns in Anti-depressant Use by Children* (U.S. G.P.O., 2005).; U.S. Senate Committee on Finance and U.S. Senate Committee on Finance, *FDA, Merck, and Vioxx: Putting Patient Safety First?* (U.S. G.P.O., 2005).

[3] Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public* (Washington, DC: National Academies Press, 2007), 2, http://www.nap.edu/catalog/11750.html.

[4] Ibid., 7.

[5] § 905 in *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, 2007, codified at 21 U.S.C. § 355(k)(3).

[6] § 901 in *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, 2007, codified at 21 U.S.C. § 355-1(b) defines the scope of safety risks of concern as follows: "The term 'serious risk' means a risk of a serious adverse drug experience" and "The term 'serious adverse drug experience' is an adverse drug experience that (A) results in—(i) death; (ii) an adverse drug experience that places the patient at immediate risk of death[...];(iii) inpatient hospitalization or prolongation of existing hospitalization; (iv) a persistent or significant incapacity or substantial disruption of the ability to conduct normal life functions; or (v) a congenital anomaly or birth defect; or (B) based on appropriate medical judgment, may jeopardize

therapeutic biologics, and vaccines[7] that emerge after licensure when the product is more widely used in the general population (i.e., as opposed to a carefully controlled clinical trial)[8]. Specifically, the FDA was to:

> "(i) develop methods to obtain access to disparate data sources including the data sources specified in subparagraph (C);
>
> [Subparagraph C specifies:
> (aa) federal health-related electronic data (such as data from the Medicare program and the health systems of the Department of Veterans Affairs);
> (bb) private sector health-related electronic data (such as pharmaceutical purchase data and health insurance claims data); and
> (cc) other data as the Secretary deems necessary to create a robust system to identify adverse events and potential drug safety signals;]
>
> (ii) develop validated methods for the establishment of a **postmarket risk identification and analysis system** to link and analyze safety data from multiple sources with the goals of including, in aggregate...at least 100,000,000 patients by July 1, 2012."[9] (emphasis added)

The FDA's implementation of this legislation – the Sentinel Initiative[10] – will enable users to systematically query distinct databases of patient-level data and return aggregated query results to gain knowledge on the postmarket risks and benefits of

---

the patient and may require a medical or surgical intervention to prevent an outcome described under subparagraph (A)."

[7] § 905(a)(3) in *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, 2007, codified at 21 U.S.C. § 355(k)(3)(A) clarifies that the system contains "information with respect to a drug approved under this section [section 355] or under section 351 of the Public Health Service Act." This scope technically covers blood/blood products and tissue/tissue products, but I limit this dissertation to drugs, therapeutic biologics (i.e., those regulated within the FDA's Center for Drug Evaluation and Research) and vaccines. As of July 9, 2012, the scope of the system was expanded to include medical devices, although I do not address this use in this dissertation. See § 615 in *Food and Drug Administration Safety and Innovation Act, Public Law 112-144*, 2012.

[8] Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 38., The IOM noted that medical product approval systems are characterized by an inherent "delayed availability of important safety data until a drug is used in larger and more diverse populations." Most, but not all, postmarket safety data arises from postmarket clinical experiences (as opposed to a clinical trial setting), and prior to the widespread availability of electronic healthcare data, was contained in individual case reports. See T. Brewer and G. A. Colditz, "Postmarketing Surveillance and Adverse Drug Reactions: Current Perspectives and Future Needs," *Journal of the American Medical Association* 281, no. 9 (1999): 824–829.

[9] § 905 in *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355(k)(3)(B)(i-ii) and 21 U.S.C. § 355(k)(3)(C)(i)(III).

[10] M. A. Robb et al., "The US Food and Drug Administration's Sentinel Initiative: Expanding the Horizons of Medical Product Safety," *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 9–11; R. E. Behrman et al., "Developing the Sentinel System - A National Resource for Evidence Development," *The New England Journal of Medicine* (2011); Food and Drug Administration U.S. Department of Health and Human Services, "The Sentinel Initiative" (FDA, 2010), http://www.fda.gov/downloads/Safety/FDAsSentinelInitiative/UCM233360.pdf.

medical products[11]. The FDA created a five-year pilot program – the Mini-Sentinel System – to build initial infrastructure, develop capabilities, test methods, and conduct pilot assessments.[12] It is important to understand that as of this writing, the Mini-Sentinel System is considered a "laboratory"[13] and its capabilities are at various stages of development. As a system, it is not yet *routinely* employed in postmarket safety assessments being conducted at the FDA.

First, a key feature of the Mini-Sentinel System is the *secondary* use, or repurposing, of electronic healthcare data in order to identify and/or assess postmarket safety signals. These data are currently administrative/claims data with some clinical data such as laboratory tests.[14] Second, the sizable scale and distributed architecture of this effort is another unique feature, with data that comprise nearly one-third of the privately insured U.S. population.[15] Third, the system's reusable data infrastructure facilitates the execution of conventional multi-site pharmacoepidemiologic studies[16].

Specifically, the Mini-Sentinel System eliminates the need to constitute a study database *de novo* to assess hypotheses with respect to postmarket safety signal(s) of interest. Instead, these data are continually maintained in a common, interoperable format[17] across multiple sites resulting in less time spent harmonizing data sources for each postmarket safety activity. Also, the use of these data has been designated as a

---

[11] Congress did not specify medical devices for inclusion in the active postmarket risk identification and analysis system until just recently. See § 615 of *Food and Drug Administration Safety and Innovation Act, Public Law 112-144*.

[12] R. Platt et al., "The U.S. Food and Drug Administration's Mini-Sentinel Program: Status and Direction," *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 1–8; Robb et al., "The US Food and Drug Administration's Sentinel Initiative: Expanding the Horizons of Medical Product Safety."

[13] Platt et al., "The U.S. Food and Drug Administration's Mini-Sentinel Program: Status and Direction."

[14] Ibid. Healthcare providers use administrative/claims data to charge health insurance companies for their services.

[15] See *infra* at note 113-114 for more details.

[16] Multi-site studies typically combine patient data, which requires data use agreements among the multiple sites as well as privacy and human subjects reviews at each site. For a general description of the conduct of pharmacoepidemiologic studies using databases, see U.S. Department of Health and Human Services et al., "Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets (Draft)", February 16, 2011, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM243537.pdf.

[17] L. H. Curtis et al., "Design Considerations, Architecture, and Use of the Mini-Sentinel Distributed Data System," *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 23–31.

public health activity as opposed to "research," which means that the Common Rule[18] does not apply to these activities and review by Institutional Review Boards is not required.[19] Practically, many front-end time delays and administrative requirements associated with testing drug safety hypotheses have been eliminated or reduced. Consequently, the system provides an efficient environment for the conduct of conventional pharmacoepidemiologic studies using administrative data.

More importantly, and perhaps Congress's *principal* intent, this data infrastructure also allows for new *routine* postmarket monitoring capabilities including 1) the retrieval of population-wide descriptive statistics on medical product usage[20], diagnoses, and outcomes; 2) the performance of statistical surveillance methods for the automated generation of new safety signals (e.g., data-mining and syndromic surveillance); and 3) the performance of sequential statistical analyses on pre-specified postmarket drug safety hypotheses (i.e., safety signals). As of this writing, some of these capabilities are in embryonic stages of development. However, I include them here to give the reader a broad understanding of the potential of the system. These capabilities are novel supplements to the FDA's existing systems because the data infrastructure makes data available in "near real-time,"[21] allowing for more timely evidence generation to support regulatory decision-making.

It is important to understand that "near real-time" is a *relative* concept. That is, *relative* to a conventional observational study with a singular end-of-study analysis,

---

[18] The Common Rule refers to federal regulations that protect the rights of human subjects involved in biomedical and behavioral research. It can be found in numerous instances within the entire Code of Federal Regulations but is classically located at 45 CFR § 46 in its entirety. Department of Health and Human Services, "Code of Federal Regulations Title 45 Part 46, Protection of Human Subjects", July 14, 2009, http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html.

[19] D. McGraw, K. Rosati, and B. Evans, "A Policy Framework for Public Health Uses of Electronic Health Data," *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 18–22.

[20] For many years, the FDA has had access to proprietary, nationally projected medical product utilization data. See Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 112. This capability is **new** because it will be possible to link data on medical product utilization to diagnoses and outcomes. For example, one might be able to assess "channeling" behavior, or the tendencies for certain types of patients to be prescribed particular products.

[21] "Near real-time" data refer to data on clinical experiences that arrive with a variable delay from when the experience occurred. There are two sources of delay. First, there is a processing delay, which is the time that elapses between when the experience occurs, and when it is recorded and available for analysis. Second, there is a refresh delay, which is associated with the frequency with which an originating data source renews their dataset and makes it available for analysis. These concepts will be explained in greater detail in subsection 6.2.2.

which might analyze data years after the outcomes occurred in time[22], "near real-time" data make analyses possible on the order of months[23]. This "near real-time" data stream allows regulators and public health investigators to analyze data as they accumulate, thereby generating population-wide incidence rates and risk estimates sooner than conventional observational studies. Yet, these gains in speed in a surveillance setting come with a price. They are offset by data that may be "unsettled" and later corrected[24], analyses performed with limited confounding control[25], analyses performed without full adjudication of exposures and outcomes[26], or analyses that rely on previous validation studies. Implementing more refined confounding control and/or adjudicating data increase time-to-results and cost because these activities move further away from automated analyses. However, these activities increase the quality of analyses by reducing the biases associated with the results[27], and may substantially affect regulatory decision-making. It remains to be seen whether regulators will favor speed over quality and how they will trade such quantities off for particular safety questions.

Still, if **reasonable but imperfect** information is available sooner, and it leads to regulatory action(s) that prevent adverse events, improve clinical care decisions, and conserve surveillance or research resources for other public health needs, then these gains

---

[22] For example, in a FDA-funded large retrospective observational study, the study period ended in 2005, a full five years prior to preliminary analyses. See L. A. Habel et al., "ADHD Medications and Risk of Serious Cardiovascular Events in Young and Middle-aged Adults," *JAMA : the Journal of the American Medical Association* 306, no. 24 (2011): 2673–2683; W. O. Cooper et al., "ADHD Drugs and Serious Cardiovascular Events in Children and Young Adults," *The New England Journal of Medicine* 365, no. 20 (2011): 1896–1904.

[23] While the Mini-Sentinel System's current data structure involves monthly data refreshes, older and similar distributed database networks like the Vaccine Safety Datalink collect and analyze data in weekly increments. See generally W. K. Yih et al., "Active Surveillance for Adverse Events: The Experience of the Vaccine Safety Datalink Project," *Pediatrics* 127 Suppl 1 (2011): S54–64.

[24] Recall that these data have a *primary* purpose - reimbursement of medical services – that may subject them to adjustments, rejections, corrections, re-submissions, and other changes for some period of time after their chronological occurrence.

[25] The degree of confounding control associated with the new capabilities of the Mini-Sentinel System will be discussed in greater detail in 4.2.1.1.

[26] Adjudication refers to procedures that are performed to validate the data, i.e. to ensure that the electronic record actually reflects patient experiences. It often involves medical chart abstraction and confirmation of the exposures, outcomes, and covariates of interest.

[27] For example, sequential database surveillance performed for H1N1 influenza vaccination-Guillain-Barre Syndrome (GBS) found that electronic identification of GBS had a positive predictive value of 53.3%. The resultant differences between risk estimates with chart-confirmed GBS vs. electronically identified GBS are substantial. See Appendix B in Grace M. Lee et al., "H1N1 and Seasonal Influenza Vaccine Safety in the Vaccine Safety Datalink Project," *American Journal of Preventive Medicine* 41, no. 2 (August 2011): 121–128. Biases will be described in greater detail in section 4.2.

must be weighed against the cost of operating and maintaining the Sentinel System at the scale specified by Congress[28]. Quantifying the value of this system first requires mapping the needs of the FDA in the postmarket to an analysis of the optimal use(s) of this system for each of its capabilities, and then determining the appropriate demand for such use(s).

Examining the proper, efficient, and *routine* incorporation of some of these novel capabilities into the existing postmarket surveillance systems of the FDA is the subject of this dissertation. This research develops qualitative and quantitative tools to aid the FDA in evaluating the Mini-Sentinel System's capabilities with regard to generating postmarket evidence to support regulatory decision-making. I consider drugs, therapeutic biologics, and vaccines as the exposures[29] of interest. The unit of analysis that I examine, and take as an input to my models, is a pre-specified exposure-outcome pair[30]. Thus, this dissertation does not address the optimal use of the Mini-Sentinel System for signal detection activities, which involve searching the data for new safety signals without pre-specified hypotheses of interest.

Like any technical system, the Mini-Sentinel System has design-based limitations and cannot fulfill all of the FDA's postmarket needs. These design-based limitations are summarized in a qualitative tool that the FDA can employ to determine whether the Mini-Sentinel System is well suited, *on its face*, to evaluate a pre-specified exposure-outcome pair. I call this qualitative tool the Mini-Sentinel System Pre-Screening Checklist and describe it herein in Section 4. Once an initial *qualitative* assessment is complete, quantitative tools may further be used to determine the Mini-Sentinel System's likelihood of meeting the FDA's needs for assessment. The quantitative tool I describe herein in Section 6 is the Sequential Database Surveillance Simulator. The simulator allows the FDA to explore the surveillance possibilities for a particular pre-specified exposure-

---

[28] Congress specified coverage of 100,000,000 persons by 2012 in *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*. It is unclear if a needs assessment was performed to justify this particular size.

[29] Rothman et al. define exposure as follows: "In epidemiology, it is customary to refer to potential causal characteristics as exposures. Thus, exposure can refer to a behavior (e.g., needle sharing), a treatment or other intervention (e.g., an educational program about hazards of needle sharing), a trait (e.g., a genotype), an exposure in the ordinary sense (e.g., an injection of contaminated blood), or even a disease (e.g., diabetes as the cause of death)." Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash, *Modern Epidemiology*, Third. (Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008), 52.

[30] Outcomes are health outcomes of interest, and safety surveillance particularly focuses on adverse events as health outcomes of interest. An exposure-outcome pair is a hypothesized relationship between the exposure and outcome of interest, e.g., oral anti-diabetic medications and acute myocardial infarctions.

outcome pair under a variety of potential real-world circumstances in a virtual, low-cost way. Using this tool, the FDA may begin to make assessments of the Mini-Sentinel System's sequential database surveillance capabilities for individual pairs. As the FDA considers the larger scope of exposure-outcome pairs that it needs to evaluate and uses the tool repeatedly, the FDA can begin to draw conclusions with regard to overall demand for this capability. I comment on, but do not fully examine questions of overall demand, which requires consideration of the entire scope of exposure-outcome pairs that might exist, an *n*-dimensional space.

A primary motivation to examine this topic is the public investment required to maintain the infrastructure in the long term. First, although the infrastructure is not explicitly funded beyond the pilot Mini-Sentinel System, the incorporation of this system into routine postmarket surveillance activities requires non-trivial annual maintenance and operation costs. These costs simply keep the data up-to-date and capable of being accessed. Second, each request to access these data has some processing costs that clearly vary with the number and size of the data requests in addition to how efficiently the requests can be coded, tested, and distributed to the individual sites who access their proprietary databases. Third, as the system is used for more protocol-oriented (i.e., *ad hoc*) analyses, senior scientific support (e.g., epidemiologists, clinicians, and biostatisticians) will be necessary for each of these assessments. Fourth, the FDA must scale its regulatory efforts (i.e., full-time trained staff) to process the output assessments of this system. Taken together, continued funding of this resource requires budgeting for both fixed and variable costs that are not yet well-defined.

Another motivation is the public health effects of the legal coupling[31] of the Mini-Sentinel System's capabilities to the FDA's ability to require industry-funded postmarket studies or clinical trials, known as postmarket requirements (PMRs)[32]. Specifically, when identifying or assessing a particular postmarket safety signal, the FDA must make a

---

[31] § 901 in *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at U.S.C. § 355(o)(3)(D).

[32] A postmarketing requirement is a mandate from the FDA to the manufacturer/sponsor of a particular product to perform a study of various types. See Department of Health and Human Services et al., "Guidance for Industry: Postmarketing Studies and Clinical Trials - Implementation of Section 505(o)(3) of the Federal Food, Drug, and Cosmetic Act (Final)", March 31, 2011, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM172001.pdf.

determination on whether its publicly-funded postmarket safety systems (including the Mini-Sentinel System) generate **sufficient** knowledge on postmarket safety risks to make a regulatory decision, or whether additional privately-funded (i.e. manufacturer-funded) studies are necessary. These decisions clearly hold significant implications for the FDA, manufacturers/sponsors, healthcare providers, and patients. Such decisions are likely to be scrutinized and perhaps contested, and depend on a clear scientific demonstration of suitable use of the system. The ethical implications of this decision were the subject of a recent IOM committee report, which noted,

> "When the FDA imposes a postmarketing requirement, it is expressing...a judgment that the public health interests served by requiring additional research outweigh the burdens placed on pharmaceutical manufacturers and – more important from an ethical standpoint —any risk of harm or burdens on research participants."[33]

In this dissertation, Section 2 reviews the FDA's routinely operating postmarket systems (i.e., the existing postmarket systems), and the addition of the "pilot" Mini-Sentinel System. Specifically, the Mini-Sentinel System's data infrastructure, and risk identification and analysis capabilities are reviewed. Section 3 establishes the legal/policy context in which the FDA's postmarket systems are now embedded and more thoroughly explains the aforementioned legal coupling of the Mini-Sentinel System's technical capabilities to the FDA's regulatory ability to require privately-funded postmarket studies. That is, it examines the legal requirements with regard to the scientific quality of the evidence needed to support various regulatory actions. Additionally, Section 3 outlines the FDA's postmarket regulatory decision-making process and potential regulatory actions that may result following identification and analysis of new safety information. Section 4 presents the Mini-Sentinel System Pre-Screening Checklist, a qualitative tool designed to aid the decision-maker in evaluating whether the Mini-Sentinel System is likely to be suited, *on its face*, to evaluate particular exposure-outcome pairs. Specifically, Section 4 addresses situations when the Mini-Sentinel System may be ill-suited as an evidence generation system of interest due to its inability to overcome various biases in an observational setting, issues of sample size, or issues of generalizability to broader populations.

---

[33] Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs* (Washington, D.C.: National Academies Press, 2012), 4, http://books.nap.edu/catalog/13219.html.

Section 5 reviews the scientific and technical state-of-the-art with respect to prospective sequential database surveillance. Section 6 describes the Sequential Database Surveillance Simulator in detail. Section 7 illustrates the use of the Sequential Database Surveillance Simulator with a vaccine-based example. Particularly, Section 7 establishes the important tradeoffs associated with sample size calculations in sequential statistical analyses, particularly the tradeoff between statistical power and median sample size. It draws attention to the performance characteristics of various surveillance configurations with respect to timeliness and accuracy of signal detection. Section 8 adds complexity to the vaccine example by examining the effects of misclassification on sequential database surveillance performance. It also establishes a way to use the simulator to investigate the performance of different algorithms for detecting outcomes of interest. This simulated example is particularly important because it addresses inherent differences in the Mini-Sentinel System's component databases, particularly the existence of a small subset of databases with access to richer laboratory data. Sequential database surveillance performance is compared between this small subset with high quality data to the larger database configuration with claims-only data.

Section 9 addresses complications related to modeling medical product adoption and utilization, specifically the uptake of new molecular entities. A cohort of 40 new molecular entities is examined. A subset of this cohort is then investigated by fitting Mini-Sentinel System data to classical diffusion models. An important finding that results is that many new molecular entities are better described by "dual market" adoption patterns. In other words, the market for these medical products consists of two distinct sets of adopters. This finding raises a non-trivial generalizability concern when sequential database surveillance is performed immediately post-licensure. Section 10 is a summary of the findings of the dissertation and a discussion on future work. Appendices A-C contain supporting data and Appendix D is a glossary of terms to assist the reader.

THIS PAGE INTENTIONALLY LEFT BLANK

2 The FDA'S POSTMARKET SYSTEMS

2.1 The FDA's Routinely Operating Postmarket Systems

The Institute of Medicine (IOM)'s overview of the U.S. Food and Drug Administration (FDA)'s postmarket systems and their contribution to identifying and assessing medical product-associated risks is briefly reviewed herein. As stated earlier, because the FDA has not yet deployed the Mini-Sentinel System in *routine postmarket assessment,* the overview presented in this subsection still reflects the current state at the FDA.

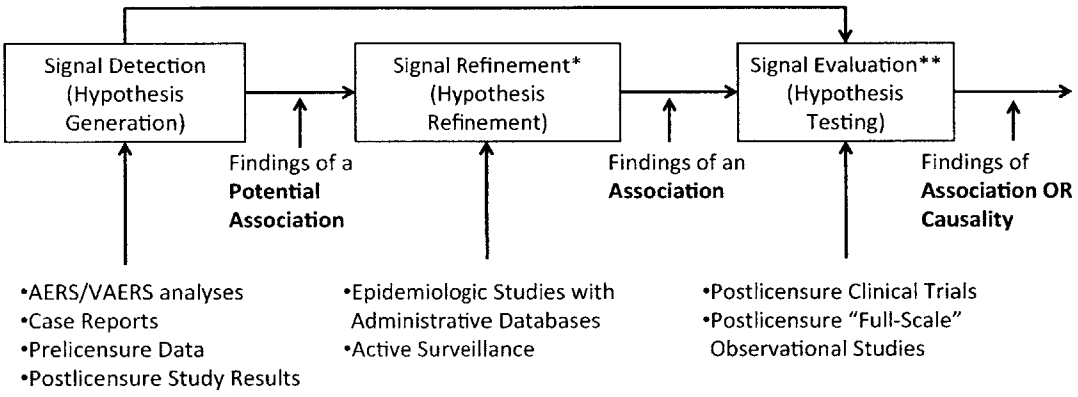

Figure 1. Identification and Adjudication of a Signal of Serious Risk in the U.S. Food and Drug Administration's Postmarket Surveillance Systems
Adapted from the Institute of Medicine.[34]
*In the original Institute of Medicine report, this stage was referred to as "signal strengthening or testing." The wording was changed to reflect a more neutral stance.
**In the original Institute of Medicine report, this stage was referred to as "signal confirmation." The wording was changed to reflect a more neutral stance.
Abbreviations: AERS, Adverse Event Reporting System; VAERS, Vaccine Adverse Event Reporting System.

Figure 1 shows the three-stage process for identification and adjudication of medical product-associated postmarket safety risks, often referred to as safety signals[35].

[34] This figure is generated from descriptions contained in Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 105–119.
[35] Two well-accepted definitions of "signal" are as follows: 1) "Information that arises from one or multiple sources (including observations and experiments) which suggests a new potentially causal association, or a new aspect of a known association, between an intervention and an event or set of related events, either adverse or beneficial, that is judged to be of sufficient likelihood to justify verificatory action." Council for International Organizations of Medical Sciences. Working Group VIII, *Practical aspects of signal*

2.1.1 Signal Detection

Signal detection[36] is the generation of a hypothesis with respect to a signal of serious risk[37] associated with a medical product. It is usually based on the discovery of new data and subsequent analyses of those data. In this first stage in the lifecycle of a safety signal, the FDA's spontaneous reporting systems - the Adverse Event Reporting System[38] and Vaccine Adverse Event Reporting System[39] - generate the bulk of hypotheses regarding potential associations[40] between medical products (i.e., exposures) and adverse events (i.e., outcomes).[41] Disproportionality analyses are used to identify exposure-outcome pairs that are reported in excess of what would be expected if these pairs were

---

detection in pharmacovigilance : report of CIOMS Working Group VIII. (Geneva: CIOMS, 2010). 2) "Reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously. Usually more than a single report is required to generate a signal, depending upon the seriousness of the event and the quality of the information." World Health Organization, *The Importance of Pharmacovigilance - Safety Monitoring of Medicinal Products* (Geneva, Switzerland: World Health Organization, 2002), 42, http://apps.who.int/medicinedocs/en/d/Js4893e/. See also Manfred Hauben and Jeffrey K Aronson, "Defining 'Signal' and Its Subtypes in Pharmacovigilance Based on a Systematic Review of Previous Definitions," *Drug Safety* 32, no. 2 (2009): 99–110.

[36] The FDA has also used the following definition of signal generation (which is generally an interchangeable term with signal detection): "an approach that uses statistical methods to identify medical product–adverse outcome associations that may be safety signals; no particular medical product exposure or adverse outcome is pre-specified." See Robb et al., "The US Food and Drug Administration's Sentinel Initiative: Expanding the Horizons of Medical Product Safety," 10.

[37] § 901 in *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*,.codified at 21 U.S.C. § 355-1(b)(6), gives the legislative definition: "The term 'signal of a serious risk' means information related to a serious adverse drug experience associated with use of a drug."

[38] See generally, Gerald J. Dal Pan, Marie Lindquist, and Kate Gelperin, "Postmarketing Spontaneous Pharmacovigilance Reporting Systems," in *Pharmacoepidemiology*, ed. Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy, Fifth. (John Wiley & Sons, 2011), 137–157; S. R. Ahmad, "Adverse Drug Event Monitoring at the Food and Drug Administration," *Journal of General Internal Medicine* 18, no. 1 (2003): 57–60.

[39] See generally, R. Ball et al., "Statistical, Epidemiological, and Risk-assessment Approaches to Evaluating Safety of Vaccines Throughout the Life Cycle at the Food and Drug Administration," *Pediatrics* 127 Suppl 1 (2011): S31–8.

[40] Associations in epidemiology are referred to as correlations in other fields of study. Strom defines an association as a statistically significant inference regarding a population. He further defines types of associations, one of which is causal when a biological inference establishes causation and all confounding is eliminated. See Brian L. Strom, "Basic Principles of Clinical Epidemiology Relevant to Pharmacoepidemiologic Studies," in *Pharmacoepidemiology*, ed. Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy, Fifth. (John Wiley & Sons, 2011), 38–43; C. H. Hennekens and D. DeMets, "Statistical Association and Causation: Contributions of Different Types of Evidence," *JAMA : the Journal of the American Medical Association* 305, no. 11 (2011): 1134–1135.

[41] Thomas J Moore, Sonal Singh, and Curt D Furberg, "The FDA and New Safety Warnings," *Archives of Internal Medicine* 172, no. 1 (January 9, 2012): 78–80; D. K. Wysowski and L. Swartz, "Adverse Drug Event Surveillance and Drug Withdrawals in the United States, 1969-2002: The Importance of Reporting Suspected Reactions," *Archives of Internal Medicine* 165, no. 12 (2005): 1363–1369.

independently distributed in the database.[42] Under certain conditions, it is also possible to identify a signal based on a single case report. For example, the FDA recently identified a single potentially medical product-associated death as a signal[43], and other outcomes may require only a few cases to generate a signal.[44]

Spontaneous reporting systems provide effective and inexpensive signal detection, particularly for rare or very rare[45] outcomes that are not detectable in clinical trials and that are unlikely to be related to the disease being treated. These types of signals are sometimes categorized as "Type B" adverse events, meaning they are considered to be idiosyncratic events that occur in patients with some (typically) unknown hypersensitivity or predisposing condition.[46] However, spontaneous reporting systems are a) known to have data quality problems; b) be subject to significant underreporting with regard to common outcomes (e.g., acute myocardial infarctions); c) be subject to significant underreporting when the outcomes are unlikely to trigger suspicion of being medical product-related; and d) be ineffective at calculating incidence or prevalence of an event in the population.[47] In summary, these systems provide a useful but limited function (i.e., they perform well for Type B signal detection).

---

[42] J S Almenoff et al., "Novel Statistical Tools for Monitoring the Safety of Marketed Drugs," *Clinical Pharmacology and Therapeutics* 82, no. 2 (August 2007): 157–166.

[43] In this case, there was a labeled, known risk regarding fingolimod-associated bradycardia, which may lead to death. To manage the risk, the label recommended increased monitoring for the first six hours following the initial dose. This death occurred in a 24-hour period following the initial dose despite compliance with the prescribed 6-hour monitoring period, suggesting an unexpected increase in the severity of risk. This increased severity was the important factor in identifying a *new* signal. See Center for Drug Evaluation and Research, "Drug Safety and Availability - FDA Drug Safety Communication: Safety Review of a Reported Death After the First Dose of Multiple Sclerosis Drug Gilenya (fingolimod)," WebContent, December 20, 2011, http://www.fda.gov/Drugs/DrugSafety/ucm284240.htm.

[44] Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 108–109. "Even a small number of reports of events that are commonly caused by drug exposure, such as liver or kidney failure, aplastic anemia, anaphylaxis, Stevens-Johnson syndrome, and so on, can constitute an important safety signal."

[45] The Council for International Organizations of Medical Sciences (CIOMS) Working Group III, *Guidelines for Preparing Core Clinical Safety Information on Drugs.* (Geneva: World Health Organization (WHO), 1995)."Rare" outcomes refer to those that occur with a frequency of greater than 1 event per 10,000 person-years, but less than 1 event per 1,000 person years. "Very Rare" outcomes occur with a frequency less than 1 event per 10,000 person-years but greater than 1 event per 100,000 person-years.

[46] R H Meyboom, M Lindquist, and A C Egberts, "An ABC of Drug-related Problems," *Drug Safety: An International Journal of Medical Toxicology and Drug Experience* 22, no. 6 (June 2000): 415–423.

[47] B. L. Strom, "Potential for Conflict of Interest in the Evaluation of Suspected Adverse Drug Reactions: a Counterpoint," *Journal of the American Medical Association* 292, no. 21 (2004): 2643–2646; U.S. General Accounting Office, *Adverse Events: The Magnitude of Health Risk Is Uncertain Because of Limited Incidence Data*, vol. GAO/HEHS-00–21 (Washington, DC: GPO, 2000); T. J. Moore, B. M. Psaty, and C. D. Furberg, "Time to Act on Drug Safety," *Journal of the American Medical Association* 279, no. 19

Signal detection with regard to more common outcomes, or those that may be related to the natural history of the disease, are typically detected from safety data in a prelicensure or postlicensure clinical trial[48], or meta-analyses of randomized clinical trial data[49]. These data may be underpowered statistically to suggest an association, but be biologically plausible and concerning enough to generate a signal. These types of signals have been referred to as "Type C" adverse effects or "statistical effects."[50] In the absence of large sample sizes, they are difficult to detect because of the high background frequency of the adverse event in the unexposed population. Notably, the Mini-Sentinel System enables such large sample sizes, implying a potential improvement to the FDA's signal detection capabilities for Type C signals once the Mini-Sentinel System is incorporated into routine operations. Signal detection in the Mini-Sentinel System is an active area of ongoing research that will be described briefly in section 2.2.2.2.

Once a safety signal has been detected, the FDA must determine whether the available evidence requires it to take immediate regulatory action, to gather more information (e.g., conduct postmarket studies[51]), or to do both simultaneously.[52]

2.1.2 Signal Refinement

The IOM identified two follow-on stages for signal adjudication after a signal has been detected. It differentiated these stages based on the type of postmarket study that was used to test the hypothesis of a medical product-associated risk. The IOM identified the second stage as a "signal strengthening and testing" stage, which is now known as

---

(1998): 1571–1573; Brewer and Colditz, "Postmarketing Surveillance and Adverse Drug Reactions: Current Perspectives and Future Needs."

[48] For example, cardiovascular safety signals for Vioxx® (rofecoxib) were identified by the FDA's Medical Officer in prelicensure clinical trials. See B. M. Psaty and C. D. Furberg, "COX-2 Inhibitors--lessons in Drug Safety," *The New England Journal of Medicine* 352, no. 11 (2005): 1133–1135.

[49] For example, a published meta-analysis of Avandia® (rosiglitazone) prompted unplanned interim analyses of trial data. See D. M. Nathan, "Rosiglitazone and Cardiotoxicity--weighing the Evidence," *The New England Journal of Medicine* 357, no. 1 (2007): 64–66; B. M. Psaty and C. D. Furberg, "The Record on Rosiglitazone and the Risk of Myocardial Infarction," *The New England Journal of Medicine* 357, no. 1 (2007): 67–69; S. E. Nissen and K. Wolski, "Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes," *The New England Journal of Medicine* 356, no. 24 (2007): 2457–2471.

[50] Meyboom, Lindquist, and Egberts, "An ABC of Drug-related Problems."

[51] Postmarket studies may include laboratory studies, animal studies, or clinical investigations. For examples, see Department of Health and Human Services et al., "Guidance for Industry: Postmarketing Studies and Clinical Trials - Implementation of Section 505(o)(3) of the Federal Food, Drug, and Cosmetic Act (Final)."

[52] The FDA's decision algorithm is discussed more thoroughly herein in section 3.

signal refinement[53]. The postmarket studies used to support evidence development in this stage include conventional pharmacoepidemiologic research using administrative databases and active surveillance of emergency rooms for adverse drug events.

Generally, pharmacoepidemiologic research using administrative databases can refer to descriptive studies that report drug utilization[54] or physician prescribing behavior[55], or to analytical studies designed to produce information on comparative drug safety[56] or effectiveness[57]. Administrative database studies can also evaluate the changes in health outcomes produced by changes in health policy such as the effect of changing insurance copayment requirements.[58] These studies sample electronic healthcare data, which can be either administrative/claims data or electronic health records. Within the drug safety realm, much has been written on the strengths and limitations of these studies.[59] Advantages include the ability to study large sample sizes inexpensively and the potential for greater generalizability than results from more exclusive randomized experiments. Disadvantages include the potential for biased results from poor quality and missing

[53] The FDA has used the following definition of signal refinement: "a process by which an identified potential safety signal is further investigated to determine whether evidence exists to support a relationship between the medical product exposure and the outcome." See Robb et al., "The US Food and Drug Administration's Sentinel Initiative: Expanding the Horizons of Medical Product Safety," 10.

[54] See, for example, Paul N Pfeiffer et al., "Depression Care Following Psychiatric Hospitalization in the Veterans Health Administration," *The American Journal of Managed Care* 17, no. 9 (September 2011): e358–364.

[55] See for example, Paul N Pfeiffer et al., "Trends in Antidepressant Prescribing for New Episodes of Depression and Implications for Health System Quality Measures," *Medical Care* 50, no. 1 (January 2012): 86–90.

[56] See, for example, Habel et al., "ADHD Medications and Risk of Serious Cardiovascular Events in Young and Middle-aged Adults"; Cooper et al., "ADHD Drugs and Serious Cardiovascular Events in Children and Young Adults."

[57] See, for example, Seo Young Kim and Daniel H Solomon, "Use of Administrative Claims Data for Comparative Effectiveness Research of Rheumatoid Arthritis Treatments," *Arthritis Research & Therapy* 13, no. 5 (2011): 129.

[58] See, for example, Sujha Subramanian, "Impact of Medicaid Copayments on Patients with Cancer: Lessons for Medicaid Expansion Under Health Reform," *Medical Care* 49, no. 9 (September 2011): 842–847.

[59] W. A. Ray, "Improving Automated Database Studies," *Epidemiology (Cambridge, Mass.)* 22, no. 3 (2011): 302–304; B. L. Strom, "Methodologic Challenges to Studying Patient Safety and Comparative Effectiveness," *Medical Care* 45, no. 10 Supl 2 (2007): S13–5; S. Schneeweiss and J. Avorn, "A Review of Uses of Health Care Utilization Databases for Epidemiologic Research on Therapeutics," *Journal of Clinical Epidemiology* 58, no. 4 (2005): 323–337; R. Temple, "Meta-analysis and Epidemiologic Studies in Drug Development and Postmarketing Surveillance," *JAMA : the Journal of the American Medical Association* 281, no. 9 (1999): 841–844.

data.[60] As the IOM states, "precise but biased estimates of risk are not generally useful."[61]

Consequently, it is not unusual for independently conducted pharmacoepidemiologic studies to produce variation in their estimates of risk[62], and although the same possibility exists with randomized studies[63], there is more skepticism of pharmacoepidemiologic studies.[64] Some attribute these problems to growing pains of the discipline.[65] On that note, there have been efforts to increase the internal validity and repeatability of these studies.[66] Despite these potential shortcomings, the FDA funds a limited number of pharmacoepidemiologic studies via contract.[67] Manufacturers also conduct these studies

---

[60] Biases associated with observational data are discussed extensively herein in section 4.2.

[61] Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 114.

[62] See, for example, the FDA's struggle to interpret the results of seven observational studies studying the risks of venous thromboembolism associated with drospirenone-containing contraceptives. See Center for Drug Evaluation and Research, "Drug Safety and Availability - FDA Drug Safety Communication: Updated Information About the Risk of Blood Clots in Women Taking Birth Control Pills Containing Drospirenone," WebContent, April 10, 2012, http://www.fda.gov/Drugs/DrugSafety/ucm299305.htm; Center for Drug Evaluation and Research, "Reproductive Health Drugs Advisory Committee - Briefing Information for the December 9, 2011 Joint Meeting of the Advisory Committee for Reproductive Health Drugs and the Drug Safety and Risk Management Advisory Committee," WebContent, December 9, 2011, http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ReproductiveHealthDrugsAdvisoryCommittee/ucm282631.htm.

[63] D. Jane-wit, R. I. Horwitz, and J. Concato, "Variation in Results from Randomized, Controlled Trials: Stochastic or Systematic?," *Journal of Clinical Epidemiology* 63, no. 1 (2010): 56–63.

[64] D. A. Lawlor et al., "Those Confounded Vitamins: What Can We Learn from the Differences Between Observational Versus Randomised Trial Evidence?," *Lancet* 363, no. 9422 (2004): 1724–1727; S. J. Pocock and D. R. Elbourne, "Randomized Trials or Observational Tribulations?," *The New England Journal of Medicine* 342, no. 25 (2000): 1907–1909.

[65] J. Avorn, "In Defense of Pharmacoepidemiology--embracing the Yin and Yang of Drug Research," *The New England Journal of Medicine* 357, no. 22 (2007): 2219–2221; E. von Elm and M. Egger, "The Scandal of Poor Epidemiological Research," *BMJ (Clinical Research Ed.)* 329, no. 7471 (2004): 868–869.

[66] E. von Elm et al., "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies," *Annals of Internal Medicine* 147, no. 8 (2007): 573–577; G. C. Hall et al., "Guidelines for Good Database Selection and Use in Pharmacoepidemiology Research," *Pharmacoepidemiology and Drug Safety* 21, no. 1 (2012): 1–10; U.S. Department of Health and Human Services et al., "Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets (Draft)."

[67] U.S. Government Accountability Office, *FDA Has Begun Efforts to Enhance Postmarket Safety, but Additional Actions Are Needed*, vol. GAO-10–68 (Washington, DC: GPO, 2009), 28–30, http://www.gao.gov/products/GAO-10-68."Since FDA initially awarded about $5.4 million in total to these companies in fiscal year 2005, these contracts have yielded five completed epidemiologic studies on drug safety, including a study on how antidepressant use in pregnancy affects the health of newborns. In fiscal year 2008, FDA added about $9 million in total to the four contracts."

to fulfill postmarket study commitments or requirements to assess particular safety signals.[68]

Active surveillance of emergency rooms via the National Electronic Injury Surveillance System Cooperative Adverse Drug Event Surveillance program is another means of generating descriptive statistics with respect to physician-identified adverse drug experiences. Implemented as a joint venture of the FDA, the Centers for Disease Control and Prevention and the Consumer Products Safety Commission, this program employs trained personnel to routinely review patient records to document physician-identified adverse drug experiences in a nationally representative sample of emergency departments.[69] This type of active surveillance establishes incidence and prevalence data, particularly among medical products with a "narrow therapeutic range" that cause easily-recognized serious adverse events when used at improperly high dosages (i.e., overdoses).[70] These medical product-associated adverse events – also designated commonly as "Type A" adverse effects[71] - are predictable, and normally resolve or do not recur when the dose is reduced. Generally, there is no question of causality, and these statistics may contribute to the FDA's efforts to manage known risks associated with supratherapeutic effects.

Overall, the IOM described the output of the signal refinement stage as:

> "provid[ing] guidance for the development of further studies or provid[ing] sufficient information to narrow the uncertainty about drug-related risks and benefits and guide regulatory actions and the decisions of patients and providers."[72]

---

[68] For example, see D. D. Dore et al., "A Cohort Study of Acute Pancreatitis in Relation to Exenatide Use," *Diabetes, Obesity & Metabolism* 13, no. 6 (2011): 559–566; D. D. Dore, J. D. Seeger, and K. Arnold Chan, "Use of a Claims-based Active Drug Safety Surveillance System to Assess the Risk of Acute Pancreatitis with Exenatide or Sitagliptin Compared to Metformin or Glyburide," *Current Medical Research and Opinion* 25, no. 4 (2009): 1019–1027. These studies were used to fulfill postmarket commitments. See commitment IDs 221528-221541 in Center for Drug Evaluation and Research, "Postmarket Requirements and Commitments Database", n.d., http://www.accessdata.fda.gov/scripts/cder/pmc/index.cfm.

[69] D. S. Budnitz et al., "National Surveillance of Emergency Department Visits for Outpatient Adverse Drug Events," *Journal of the American Medical Association* 296, no. 15 (2006): 1858–1866.

[70] For example, "insulin or warfarin was implicated in more than one quarter of all estimated hospitalizations." In Ibid., 1863. The denominator refers to all emergency department visits with identifiable adverse drug events that then resulted in hospitalizations.

[71] Meyboom, Lindquist, and Egberts, "An ABC of Drug-related Problems."

[72] Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 115.

In other words, the outputs of the signal refinement stage can simply precede further investigation or can terminate in regulatory action.

2.1.3 Signal Evaluation

The IOM identified "signal confirmation" as the third stage, which is now termed signal evaluation[73]. During signal evaluation, evidence development is generated with the conduct of "full-scale observational studies and clinical trials."[74] The intent of such studies is to establish causal relationships as opposed to statistical associations.[75] It is unclear exactly what the IOM meant when describing "full-scale" observational studies.[76] Moreover, the committee then references two *randomized* comparative safety and effectiveness studies as exemplars of this stage. These two National Institutes of Health-funded studies[77] are noteworthy for their comparison of multiple therapeutic options, their wide exploration of both safety/effectiveness endpoints[78], and their cost, at $125 million and $725 million.[79] In general, the costs and operational feasibility of these

---

[73] The FDA has used the following definition: "*Signal evaluation* consists of the implementation of a full epidemiological analysis to more thoroughly evaluate the causal relationship between exposure to the medical product and the adverse outcome of interest." Robb et al., "The US Food and Drug Administration's Sentinel Initiative: Expanding the Horizons of Medical Product Safety," 10.

[74] Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 115.

[75] For a definition of association, see *supra* at note 40.

[76] While the IOM is not clear on these distinctions, they did segregate "full-scale" observational studies from retrospective studies in administrative databases. This designation is unsettled and the subject of consternation among many epidemiologists who believe that properly executed database studies with significant validation and confounding control can be used for causal inference. See *supra* at note 73, implying that database studies can support causal inference. I hypothesize that the IOM's idea of full-scale observational studies likely involved primary data collection, but there is little supporting evidence to confirm or refute this conjecture.

[77] ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial, "Major Outcomes in High-risk Hypertensive Patients Randomized to Angiotensin-converting Enzyme Inhibitor or Calcium Channel Blocker Vs Diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT)," *Journal of the American Medical Association* 288, no. 23 (2002): 2981–2997; J. E. Rossouw et al., "Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results From the Women's Health Initiative Randomized Controlled Trial," *Journal of the American Medical Association* 288, no. 3 (2002): 321–333.

[78] Dual safety/effectiveness endpoints occur when considering therapeutics that are licensed to lower cardiovascular risk. These study endpoints are, by nature, both related to safety and effectiveness. In comparison, Vioxx® (rofecoxib) was licensed to treat osteoarthritis and so cardiovascular risk associated with it is a more strictly safety-related endpoint.

[79] Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 115.

studies result in their undersupply.[80] However, therapeutics approved with surrogate endpoints[81] under the FDA's accelerated approval process require such trials, which are primarily focused on demonstrating efficacy as opposed to safety.

Absent the accelerated approval mechanism or other special program, prior to the 2007 Food and Drug Administration Amendments Act (FDAAA), the FDA had no legal authority to mandate postmarket studies – either database studies, "full-scale" observational studies or randomized clinical trials - and relied on voluntary commitments from sponsors.[82] For example, if a therapeutic was approved with a surrogate endpoint under the traditional approval process, then the FDA secured a non-enforceable postmarket commitment to complete follow-up studies.[83] Many of these studies were never initiated or completed.[84] The IOM attributed this low completion rate to inadequately and hastily designed studies that were not practical.[85] The FDAAA broadened the FDA's authority to mandate these studies[86] and required the FDA to investigate the backlog of postmarket commitments annually.[87] Since then, the FDA has

---

[80] R. F. Reynolds et al., "Is the Large Simple Trial Design Used for Comparative, Post-approval Safety Research? A Review of a Clinical Trials Registry and the Published Literature," *Drug Safety : an International Journal of Medical Toxicology and Drug Experience* 34, no. 10 (2011): 799–820; D. Carpenter, "A Proposal for Financing Postmarketing Drug Safety Studies by Augmenting FDA User Fees," *Health Affairs (Project Hope)* Suppl Web Exclusives (2005): W5–469–80; S. R. Tunis, D. B. Stryer, and C. M. Clancy, "Practical Clinical Trials: Increasing the Value of Clinical Research for Decision Making in Clinical and Health Policy," *Journal of the American Medical Association* 290, no. 12 (2003): 1624–1632.

[81] A surrogate endpoint is hypothesized to be a proxy measure of a clinical endpoint that defines the therapeutic's benefit. For example, progression-free survival is an oft used surrogate endpoint for morbidity/mortality associated with cancer. See U.S. Government Accountability Office, *FDA Needs to Enhance Its Oversight of Drugs Approved the Basis of Surrogate Endpoints*, GAO-09-866 (Washington, DC: GPO, 2009), http://www.gao.gov/products/GAO-09-866.

[82] Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 155–157; Department of Health and Human Services et al., "Guidance for Industry: Postmarketing Studies and Clinical Trials - Implementation of Section 505(o)(3) of the Federal Food, Drug, and Cosmetic Act (Final)."

[83] U.S. Government Accountability Office, *FDA Needs to Enhance Its Oversight of Drugs Approved the Basis of Surrogate Endpoints.*

[84] Department of Health and Human Services Office of the Inspector General, *FDA's Monitoring of Postmarket Study Commitments*, vol. OEI-01–04–00390 (Washington, DC: OIG, 2006).

[85] Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 111–116.

[86] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355(o)(3).

[87] § 921 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355(k)(5)(C).

substantially changed their tracking system for these studies[88] and many commitments have been fulfilled or cancelled after further review.[89]

In general, signal evaluation studies designed explicitly for safety endpoints, as opposed to following up surrogate endpoints, are rare.[90] There is active debate about whether randomized controlled trials are the most appropriate ethical and scientific approach to address safety issues in the postmarket.[91] Concerns include the equipoise among participants when the purpose of the study is to establish proof of harm, the deviations from the ideal randomized controlled trial design that change statistical inferences, and the appropriateness of particular analyses (e.g., intention-to-treat[92]). While randomized controlled trials are commonly accepted as the gold standard for proof of efficacy, internal disagreements at the FDA regarding the type of evidence required to adjudicate the rosiglitazone-cardiovascular outcomes safety signal prompted the FDA to reach out to the IOM for guidance on the matter.[93] Essentially, the IOM found legitimate ethical and scientific arguments on both sides, and stressed the needs both to reach out to advisory committees and the public for guidance and to act transparently regarding these more contentious decisions on a case-by-case basis.

The FDA has taken this case-by-case approach. With respect to rosiglitazone, the FDA canceled an ongoing randomized controlled trial and issued a risk evaluation and

---

[88] See Center for Drug Evaluation and Research, "Postmarket Requirements and Commitments Database."
[89] See Center for Drug Evaluation and Research, "Postmarketing Requirements and Commitments: Reports," WebContent, n.d., http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/PostmarketingPhaseIVCommitments/ucm064436.htm; U.S. Government Accountability Office, *FDA Needs to Enhance Its Oversight of Drugs Approved the Basis of Surrogate Endpoints*.
[90] Reynolds et al., "Is the Large Simple Trial Design Used for Comparative, Post-approval Safety Research? A Review of a Clinical Trials Registry and the Published Literature." Many phase IV randomized controlled trials or large simple trials are designed with the primary purpose of the study being to assess an efficacy or effectiveness endpoint but where data on the general safety profile were also collected.
[91] See Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*; J. P. Vandenbroucke, "Why Do the Results of Randomised and Observational Studies Differ?," *BMJ (Clinical Research Ed.)* 343 (2011): d7020; P. N. Papanikolaou, G. D. Christidi, and J. P. Ioannidis, "Comparison of Evidence on Harms of Medical Interventions in Randomized and Nonrandomized Studies," *CMAJ : Canadian Medical Association Journal* 174, no. 5 (2006): 635–641; J. P. Vandenbroucke, "What Is the Best Evidence for Determining Harms of Medical Treatment?," *CMAJ : Canadian Medical Association Journal* 174, no. 5 (2006): 645–646.;
[92] See Miguel A Hernán and Sonia Hernández-Díaz, "Beyond the Intention-to-treat in Comparative Effectiveness Research," *Clinical Trials (London, England)* 9, no. 1 (February 2012): 48–55.
[93] See Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, xvii.

mitigation strategy, which significantly limits utilization of rosiglitazone.[94] More recently, the FDA held several advisory committee meetings to address the type of signal evaluation study that would be required to evaluate a long-acting beta agonist-death signal in children.[95] In this case, the FDA required four manufacturers to conduct randomized controlled trials, and to collect data in a way that makes the results amenable to future meta-analysis.[96] In general, both Congress and the IOM have placed a heavy justification burden on the FDA regarding engaging in signal evaluation studies.

2.1.4 Other Studies

Notably, the IOM does not describe how the results of meta-analyses[97] should be regarded in its three-phase structure. Attention to meta-analyses has increased following publication of results that identified the rosiglitazone-cardiovascular outcomes safety signal.[98] Concerns about meta-analyses relate to the heterogeneity of contributing studies and specifically, how exposures, outcomes, and covariates are classified; and what types of patients are included/excluded.[99] Essentially, combining potentially dissimilar data requires many judgments from investigators with little established guidance regarding how to pool data. In a recent study, Golder et al. found no comparable difference in risk estimates derived from meta-analyses of randomized controlled trials as opposed to meta-

---

[94] Center for Drug Evaluation and Research, "Postmarket Drug Safety Information for Patients and Providers - HHS FDA: Briefing on Avandia," WebContent, September 23, 2010, http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm227934.htm.

[95] Center for Drug Evaluation and Research, "Pulmonary-Allergy Drugs Advisory Committee - Briefing Information for the March 10-11, 2010 Joint Meeting of the Pulmonary-Allergy Drugs Advisory Committee and Drug Safety and Risk Management Committee," WebContent, March 10, 2010, http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/Pulmonary-AllergyDrugsAdvisoryCommittee/ucm202692.htm.

[96] Food and Drug Administration and Center for Drug Evaluation and Research, "FDA Drug Safety Communication: FDA Requires Post-market Safety Trials for Long-Acting Beta-Agonists (LABAs)," WebContent, April 15, 2011, http://www.fda.gov/Drugs/DrugSafety/ucm251512.htm.

[97] Berlin et al. define meta-analysis as "the statistical analysis of a collection of analytic results for the purpose of integrating the findings." Jesse A. Berlin, M. Soledad Cepeda, and Carin J. Kim, "The Use of Meta-analysis in Pharmacoepidemiology," in *Pharmacoepidemiology*, ed. Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy, Fifth. (John Wiley & Sons, 2011), 723.

[98] Nissen and Wolski, "Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes."

[99] Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 128–131.

analyses of observational studies.[100] However, the FDA published findings particularly noting the difficulties in using meta-analyses of randomized controlled trials to study drug safety questions.[101] As part of performance goals in the Prescription Drug and User Fee Amendments of 2012 (PDUFA V), the FDA has agreed to develop guidance on the role of meta-analyses in regulatory decision-making.[102]

2.2 Mini-Sentinel System: Active Postmarket Risk Identification and Analysis

In general, Congress chose to focus its public health resources in the signal detection and refinement stages. Recall that the legislatively mandated addition of the Mini-Sentinel System to the FDA's armamentarium of postmarket safety systems is a direct response to the IOM's recommendation regarding strengthening the FDA's ability to test drug safety hypotheses. Essentially, the data structure of the Mini-Sentinel System both facilitates, and significantly scales, the FDA's existing capacity to perform pharmacoepidemiologic studies using administrative databases. Figure 2 shows a hypothesized placement of the Mini-Sentinel System and its potential capabilities within the safety signal adjudication framework. In comparison to Figure 1, the Mini-Sentinel System now enlarges the FDA's signal refinement capabilities and, in the future, may add to the FDA's signal detection capabilities. This next subsection describes the Mini-Sentinel System's data infrastructure and proposed capabilities that will supplement the FDA's existing postmarket safety systems.

---

[100] Su Golder, Yoon K Loke, and Martin Bland, "Meta-analyses of Adverse Effects Data Derived from Randomised Controlled Trials as Compared to Observational Studies: Methodological Overview," *PLoS Medicine* 8, no. 5 (May 2011): e1001026. See also Jan P Vandenbroucke and Bruce M Psaty, "Benefits and Risks of Drug Treatments: How to Combine the Best Evidence on Benefits with the Best Data About Adverse Effects," *JAMA: The Journal of the American Medical Association* 300, no. 20 (November 26, 2008): 2417–2419.

[101] Tarek A Hammad, Simone P Pinheiro, and George A Neyarapally, "Secondary Use of Randomized Controlled Trials to Evaluate Drug Safety: a Review of Methodological Considerations," *Clinical Trials (London, England)* 8, no. 5 (October 2011): 559–570.

[102] Food and Drug Administration and Center for Drug Evaluation and Research, "Prescription Drug User Fee Act (PDUFA) V: Reauthorization Performance Goals and Procedures; Fiscal Years 2013 Through 2017," WebContent, July 19, 2012, http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM270412.pdf.

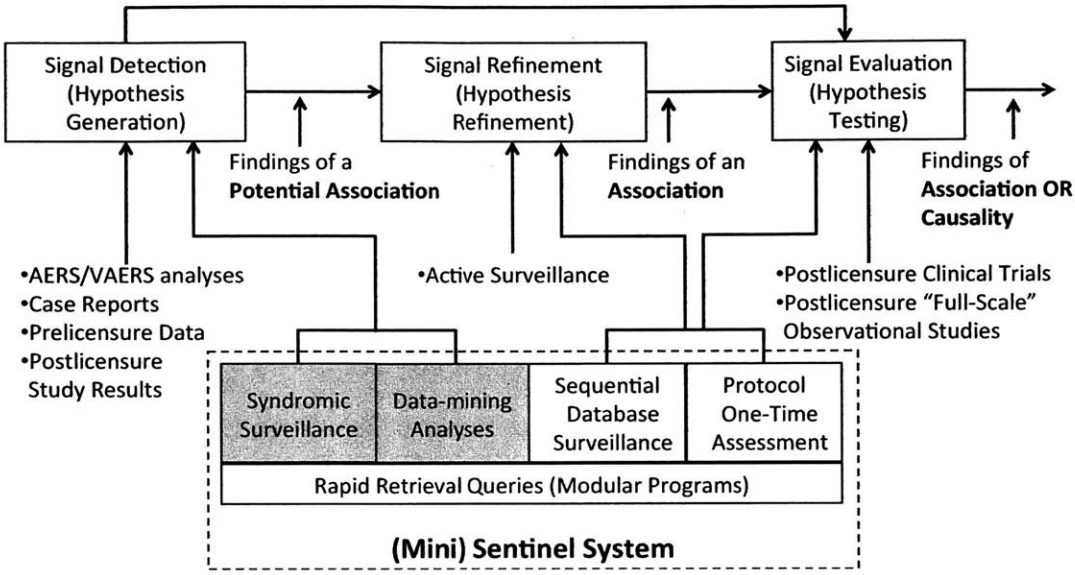

Figure 2. Hypothesized Identification and Adjudication of a Signal of Serious Risk in the U.S. Food and Drug Administration's Postmarket Surveillance Systems Including the Mini-Sentinel System
Adapted from Figure 1 based on material elsewhere.[103]
Abbreviations: AERS, Adverse Event Reporting System; VAERS, Vaccine Adverse Event Reporting System.

2.2.1 Mini-Sentinel System Data Infrastructure

At the outset, Congress's mandate - "to link and analyze safety data from multiple sources with the goals of including, in aggregate...at least 100,000,000 patients by July 1, 2012"[104] – necessitated the cooperation of multiple data partners with access to sensitive and legally protected health data. To facilitate their participation, the Mini-Sentinel System was envisioned as a distributed data environment as opposed to a centralized data repository.[105] The advantages of a distributed data environment are discussed elsewhere[106], but the main advantage is that it allows data partners to maintain control over their data and its uses. It also mitigates many legal, proprietary, privacy, and security concerns with respect to dealing with privately held, protected, and identifiable patient data.

---

[103] Platt et al., "The U.S. Food and Drug Administration's Mini-Sentinel Program: Status and Direction."
[104] *Food and Drug Administration Amendments Act of 2007, Public Law 110-85.*
[105] R. Platt et al., "The New Sentinel Network--improving the Evidence of Medical-product Safety," *The New England Journal of Medicine* 361, no. 7 (2009): 645–647.
[106] J. C. Maro et al., "Design of a National Distributed Health Data Network," *Annals of Internal Medicine* 151, no. 5 (2009): 341–344.

There were few existing distributed data models to rely on to guide development of the proposed infrastructure. The most prominent was the Centers for Disease Control and Prevention's Vaccine Safety Datalink (VSD), which performs vaccine safety research and surveillance, using administrative and claims data from participating health plans covering approximately 8.8 millions persons.[107] Other health data infrastructures that had utilized a distributed data environment included the HMO Research Network (HMORN)[108], the Meningococcal Vaccine Study[109], and the concurrently developed Post-licensure Immunization Safety Monitoring (PRISM) project[110]. All of these projects employ similar data architectures that are generically depicted in Figure 3 below.

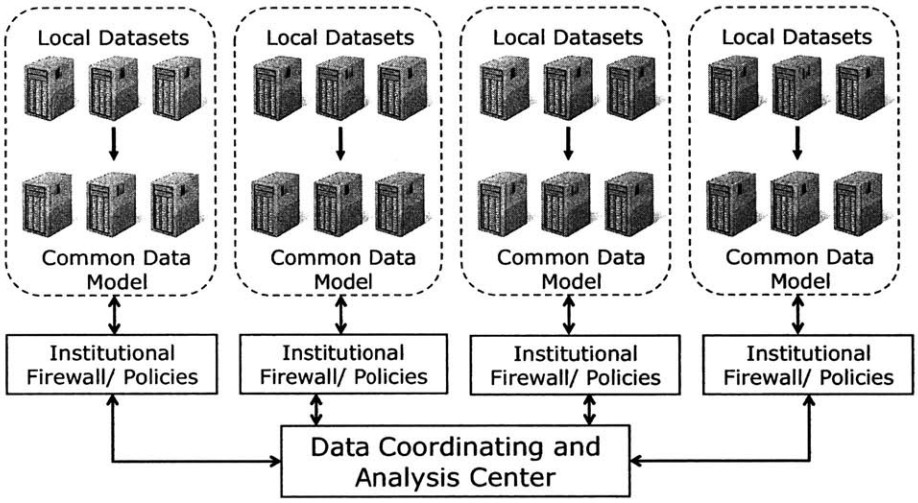

Figure 3. Generic Depiction of a Distributed Data Network Architecture

In a distributed data environment, participating data partners/sites periodically use an extract, transfer, and load procedure to generate copies of data and store them in a separate, firewalled location onsite. The frequency of dataset refreshes varies by data

[107] J. Baggs et al., "The Vaccine Safety Datalink: a Model for Monitoring Immunization Safety," *Pediatrics* 127 Suppl 1 (2011): S45–53.
[108] R. Platt et al., "Multicenter Epidemiologic and Health Services Research on Therapeutics in the HMO Research Network Center for Education and Research on Therapeutics," *Pharmacoepidemiology and Drug Safety* 10, no. 5 (2001): 373–377.
[109] P. Velentgas et al., "A Distributed Research Network Model for Post-marketing Safety Studies: The Meningococcal Vaccine Study," *Pharmacoepidemiology and Drug Safety* 17, no. 12 (2008): 1226–1234.
[110] M. Nguyen et al., "The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring Program: Strengthening the Federal Vaccine Safety Enterprise," *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 291–297.

partner, but is generally quarterly and may be annual.[111] These datasets have been transformed in order to adhere to a common data model, thereby ensuring identical file structures, data fields, and coding systems. Important elements of the common data model include demographic data, enrollment data, outpatient pharmacy dispensing data, and healthcare encounter information including diagnoses and medical procedures (e.g., results from inpatient, outpatient, and emergency department visits). The Mini-Sentinel System's common data model is described in detail elsewhere.[112] As of July 2011, it contains data from 17 partners covering over 300 million person-years of observation time, 2.4 billion unique encounters including 38 million acute inpatient stays, and 2.9 billion dispensings of prescriptions.[113] Further, the FDA reported that it met the 100 million persons requirement in the initiating legislation in December 2011.[114]

For the most part, the Mini-Sentinel System and the antecedent distributed data networks primarily rely on administrative claims data as opposed to clinical data obtained via electronic health records (EHRs). Others[115] have noted the relative strengths and weaknesses associated with each type of data and I review them here. Administrative claims data are advantageous because their primary purpose – billing for the utilization of health services – ensures that common coding conventions and interoperable systems are available for large sample sizes. Additionally, healthcare providers have strong incentives to submit claims for all care provided and health insurers record nearly all medical activities associated with their enrollees. Consequently, such records may provide more complete capture than data contained in stand-alone EHRs originating in individual physician practices or hospitals.

---

[111] Curtis et al., "Design Considerations, Architecture, and Use of the Mini-Sentinel Distributed Data System."
[112] Ibid.
[113] Platt et al., "The U.S. Food and Drug Administration's Mini-Sentinel Program: Status and Direction."
[114] Melissa Robb, "FDA's Mini-Sentinel Exceeds 100 Million Lives (and Counting)... A Major Milestone in Developing a Nationwide Rapid-response Electronic Medical Product Safety Surveillance Program," *FDA Voice*, June 29, 2012, http://blogs.fda.gov/fdavoice/index.php/2012/06/fdas-mini-sentinel-exceeds-100-million-lives-and-counting-a-major-milestone-in-developing-a-nationwide-rapid-response-electronic-medical-product-safety-surveillance-program/.
[115] Brian L. Strom, "Overview of Automated Databases in Pharmacoepidemiology," in *Pharmacoepidemiology*, ed. Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy, Fifth. (John Wiley & Sons, 2011), 158–162; John Seeger and Gregory W. Daniel, "Commercial Insurance Databases," in *Pharmacoepidemiology*, ed. Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy, Fifth. (John Wiley & Sons, 2011), 189–208; Schneeweiss and Avorn, "A Review of Uses of Health Care Utilization Databases for Epidemiologic Research on Therapeutics."

Some principal disadvantages of administrative claims data are the lack of clinical richness (e.g., clinical notes) that may be found in EHRs and the potential lack of generalizability since uninsured populations simply are not captured. Given the slow adoption of EHRs in the United States[116] (particularly ones that are interoperable) and the need for additional technologies to effectively capture data in EHRs (e.g., natural language processing), the routine inclusion of these data is still far off.[117] However, some participating data partners that act as integrated delivery systems – i.e., those that both insure patients and deliver their care - have supplemental data arising from laboratory results and EHRs. Additionally, claims data may suffer from data quality issues due to attempts to manipulate the payment system to secure more favorable reimbursement, known as "upcoding."[118]

Finally, the PRISM project within the Mini-Sentinel System linked data from Immunization Information Systems (e.g., registries) that cover 14 million persons in eight states.[119] This linkage was designed to capture immunizations received outside the medical home (e.g., in a retail setting). In the future, attempts to link disease-based registries, birth indices, death indices, and other sources of data are planned.[120]

2.2.2 Mini-Sentinel System Capabilities

As shown in Figure 2, there are five capabilities of the Mini-Sentinel System, some currently unrealized, which will be discussed in turn: 1) the development of population-wide descriptive statistics with regard to treatments, medical conditions, and outcomes via use of rapid retrieval queries; 2) the retrospective identification of *new* potential signals of serious risk via data-mining; 3) the prospective identification of *new* potential signals of serious risk via syndromic surveillance; 4) the retrospective assessment of pre-specified safety signals using a singular analysis, or "one-time protocol-based

---

[116] A. K. Jha et al., "Use of Electronic Health Records in U.S. Hospitals," *The New England Journal of Medicine* 360, no. 16 (2009): 1628–1638.

[117] However, new "meaningful use" regulations may hasten adoption. See Chun-Ju Hsiao et al., "Electronic Health Record Systems and Intent to Apply for Meaningful Use Incentives Among Office-based Physician Practices: United States, 2001-2011," *NCHS Data Brief*, no. 79 (November 2011): 1–8.

[118] Christopher S Brunt, "CPT Fee Differentials and Visit Upcoding Under Medicare Part B," *Health Economics* 20, no. 7 (July 2011): 831–841.

[119] D. A. Salmon et al., "Immunization-safety Monitoring Systems for the 2009 H1N1 Monovalent Influenza Vaccination Program," *Pediatrics* 127 Suppl 1 (2011): S78–86.

[120] Platt et al., "The U.S. Food and Drug Administration's Mini-Sentinel Program: Status and Direction."

assessment"; and 5) the prospective assessment of pre-specified safety signals using sequential database surveillance.

2.2.2.1 Rapid Retrieval Queries (Summary Tables and Modular Programs)

Rapid retrieval queries are standardized executable computer programs that are distributed to Mini-Sentinel System data partners via the Mini-Sentinel Operations Center (MSOC). These queries generate descriptive statistics regarding medical product usage, diagnoses, and procedures. The data partners may execute these programs onsite and send results back (via the MSOC) to the initiator for analysis, or may opt out of particular queries. Only authorized users with appropriate credentials may distribute queries.

The simplest of these queries run against "summary tables" that are annual prevalence counts of enrollment, diagnoses, procedures, and drug utilization.[121] The prevalence counts are stratified by year, sex, and age group. These queries are broad and describe simple phenomena, e.g., how many females had acute myocardial infarctions in the years 2005-2008.[122] They can be used for rapid feasibility checks[123] and as part of sample size calculations. They typically can be completed within a week of the initiating request from the FDA. The MSOC reported that over 50 summary table queries were performed in June and July 2011.[124] The MSOC has posted the results[125] of some of these summary tables including hip implant procedures, diagnoses of progressive multifocal leukoencephalopathy, and asthma.

More complex queries are referred to as "modular programs" and these queries operate in the same fashion but run against patient-level data as opposed to summary tables.[126] A key feature of modular programs is the parameterized nature of the code,

---

[121] Mini-Sentinel Operations Center, "Distributed Query Tool - Summary Tables", August 10, 2011, http://mini-sentinel.org/data_activities/details.aspx?ID=117.

[122] Mini-Sentinel Operations Center, "Assessments of Diagnoses and Medical Procedures | Acute Myocardial Infarctions", April 27, 2012, http://mini-sentinel.org/assessments/diagnoses_and_medical_procedures/details.aspx?ID=132.

[123] A feasibility check generally considers whether a database contains sufficient exposures or outcomes in appropriate populations to answer the study question.

[124] Curtis et al., "Design Considerations, Architecture, and Use of the Mini-Sentinel Distributed Data System."

[125] Mini-Sentinel Operations Center, "Assessments of Diagnoses and Medical Procedures", n.d., http://mini-sentinel.org/assessments/diagnoses_and_medical_procedures/default.aspx.

[126] For a brief summary, see Table 3, in Curtis et al., "Design Considerations, Architecture, and Use of the Mini-Sentinel Distributed Data System."

meaning it can be used repeatedly with modifications to the input information. They allow for more complicated questions, e.g., how many females prescribed new/incident oral anti-diabetic medications had acute myocardial infarctions? There are four established modular programs: 1) describes incident/prevalent medication use or procedures;[127] 2) describes incident/prevalent medication use or procedures among individuals with particular diagnoses;[128] 3) describes outcomes among incident users of medications with/without particular diagnoses;[129] 4) describes concomitant medication use among incident/prevalent users of another medication with/without particular diagnoses.[130] The current capacity for modular program requests is one per week as part of a service agreement with data partners and the FDA.[131] Modular program requests can typically be completed within a few weeks of the initiating request from the FDA.

Modular programs 1, 2, and 4 most closely mirror previous studies of medical product utilization. They can be useful for understanding prescribing patterns or "channeling" behaviors, or for observing usage patterns related to specific subgroups. Modular program 3 (MP3) is a more unique and powerful program for safety signal assessment because it describes crude (i.e., mostly unadjusted) *potentially* medical product-associated adverse event rates. The results of modular program 3 could be used to develop crude risk estimates. The MSOC has posted the results of two MP3 queries evaluating smoking cessation drugs and cardiovascular outcomes[132], and angiotensin-II receptor blockers and celiac disease[133]. Several new modular programs are actively being developed as well as enhancements to MP3.

---

[127] Mini-Sentinel Operations Center, "Module 1: Drug Use - General Characterization", August 17, 2011, 1, http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Modular-Program-1_v1.0.pdf.

[128] Mini-Sentinel Operations Center, "Module 2: Drug Use - By Medical Condition", August 17, 2011, 2, http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Modular-Program-2_v1.0.pdf.

[129] Mini-Sentinel Operations Center, "Module 3: Drug Use - Incident Outcomes", August 17, 2011, 3, http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Modular-Program-3_v1.0.pdf.

[130] Mini-Sentinel Operations Center, "Module 4: Drug Use - Concomitant Use", August 17, 2011, 4, http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Modular-Program-4_v1.0.pdf.

[131] Brown, J.S., Personal communication, March 5, 2012.

[132] Mini-Sentinel Operations Center, "Smoking Cessation Drugs & Cardiovascular Outcomes", January 17, 2012, http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Smoking-Cessation-Drugs-and-Selected-Cardiovascular-Outcomes.pdf.

[133] Mini-Sentinel Operations Center, "Angiotensin II Receptor Blockers & Celiac Disease", January 17, 2012, http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Angiotensin-II-Receptor-Blockers-and-Celiac-Disease.pdf.

2.2.2.2 Data-mining and Syndromic Surveillance

Data-mining and syndromic surveillance are both signal detection methods, which are performed without *a priori* specification of a hypothesized medical product-associated risk. Table 1 is a convenient way to classify these activities based on 1) their temporal perspective (i.e., when the data are sampled relative to when the sampling design is specified) and 2) the existence of a pre-specified hypothesis of medical product-associated risk, implying the existence of supporting data that suggest a potential risk.

		Temporal Perspective	
		Retrospective	Prospective
Exposure-Outcome Pair of Interest	Pre-Specified	**Protocol based one-time assessments**	**Sequential Database Surveillance**
	Non pre-specified	**Data-mining**	**Syndromic Surveillance**

Table 1. Types of Active Postmarket Risk Identification and Analysis[134]

Data-mining is a retrospective assessment of an existing database, and might likely employ disproportionality analyses similar to techniques currently used to analyze data in the spontaneous reporting system.[135] Syndromic surveillance is prospective monitoring for certain outcomes, irrespective of medication usage, that are present beyond some baseline level. In the past, the Centers for Disease Control and Prevention and other local public health agencies have monitored less serious outcomes, e.g., fever and influenza-like symptoms, to track infectious disease outbreaks.[136] Regulatory authorities might adapt this technique, by generating a list of known medical product-associated outcomes such as aplastic anemia or Stevens Johnson Syndrome, and create an alerting system triggered by the arrival of new data, i.e. a data refresh.

---

[134] Sebastian Schneeweiss and Jennifer Nelson, "Mini-Sentinel Methods Core: Accomplishments and Lessons Learned" (presented at the International Society of Pharmacoepidemiology (ISPE), Chicago, IL, August 17, 2011).

[135] Almenoff et al., "Novel Statistical Tools for Monitoring the Safety of Marketed Drugs."

[136] W Katherine Yih et al., "Evaluating Real-time Syndromic Surveillance Signals from Ambulatory Care Data in Four States," *Public Health Reports (Washington, D.C.: 1974)* 125, no. 1 (February 2010): 111–120; Jian Xing, Howard Burkom, and Jerome Tokars, "Method Selection and Adaptation for Distributed Monitoring of Infectious Diseases for Syndromic Surveillance," *Journal of Biomedical Informatics* 44, no. 6 (December 2011): 1093–1101.

Both data-mining and syndromic surveillance are in the very earliest stages of development in Mini-Sentinel System, and have not been piloted.[137] They are shown grayed in Figure 2 and Table 1 to reflect their embryonic development. However, once these capabilities are developed, they could serve as a technical solution to another of the IOM's recommendations: "systematically implement statistical-surveillance methods on a regular and routine basis for the automated generation of new safety signals."[138] For the purpose of analyzing the optimal use of the Mini-Sentinel System, this dissertation presumes safety signals have been detected and thus, will not further address these capabilities. However, when considering questions regarding the appropriate size of the future Sentinel System and the value of this system in fulfilling the FDA's responsibilities, the speed and thoroughness of safety signal detection beyond the existing passive spontaneous reporting systems merits more research.

2.2.2.3 Protocol-Based One-Time Assessments

Protocol-based one-time assessments refer to retrospective studies to assess particular exposure-outcome pairs using a singular end-of-study analysis. These assessments use customized protocols that adjust for confounding beyond age and sex using more advanced techniques[139]; use a variety of epidemiologic study designs[140]; and generate analytic results. These assessments may include medical chart adjudication of exposure,

---

[137] Davis, R. and Kulldorff, M., "Statistical Methods Development Details | Vaccine Safety Monitoring - Adverse Events", November 16, 2010, http://mini-sentinel.org/methods/methods_development/details.aspx?ID=1028.

[138] Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 7.

[139] J. A. Rassen and S. Schneeweiss, "Using High-dimensional Propensity Scores to Automate Confounding Control in a Distributed Medical Product Safety Surveillance System," *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 41–49; J. A. Rassen, J. Avorn, and S. Schneeweiss, "Multivariate-adjusted Pharmacoepidemiologic Analyses of Confidential Information Pooled from Multiple Health Care Utilization Databases," *Pharmacoepidemiology and Drug Safety* 19, no. 8 (2010): 848–857; R. J. Glynn, J. J. Gagne, and S. Schneeweiss, "Role of Disease Risk Scores in Comparative Effectiveness Research with Emerging Therapies," *Pharmacoepidemiology and Drug Safety* 21 Suppl 2 (2012): 138–147; J. A. Rassen et al., "Privacy-maintaining Propensity Score-based Pooling of Multiple Databases Applied to a Study of Biologics," *Medical Care* 48, no. 6 Suppl (2010): S83–9.

[140] J. J. Gagne et al., "Design Considerations in an Active Medical Product Safety Monitoring System," *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 32–40.

outcomes, or covariates.[141] In short, these assessments have much in common with traditional multi-site pharmacoepidemiologic studies using administrative databases. One difference is that the distributed database model may limit typical multi-variate regression-based analyses on patient-level data, which would require pooling of patient-level data across data partners.[142] However, development of distributed regression techniques, i.e., regression that can be performed on non-pooled data, is on the horizon[143], and some Mini-Sentinel System investigators are actively researching more sophisticated distributed approaches. In the interim, propensity scores and other summary statistics are used to allow for a greater inclusion of potential confounding variables.[144] Three protocol-based one-time assessments have been initiated as pilots: 1) venous thromboembolism following quadrivalent human papilloma virus vaccine;[145] and 2) angioedemia following administration of particular anti-hypertensives;[146] and 3) intussusception following rotavirus vaccine.[147]

2.2.2.4 Sequential Database Surveillance

Sequential database surveillance is another means of assessing evidence regarding a statistical association with respect to a pre-specified exposure-outcome pair. It differs from protocol-based one-time assessments in that it is performed prospectively with multiple, repeated assessments or hypothesis tests. Generally, to conduct a sequential database surveillance evaluation, one prospectively gathers data from multiple databases (e.g., population-based health data) to monitor the incidence rate of a medical product-

---

[141] See, for example, Michael D. Nguyen et al., "Monitoring for Venous Thromboembolism After Gardasil Vaccination," *Mini-Sentinel*, March 6, 2012, http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Gardasil-and-Venous-Thromboembolism-Protocol.pdf.

[142] See Darren Toh et al., "Protocol for Signal Refinement of Angioedema Events in Association with Use of Drugs That Act on the Renin-Angiotensin-Aldosterone System," *Mini-Sentinel*, July 18, 2011, http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Angioedema-and-RAAS_Protocol.pdf.... "Performing a centralized, conventional multivariable-adjusted analysis to obtain [Mini-Sentinel System]-wide estimates [of risk] may not be the preferred approach because it requires transferring of potentially identifiable individual-level information."

[143] Alan F. Karr et al., "Privacy-Preserving Analysis of Vertically Partitioned Data Using Secure Matrix Products," *Journal of Official Statistics* 25, no. 1 (2009): 125–138.

[144] M. A. Brookhart et al., "Confounding Control in Healthcare Database Research: Challenges and Potential Approaches," *Medical Care* 48, no. 6 Suppl (2010): S114–20.

[145] Nguyen et al., "Monitoring for Venous Thromboembolism After Gardasil Vaccination."

[146] Toh et al., "Protocol for Signal Refinement of Angioedema Events in Association with Use of Drugs That Act on the Renin-Angiotensin-Aldosterone System."

[147] Greene, Sharon K. Personal communication. June 14, 2012.

adverse event pair under surveillance. One then compares the observed incidence rate to an expected rate, which is calculated based on either a concurrent-, historical-, or self-controlled group. Comparisons are made at regular intervals as data accrue using sequential statistical tests with pre-specified signaling thresholds. If the test statistic exceeds the threshold, then a statistical signal of excess risk is identified, the hypothesized exposure-outcome association is strengthened, and the null hypothesis of no excess risk is rejected. This signal is ordinarily followed by confirmatory assessments and review to validate or refute the finding.

Sequential database surveillance is not a new technique in medical product surveillance although it has largely been performed to study vaccines[148] that are administered to healthy people and where there is less potential confounding. One sequential database surveillance protocol is being tested in the Mini-Sentinel System: the risk of acute myocardial infarction associated with oral anti-diabetic agents.[149] Another protocol is in development for sequential analyses related to influenza vaccine safety.[150] The literature on prospective sequential database surveillance is reviewed herein in Section 5.

This dissertation is focused on developing qualitative and quantitative tools to aid regulators in utilizing the sequential database surveillance capabilities in the Mini-Sentinel System. However, prior to engaging in the specifics of using the Mini-Sentinel System for these purposes, it is necessary to first examine how the FDA uses postmarket evidence to support regulatory decision-making. I turn to that topic next.

---

[148] Yih et al., "Active Surveillance for Adverse Events: The Experience of the Vaccine Safety Datalink Project."

[149] B. Fireman et al., "A Protocol for Active Surveillance of Acute Myocardial Infarction in Association with the Use of a New Antidiabetic Pharmaceutical Agent," *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 282–290.

[150] Greene, Sharon K. Personal communication, June 14, 2012.

3 FDA REGULATORY DECISION-MAKING

This section connects the U.S. Food and Drug Administration (FDA)'s evidence-generation activities[151] via its postmarket systems (as generally discussed in the previous section) to regulatory decision-making while embedding both of these activities in the new legal framework enacted in the Food and Drug Administration Amendments Act. As will be explained next, the emergence of new safety information[152] in the postmarket compels the FDA to re-evaluate a medical product's benefit-risk profile and make a regulatory decision. Such a decision may involve a) taking regulatory actions to influence medical product utilization or a deliberate decision to forego such actions, or b) taking regulatory actions to generate new knowledge on a product's benefits and risks. It can also pursue these two courses simultaneously (and often does) when a safety signal cannot be fully resolved based on the existing evidence, but interim regulatory actions are taken while awaiting additional analyses.

It is important to distinguish the pursuit of evidence generation for the sake of science and the pursuit of evidence generation for the sake of regulatory decision-making. With regard to the former, the Institute of Medicine has said that:

> "The science of drug safety concerns questions of causal, not just statistical, relationships. That is, the important drug-safety question is whether drug exposure actually *causes* an adverse outcome, not simply whether such an outcome occurs more frequently in people who choose to take the drug."[153]

Yet, safety signals of pressing public health importance may necessitate regulatory action, and time and ethical constraints may rule out research more conducive to strict causal inference.[154] This mismatch between the need for action and the lack of ideal evidence to support it has led many public health policymaking entities - e.g., the Environmental Protection Agency, the Agency for Healthcare Research and Quality - to

---

[151] There are additional evidence generation activities that have not been explicitly discussed herein such as clinical pharmacology activities, pharmacogenomic activities, and other animal and laboratory studies. These studies also support regulatory decision-making, but are beyond the scope of this dissertation.

[152] "New safety information" is a regulatory term which is explained *infra* at note 216.

[153] Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 115.

[154] Causal inference is discussed in greater detail herein in subsection 3.4.

establish frameworks[155] designed to systematically grade the strength of a body of evidence and make determinations about causal inference.[156] The FDA, as of yet, has not systematically adopted one of these frameworks for postmarket evidence.[157] Interestingly, these frameworks are used to assess evidence that has already been developed and draw conclusions regarding its strength. Congress has directed the FDA to proceed in the reverse order: to set a goal or standard of sufficient evidence, and then assess the ability of evidence-generating systems to attain that goal.

---

[155] For example, see U.S. EPA, *Guidelines for Carcinogen Risk Assessment* (Washington, D. C.: U.S. Environmental Protection Agency, 2005), http://www.epa.gov/cancerguidelines/; D. K. Owens et al., *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*, AHRQ Methods for Effective Health Care (Rockville (MD): Agency for Healthcare Research and Quality (US), 2008), http://www.ncbi.nlm.nih.gov/pubmed/21433399.

[156] Complete coverage of this literature is beyond the scope of this dissertation. What is important is that the reader understands the potential outcomes that result after this grading exercise. For example, the Institute of Medicine has proposed "the following categorization of the strength of the *overall evidence* for or against a *causal relationship* from exposure to disease:

1. *Sufficient*: The evidence is sufficient to conclude that a causal relationship exists.
2. *Equipoise and Above*: The evidence is sufficient to conclude that a causal relationship is at least as likely as not, but not sufficient to conclude that a causal relationship exists.
3. *Below Equipoise*: The evidence is not sufficient to conclude that a causal relationship is at least as likely as not, or is not sufficient to make a scientifically informed judgment.
4. *Against*: The evidence suggests the lack of a causal relationship."

For more, see Institute of Medicine (IOM), *Improving the Presumptive Disability Decision-Making Process for Veterans*, ed. Catherine C. Bodurow and Jonathan M. Samet (The National Academies Press, 2008), 189, http://www.nap.edu/openbook.php?record_id=11908; Institute of Medicine (IOM), *Adverse Effects of Vaccines: Evidence and Causality* (Washington, D.C.: The National Academies Press, 2012); Neal A Halsey et al., "Algorithm to Assess Causality After Individual Adverse Events Following Immunizations," *Vaccine* 30, no. 39 (August 24, 2012): 5791–5798.

[157] For initial (and continued) licensure of products, the FDA does set the minimum standard to be "substantial evidence." Substantial evidence is defined as "evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof. If the Secretary determines, based on relevant science, that data from one adequate and well-controlled clinical investigation and confirmatory evidence (obtained prior to or after such investigation) are sufficient to establish effectiveness, the Secretary may consider such data and evidence to constitute substantial evidence for purposes of the preceding sentence." Codified at 21 U.S.C. § 355(d). Additionally, with regard to therapies for diabetes mellitus and cardiovascular events, the FDA is quite specific: "If the premarketing application contains clinical data that show that the upper bound of the two-sided 95 percent confidence interval for the estimated increased risk (i.e., risk ratio) is between 1.3 and 1.8, and the overall risk-benefit analysis supports approval, a postmarketing trial generally will be necessary to definitively show that the upper bound of the two-sided 95 percent confidence interval for the estimated risk ratio is less than 1.3." See U.S. Department of Health and Human Services, Food and Drug Administration, and Center for Drug Evaluation and Research, "Guidance for Industry: Diabetes Mellitus: Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes (Final)", December 17, 2008, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071627.pdf.

Recall that per the Food and Drug Administration Amendments Act, when the FDA is confronted with new safety information in the postmarket, it must make a determination on whether its publicly-funded postmarket safety systems (including the Mini-Sentinel System) will generate **sufficient** knowledge to make a regulatory decision, or whether additional privately-funded (i.e. manufacturer-funded) studies are necessary.[158] As of this writing, the FDA has not yet stated the basis on which determinations of sufficiency will be made. Presumably, the Mini-Sentinel System's "laboratory" status has allowed the FDA to defer construction of a sufficiency standard on the grounds that the capabilities of the Mini-Sentinel System as an evidence-generating system are unknown at this time. Still, the legislation directs the FDA to anticipate the strength of evidence needed to support particular regulatory actions (or deliberate choices not to take such actions), and then to choose appropriate systems to generate that evidence. The Mini-Sentinel System is just one of such systems.

In this section, I draw the reader's attention to the statutes, regulations, and guidance that inform and govern the FDA's options for regulatory action, and specifically, what these texts require with respect to findings of causality and the strength of the evidence. As the reader will discover, with the exception of one particular regulatory action, the law is unclear on this issue, leaving much to the judgment of the FDA.[159] In general, as will be discussed more thoroughly in subsection 3.4, the ability to infer causal relationships from observational data - e.g., the data available in the Mini-Sentinel System – is challenging at best.[160] However, if the FDA determines that robust findings on causality (either affirmation or rejection) are necessary precedents to particular regulatory actions, then the FDA's postmarket systems (including the Mini-Sentinel System) must be

---

[158] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355(o)(3)(B).

[159] This ambiguity leads to scientific and policy disagreements within the FDA. See, for example, a summary of discordant views with respect to the rosiglitazone-cardiovascular outcomes safety signal in Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 131–133. Incidentally, the Agency for Healthcare Research and Quality found similar disagreements among their experts even when using a common framework. See Nancy D N D Berkman et al., *Reliability Testing of the AHRQ EPC Approach to Grading the Strength of Evidence in Comparative Effectiveness Reviews*, AHRQ Methods for Effective Health Care (Rockville (MD): Agency for Healthcare Research and Quality (US), 2012), http://www.ncbi.nlm.nih.gov/pubmed/22764383.

[160] See generally Miguel Hernán and Jamie Robins, *Causal Inference*, v1.10.17 ed. (Chapman & Hall/CRC, 2012), 25–40, http://www.hsph.harvard.edu/faculty/miguel-hernan/causal-inference-book/; Institute of Medicine (IOM), *Improving the Presumptive Disability Decision-Making Process for Veterans*, Chapter 7 and Appendix J.

evaluated with respect to their ability to generate evidence to support such findings. In other words, the legal standards to support particular regulatory decision-making factor into whether the Mini-Sentinel System will be deemed a **sufficient** evidence-generation system.

3.1 Pre-requisites to Sufficiency: New Safety Information and Purpose

Per the Food and Drug Administration Amendments Act, two necessary precursors to the FDA's sufficiency decision are the establishment of new safety information[161] and the designation of a purpose (i.e., a goal) for evidence generation:

> "(i) To assess a known serious risk related to the use of the drug involved.
> (ii) To assess signals of serious risk related to the use of the drug.
> (iii) To identify an unexpected serious risk when available data indicates the potential for a serious risk."[162]

The determination of sufficiency will logically be different for the three aforementioned purposes in the legislation. These three purposes each imply a differing "maturity" to the existing evidence regarding the pre-specified exposure-outcome pair.

The first purpose, assessing known serious risks, implies a causal relationship is known and the risk is likely included in the FDA-approved product labeling. Examples of known serious risks that might still require regulatory action include an increased incidence of supratherapeutic effects, e.g., insulins-hypoglycemia, that change the benefit-risk profile of the product.

The second purpose – to assess signals of serious risk - implies that a signal has been detected, but has not yet been adequately quantified and a determination of causality, if attempted, has not been conclusive. Examples of potential signals of serious risk are

---

[161] New safety information is defined as "information derived from a clinical trial, an adverse event report, a postapproval study (including a study under section 505(o)(3)), or peer-reviewed biomedical literature; data derived from the postmarket risk identification and analysis system under section 505(k); or other scientific data deemed appropriate by the Secretary about— (A) a serious risk or an unexpected serious risk associated with use of the drug that the Secretary has become aware of (that may be based on a new analysis of existing information) since the drug was approved, since the risk evaluation and mitigation strategy was required, or since the last assessment of the approved risk evaluation and mitigation strategy for the drug; or (B) the effectiveness of the approved risk evaluation and mitigation strategy for the drug obtained since the last assessment of such strategy." Codified at 21 U.S.C. § 355-1(b)(3).

[162] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355(o)(3)(B).

published quarterly by the FDA on its website.[163] These signals may or may not be included in the product's label, but often result in regulatory action in the form of safety-related label changes.[164]

The third purpose – identifying an unexpected serious risk[165] when available data indicates the potential for a serious risk – implies that some pre-specification of an exposure-outcome pair has occurred, that this outcome is not (fully) indicated in the labeling[166], and that signal detection (i.e., identification) is ongoing. A good example is referenced herein[167] regarding a fingolimod-bradycardia signal, which was more severe than first anticipated. Other examples might be medical products presumed to cause teratogenic effects that were not directly observed (e.g., perhaps based on evidence from animal studies), or medical products that are part of a class when class-wide risks have been detected.

Typically, these three purposes are associated with traditional safety issues, i.e., the emergence of adverse events. Also, a manufacturing quality issue in which the active pharmaceutical ingredient is omitted or undersupplied might be generally regarded as a safety issue. However, Evans[168] goes further and concludes that the FDA may require a postmarket requirement in accordance with the three purposes above based on data suggesting a lack of benefit (i.e., what she refers to as "non-response" or "efficacy failure"[169]). She grounds her argument in the legislation's definition of adverse drug

---

[163] Center for Drug Evaluation and Research, "Adverse Events Reporting System (AERS) - Potential Signals of Serious Risks/New Safety Information Identified from the Adverse Event Reporting System (AERS)," WebContent, n.d., http://www.fda.gov/drugs/guidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082196.htm.

[164] Abbey Powers and G Elliott Cook, "Potential Safety Signals and Their Significance," *Archives of Internal Medicine* 172, no. 1 (January 9, 2012): 72–73.

[165] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*. "The term 'unexpected serious risk' means a serious adverse drug experience that is not listed in the labeling of a drug, or that may be symptomatically and pathophysiologically related to an adverse drug experience identified in the labeling, but differs from such adverse drug experience because of greater severity, specificity, or prevalence." Codified at 21 U.S.C. § 355-1(b)(8).

[166] The legislative language on this point is difficult to parse since it allows for the outcome to be unlabeled or perhaps "under-labeled" in a sense that the severity, specificity or prevalence is not well-defined.

[167] See *supra* at note 43. This example might also legitimately be an example of the first purpose: assessing a known serious risk.

[168] B. J. Evans, "Seven Pillars of a New Evidentiary Paradigm: The Food, Drug, and Cosmetic Act Enters the Genomic Era," *Notre Dame Law Review* 85, no. 2 (2010): 480,498–500.

[169] To be clear, non-response or efficacy failure referenced in Evans is not associated with manufacturing problems. Rather, efficacy failure varies by the individual and is likely attributable to unknown scientific

experience[170], which includes, in part, "any failure of expected pharmacological action of the drug." She argues that a lack of benefit would meet the requirements of one of the three purposes above if the lack of benefit resulted in serious harm[171]. The IOM has endorsed this interpretation[172] and is particularly concerned with therapeutics approved based on surrogate endpoints in this regard.

Herein, I do not explicitly consider the use of the FDA's postmarket safety systems (including the Mini-Sentinel System) as an evidence generation system to support regulatory decision-making when the precipitating issue is a lack of benefit. Others have suggested that systems like the Mini-Sentinel System – large-scale administrative database networks – could support such evidence generation.[173]

3.2 FDA Guidance: Tracked Safety Issues

Given new safety information, the FDA can generate a tracked safety issue, which is a management tool that is a convenient operational unit of analysis for this dissertation.[174] The FDA has stated that a tracked safety issue will be established if the identified safety issue has the potential to lead to any of the following actions:

- "Withdrawal of FDA approval of a drug
- Withdrawal of an approved indication
- Limitations on a use in a specific population or subpopulation

---

factors involving an individual's genetic makeup. The serious harm that may result *is not* the harm caused by the drug itself, but the harm caused because the patient has foregone other therapeutic options. Evans refers to this as a "lost-chance" injury, based on "lost-chance" doctrine, which, in some states, allows lawsuits from patients who have suffered irreversible disease progression as a result of a delay in treating or diagnosing their disease. Evans, "Seven Pillars of a New Evidentiary Paradigm," 499.

[170] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*. "The term 'adverse drug experience' means any adverse event associated with the use of a drug in humans, whether or not considered drug related, including—(A) an adverse event occurring in the course of the use of the drug in professional practice; (B) an adverse event occurring from an overdose of the drug, whether accidental or intentional;(C) an adverse event occurring from abuse of the drug; (D) an adverse event occurring from withdrawal of the drug; and (E) any failure of expected pharmacological action of the drug." Codified at 21 U.S.C. § 355-1(b)(1).

[171] See *supra* at note 6.

[172] Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 31. "The present committee considers providing FDA with that authority [to investigate any failure of the expected pharmacological action of the drug] to be in the interest of the public's health. When questions arise about the health benefits of a drug, studies to document a drug's effectiveness may be as critical for ensuring that the benefit–risk profile of a drug remains favorable as studies that investigate its risks."

[173] S. Toh et al., "Comparative-effectiveness Research in Distributed Health Data Networks," *Clinical Pharmacology and Therapeutics* 90, no. 6 (2011): 883–887.

[174] A tracked safety issue implies that some degree of signal detection is complete, which is important in this dissertation since I am primarily concerned with FDA decision-making after signal detection has suggested an exposure-outcome pair to monitor.

- Additions or modifications to the Warnings and Precautions, or Contraindications sections of the labeling, or the Medication Guide or other required Patient Package Insert, including safety labeling changes required under the Food and Drug Administration Amendments Act (FDAAA)
- Establishment of or changes to the proprietary name/container label/labeling/packaging to reduce the likelihood of medication errors
- Establishment or modification of a risk evaluation and mitigation strategy (REMS)
- A requirement that a sponsor conduct a safety-related postmarketing trial or study
- The conduct of a safety-related observational epidemiological study by FDA."[175]

A tracked safety issue relates to a particular exposure-outcome pair and is classified as emergency, priority, or standard based on a number of characteristics that define the safety issue.[176] Once classified, the FDA examines the existing evidence, re-evaluates the risk-benefit profile for the drug, and determines whether regulatory action is necessary. Figure 4 illustrates the series of decisions the FDA must make once a tracked safety issue is evaluated.

The FDA's first decision (marked in Figure 4 as Decision 1) concerns initial regulatory actions. To be clear, a regulatory action can encompass a) actions intended to inform/influence patient and prescriber behavior (e.g., labeling changes, restriction of access to particular therapies), b) deliberate choices to maintain the status quo, which is routine monitoring, or c) actions that generate additional evidence (e.g., issuance of a postmarket requirement). Additionally, the FDA can pursue the first and third courses simultaneously. Should Decision 1 involve an evidence-generating regulatory action and there is a desire to pursue privately-funded postmarket requirements, the FDA is required to make a sufficiency finding as discussed previously (shown in Figure 4 as Decision 2).

[175] U.S. Department of Health and Human Services, Food and Drug Administration, and Center for Drug Evaluation and Research, "Guidance: Classifying Significant Postmarketing Drug Safety Issues (Draft)", March 8, 2012, 3, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM295211.pdf.

[176] U.S. Department of Health and Human Services, Food and Drug Administration, and Center for Drug Evaluation and Research, "Guidance: Classifying Significant Postmarketing Drug Safety Issues (Draft)."

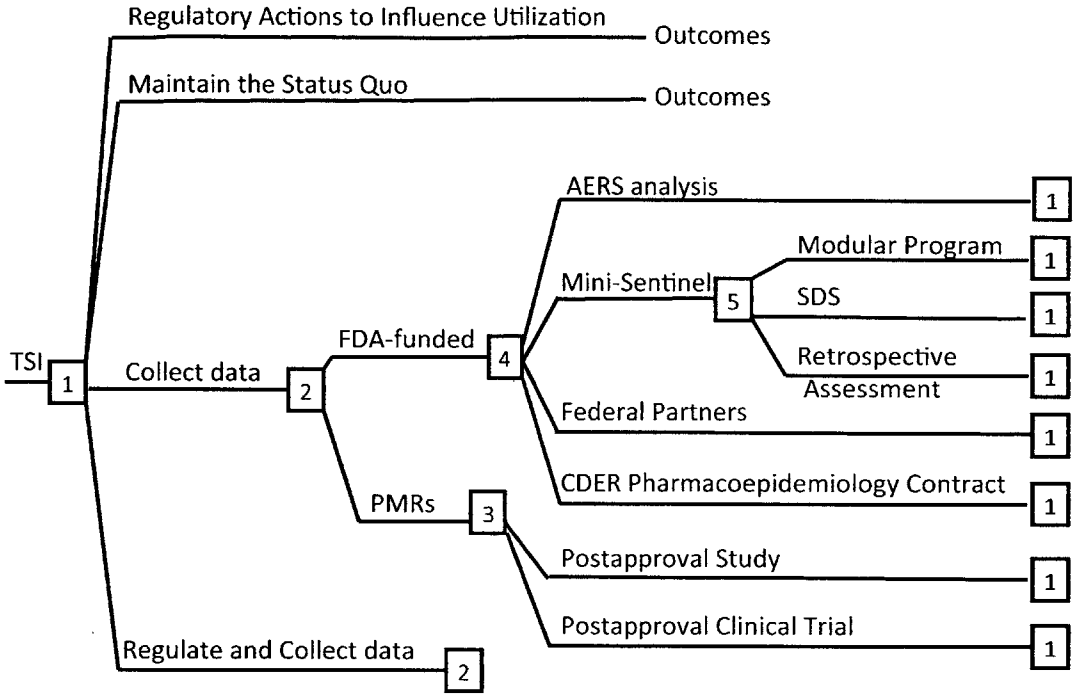

Figure 4. FDA Decision Process Following Identification and Evaluation of a Tracked Safety Issue
Abbreviations: TSI, tracked safety issue; FDA, Food and Drug Administration; PMR, postmarket requirement; CDER, Center for Drug Evaluation and Research; AERS, Adverse Event Reporting System; SDS, sequential database surveillance.

3.3 Plausible Regulatory Actions

Of the potential regulatory actions described in the tracked safety issue guidance, I do not consider issues related to medication error management. For the remaining actions, I review the legal and regulatory standards for evidence to support such actions. To be clear, I review these standards because a statutorily required sufficiency determination on the evidence generation capabilities of the FDA's postmarket systems *depends* on consideration of the strength of evidence necessary to support particular regulatory actions. For example, if regulators determine that a strong finding (either affirmation or rejection) of causality is needed to support a particular regulatory action, then the FDA's postmarket systems (including the Mini-Sentinel System) must be evaluated with respect to their likelihood of generating that finding.

3.3.1 Safety-Related Labeling Changes

The most common regulatory action performed by the FDA is to make changes to the product's FDA-approved labeling. This labeling is the FDA's principal means of communicating benefits and risks associated with a medical product. In general, changes to a product's label are normal and expected. These changes reflect the evolving nature of the benefit-risk profile of a medication as it is used in larger and more diverse populations. However, because a label is intended to influence a medication's adoption and use, historically, its contents have been a source of conflict between the FDA and manufacturers.[177] In 2007, Congress eliminated many potential conflicts by expanding the FDA's authority to mandate safety-related labeling changes given emergent safety information.[178] The FDA has issued guidance with respect to how it will use this new authority[179], but the evidence bar for certain safety-related label changes was already established in FDA regulations.

Specifically, the *Warnings and Precautions* section of the label is intended to document serious, or otherwise clinically significant, adverse drug reactions[180], and:

> "the labeling must be revised to include a warning about a clinically significant hazard as soon as there is reasonable evidence of a causal association with a drug; a causal relationship need not have been definitely established."[181]

The reader should note that this regulatory action is the only regulatory action with *explicit statements with regard to causality.* The FDA's guidance on this section also directs the manufacturer to include "anticipated" adverse drug reactions that have not yet been observed but might be reasonably suspected to occur based on a) known pharmacologic effects, chemical effects, or class-based effects; or b) animal studies (e.g.,

---

[177] See, for example, a summary of the FDA's negotiations with Merck regarding the Vioxx® (rofecoxib) label in David A. Kessler and David C. Vladeck, "A Critical Examination of the FDA's Efforts To Preempt Failure-To-Warn Claims," *Georgetown Law Journal* 96 (2008): n82.

[178] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355(o)(4).

[179] Department of Health and Human Services et al., "Guidance for Industry: Safety Labeling Changes - Implementation of Section 505(o)(4) of the Federal Food, Drug, and Cosmetic Act (Draft)", April 12, 2011, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM250783.pdf.

[180] U.S. Department of Health and Human Services et al., "Guidance for Industry: Warnings and Precautions, Contraindications, and Boxed Warning Sections of Labeling for Human Prescription Drug and Biological Products - Content and Format (Final)", October 11, 2011, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM075096.pdf.

[181] 21 CFR § 201.57(c)(6) and 21 CFR § 201.80(e), April 1, 2012 edition.

studies that might indicate teratogenicity).[182] The "anticipated" language bears a similarity to the definition of "unexpected serious risk" in the FDAAA. Both imply that prior knowledge makes the outcome likely even if it is not yet observed. Also, the FDA directs the manufacturer to include adverse drug reactions that may be associated with unapproved (i.e., off-label) uses.[183]

In contrast, the *Contraindications* section of the label requires that "known hazards and not theoretical possibilities must be listed" and is intended to describe situations when "the drug should not be used because the risk of use (e.g., certain potentially fatal adverse reactions) clearly outweighs any possible therapeutic benefit."[184] The FDA guidance contains similar language regarding "anticipated adverse drug reactions" in the *Contraindications* section.[185] To distinguish the *Warnings and Precautions* section from the *Contraindications* section, information in the former is intended to qualify/moderate use of a medical product (e.g., certain subpopulations may be at higher risk for certain adverse events) whereas information in the latter is intended to eliminate particular uses (e.g., certain subpopulations should never use this product). The *Boxed Warnings* section of the label is a subset of information already included in the either the *Warnings and Precautions* or the *Contraindications* section.

For completeness, the *Adverse Reactions* section of the label is required to list the adverse reactions that occur with the drug and with drugs in the same pharmacologically active and chemically related class.[186] This list is limited to "adverse events for which there is some basis to believe there is a causal relationship between the drug and the occurrence of the adverse event."[187] Despite FDA's guidance to the contrary[188], this

---

[182] U.S. Department of Health and Human Services et al., "Guidance for Industry: Warnings and Precautions, Contraindications, and Boxed Warning Sections of Labeling for Human Prescription Drug and Biological Products - Content and Format (Final)."
[183] Ibid.
[184] 21 CFR § 201.57(c)(5) and 21 CFR § 201.80(d), April 1, 2012 edition.
[185] U.S. Department of Health and Human Services et al., "Guidance for Industry: Warnings and Precautions, Contraindications, and Boxed Warning Sections of Labeling for Human Prescription Drug and Biological Products - Content and Format (Final)."
[186] 21 CFR § 201.57(c)(7)(i), April 1, 2012 edition.
[187] 21 CFR §201.57(c)(7), April 1, 2012 edition.
[188] U.S. Department of Health and Human Services et al., "Guidance for Industry: Adverse Reactions Section of Labeling for Human Prescription Drug and Biological Products-Content and Format (Final)", January 18, 2006, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM075057.pdf.

section is more a laundry list of documented adverse reactions. The FDA also indicates in guidance that it does not believe that changes to the *Adverse Reactions* section *alone* would normally trigger a "safety-related label change" action per the 2007 legislation.[189]

Powers et al. documented the percentage of potential signals of serious risk identified by the Adverse Event Reporting System (AERS) that resulted in safety-related labeling changes, and a majority of these changes were in the *Warnings and Precautions* section.[190] In her study of AERS-identified signals, only two resulted in changes to the *Contraindications* section: Exjade® (deferasirox) and Saphris® (asenapine). It is unclear whether evidence beyond the AERS signal was required to generate changes to the Contraindications section. Others[191] have taken a different approach by examining safety-related labeling changes and working backward to identify the evidence used to initiate them. Both confirm changes to the *Warnings and Precautions* section to be a more prevalent regulatory action than to the *Contraindications* section.

In general, while the medication's label is a comprehensive summary of a medication's risks and benefits, it includes technical language that may seem confusing to patients. "Patient labeling" – in the form of Medication Guides and Patient Package Inserts – are benefit-risk communications written explicitly for patients (as opposed to healthcare providers or pharmacists). Medication Guides are required in one or more of the following circumstances:

> "(1) The drug product is one for which patient labeling could help prevent serious adverse effects. (2) The drug product is one that has serious risk(s) (relative to benefits) of which patients should be made aware because information concerning the risk(s) could affect patients' decision to use, or to continue to use, the product. (3) The drug product is important to health and patient adherence to directions for use is crucial to the drug's effectiveness."[192]

Patient Package Inserts are required for all oral contraceptives[193] and estrogen-containing

---

[189] Department of Health and Human Services et al., "Guidance for Industry: Safety Labeling Changes - Implementation of Section 505(o)(4) of the Federal Food, Drug, and Cosmetic Act (Draft)," 4. "FDA expects that information that results in changes made only to the ADVERSE REACTIONS section, but does not warrant inclusion in other sections of labeling (such as WARNINGS AND PRECAUTIONS), would not normally trigger required safety labeling changes under section 505(o)(4)."

[190] Powers and Cook, "Potential Safety Signals and Their Significance."

[191] Moore, Singh, and Furberg, "The FDA and New Safety Warnings"; Jean Lester et al., "Evaluation of FDA Safety-Related Drug Label Changes in 2010", unpublished manuscript, May 5, 2012.

[192] 21 CFR § 208.1(c), April 1, 2012 edition.

[193] 21 CFR § 310.501, April 1, 2012 edition.

therapeutics.[194] Patient Package Inserts can also be voluntary components of labeling. Both Medication Guides and mandatory Patient Package Inserts have distribution requirements.[195] The FDA maintains a list of active products that require a Medication Guide.[196] Whether Medication Guides can be required on the basis of causal associations (i.e., as *Warning and Precautions* changes are) as opposed to known risks (i.e., as *Contraindications* changes are) seems an open question, but it appears that either could form the evidence base for a Medication Guide.

The FDA has begun tracking its safety-related label changes (including Medication Guides/Patient Package Inserts) in a monthly database, and these changes are delineated by section.[197] More significant interrogation of this dataset could be used to more closely link evidentiary standards with resultant safety-related label changes.

3.3.2 Changes to Risk Evaluation and Mitigation Strategies

Another type of regulatory action available to the FDA is the issuance or revision of a risk evaluation and mitigation strategy (REMS). A REMS may consist of up to five related elements that contain individual requirements. Those elements are: a Medication Guide/Patient Package Insert; a communication plan; elements to assure safe use (ETASU) provisions; an implementation system; and a timetable for assessments.[198] I focus on the first three elements because an implementation system is not independent of ETASU provisions. REMS assessments are designed to evaluate the degree to which the strategy is meeting its goals.[199]

The evidentiary standard for issuance of REMS in the postapproval period is broad:

---

[194] 21 CFR § 310.515, April 1, 2012 edition.
[195] 21 CFR § 208.24, April 1, 2012 edition.
[196] Center for Drug Evaluation and Research, "Medication Guides," WebContent, n.d., http://www.fda.gov/Drugs/DrugSafety/ucm085729.htm.
[197] Center for Drug Evaluation and Research, "Drug Safety Labeling Changes," WebContent, n.d., http://www.fda.gov/Safety/MedWatch/SafetyInformation/Safety-RelatedDrugLabelingChanges/default.htm.
[198] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355-1(e).
[199] Department of Health and Human Services et al., "Guidance for Industry: Format and Content of Proposed Risk Evaluation and Mitigation Strategies (REMS), REMS Assessments, and REMS Modifications (Draft)", October 1, 2009, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM184128.pdf.

> "if the Secretary becomes aware of new safety information and makes a determination that such a [risk evaluation and mitigation] strategy is necessary to ensure that the benefits of the drug outweigh the risks of the drug."[200]

In general, the statute provides flexibility to the FDA, but makes it difficult to interpret what type of data or analyses would be required to make a showing that the REMS (and its subsequent assessments) assures that a drug's benefit outweighs its risk.

First, Medication Guide/Patient Package Inserts are described above. While Medication Guides can be entirely managed via the existing regulatory structure, a Medication Guide mandated as a part of a REMS ensures that an assessment of the strategy will be performed. The FDA has stated in guidance that it does not intend to issue Medication Guide-only REMS unless a "Medication Guide without a REMS will not be sufficient to ensure that the benefits of the drug outweigh the risks."[201] Further, it has stated its intention to include a Medication Guide as part of a REMS when the REMS already includes an ETASU provision.[202] As of March 2012, 125 REMS were issued as Medication Guide-only REMS, and 105 of these were later released (i.e., the REMS requirement was fulfilled/canceled).[203] To be clear, these Medication Guides still exist; they are just no longer governed by the REMS regulatory structure, which requires mandatory assessments.

Second, while Medication Guides mandate particular manufacturer-patient communications, a communication plan refers to communications between manufacturers and healthcare providers or pharmacists. Historically, these communications have been "Dear Healthcare Provider" letters sent for one or more of the following three purposes: 1) new medical product information related to a significant hazard to health; 2) new medical product information related to important changes to the product's labeling; and

---

[200] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355-1(a)(2).

[201] U.S. Department of Health and Human Services et al., "Guidance: Medication Guides - Distribution Requirements and Inclusion in Risk Evaluation and Mitigation Strategies (REMS) (Final)", November 17, 2011, 8, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM244570.pdf.

[202] U.S. Department of Health and Human Services et al., "Guidance: Medication Guides - Distribution Requirements and Inclusion in Risk Evaluation and Mitigation Strategies (REMS) (Final)."

[203] Center for Drug Evaluation and Research, "News & Events - Slides for the June 7, 2012 Risk Evaluation and Mitigation Strategy (REMS) Assessments Public Workshop," WebContent, n.d., http://www.fda.gov/Drugs/NewsEvents/ucm307675.htm.

3) correction of advertising or labeling information related to the medical product.[204] However, the FDAAA also references communications related to explaining REMS provisions[205] (i.e., specifically ETASU and implementation plans), and "disseminating information to health care providers through professional societies about any serious risks of the drug and any protocol to assure safe use."[206] As of March 2012, 40 REMS have been established with communication plans as the primary element and 7 of those have been released.[207]

The ETASU provisions and accompanying implementation systems are regulatory actions that serve to restrict access to particular medications, and thus might include the separately noted regulatory action in the tracked safety issue guidance related to "limitations on a use in a specific population or subpopulation."[208] To be clear, safety-related changes to the labeling – particularly the *Contraindications* section – may also accomplish this goal in a less burdensome fashion. The FDA has implemented restricted distribution programs in a variety of circumstances in the past[209], albeit with uncertain legal authority to do so except under certain conditions. Specifically, these ETASU provisions pertain to *known serious risks* and restrict utilization in one of the following ways:

> "(A) health care providers who prescribe the drug have particular training or experience, or are specially certified (the opportunity to obtain such training or certification with respect to the drug shall be available to any willing provider from a frontier area in a widely available training or certification method (including an on-line course or via mail) as approved by the Secretary at reasonable cost to the provider); (B) pharmacies, practitioners, or health care settings that dispense the drug are specially certified (the opportunity to obtain such certification shall be available to any willing provider from a frontier area); (C) the drug be dispensed to patients only in certain health care settings, such as hospitals; (D) the drug be dispensed to patients with evidence or other

---

[204] 21 CFR § 200.5 See also K Uhl and P Honig, "Risk Management of Marketed Drugs: FDA and the Interface with the Practice of Medicine," *Pharmacoepidemiology and Drug Safety* 10, no. 3 (May 2001): 205–208.

[205] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355-1(e)(2)... "disseminating information about the elements of the risk evaluation and mitigation strategy to encourage implementation by health care providers of components that apply to such health care providers, or to explain certain safety protocols (such as medical monitoring by periodic laboratory tests);"

[206] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355-1(e)(3).

[207] Center for Drug Evaluation and Research, "News & Events - Slides for the June 7, 2012 Risk Evaluation and Mitigation Strategy (REMS) Assessments Public Workshop."

[208] U.S. Department of Health and Human Services, Food and Drug Administration, and Center for Drug Evaluation and Research, "Guidance: Classifying Significant Postmarketing Drug Safety Issues (Draft)."

[209] Judith C. Maro, "Development of a public health information infrastructure for postmarket evidence" (Thesis, Massachusetts Institute of Technology, 2009), 58–61, http://dspace.mit.edu/handle/1721.1/53058.

> documentation of safe-use conditions, such as laboratory test results; (E) each patient using the drug be subject to certain monitoring; or (F) each patient using the drug be enrolled in a registry."[210]

The evidentiary standard to issue a REMS with an ETASU provision specifically states that:

> "(A) the drug, which has been shown to be effective, but is associated with a serious adverse drug experience, can be approved only if, or would be withdrawn unless, such elements are required as part of such strategy to mitigate a specific serious risk listed in the labeling of the drug; and (B) for a drug initially approved without elements to assure safe use, other elements under subsections (c), (d), and (e) are not sufficient to mitigate such serious risk."[211]

As of April 2012, the FDA reports that 64 medical products are subject to ETASU provisions.[212] Among these, rosiglitazone-containing medications, long acting/extended release opioids, and transmucosal immediate-release fentanyl products are examples of products for which REMS were created postapproval to manage risks.[213] The latter two are class-wide REMS for opioid-containing medications, in which the known risks to be mitigated are supratherapeutic or "Type A" effects: addiction, abuse, misuse, overdose, and death. The REMS for rosiglitazone is notable in that continually refers to the "**potential** increased risk of myocardial infarction,"[214] (emphasis added) despite the legislative language requiring **known** serious risks.

3.3.3 Issuance of a Postmarket Requirement and the conduct of a study by FDA

In the Food and Drug Administration Amendments Act, Congress also granted the FDA new legal authorities to require manufacturers to conduct postmarket studies.

---

[210] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355-1(f)(3).

[211] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355-1(f)(1).

[212] Center for Drug Evaluation and Research, Food and Drug Administration, and U.S. Department of Health and Human Services, "Advances in FDA's Safety Program for Marketed Drugs" (FDA, April 2012), 10, http://www.fda.gov/downloads/Drugs/DrugSafety/UCM300946.pdf.

[213] See Center for Drug Evaluation and Research, "Postmarket Drug Safety Information for Patients and Providers - Approved Risk Evaluation and Mitigation Strategies (REMS)," WebContent, n.d., http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm111350.htm.

[214] I point to this example to demonstrate the difficulty in reaching consensus on causality. Center for Drug Evaluation and Research, "Approved Risk Evaluation and Mitigation Strategies (REMS): Avandia (rosiglitazone)," WebContent, n.d., http://www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/UCM255624.pdf.

However, it did not do so unconditionally. First, the FDA can exercise these new legal authorities either at the time of approval (i.e., licensure)[215], or in the postmarket period. In the latter case, the legal authority can be exercised "only if the Secretary [of Health and Human Services] becomes aware of new safety information."[216] Second, the FDA is not permitted to require a manufacturer to perform new postmarket requirements in the form of postapproval studies or postapproval clinical trials:

> "**unless** the Secretary [of Health and Human Services] makes a determination that the reports under subsection (k)(1) [i.e., spontaneous reporting systems] and the active postmarket risk identification and analysis system as available under subsection (k)(3) **will not be sufficient** to meet the purposes set forth in subparagraph (B)" [217]
> (emphasis added)

To summarize, the FDA has three supporting assertions to make prior to generating a postmarket requirement: a) the data (either the scientific data available pre-licensure or new safety information) support the postmarket requirement's generation; b) the purpose of the postmarket requirement satisfies one of the three purposes discussed in subsection 3.1 and c) the findings of sufficiency regarding the spontaneous reporting systems and active postmarket risk identification and analysis system.[218] For postmarket requirements generated at the time of approval, sufficiency determinations have been made independent of actual use of the system.[219] Letters informing manufacturers of postmarket requirements generated postapproval are not public, and therefore it is unclear how these sufficiency determinations have been made.

---

[215] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*, codified at 21 U.S.C. § 355(o)(3)(A). At the time of licensure, the FDA can require a postmarket requirement "on the basis of scientific data deemed appropriate by the Secretary, including information regarding chemically-related or pharmacologically-related drugs."

[216] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85,* codified at 21 U.S.C. § 355(o)(3)(C). For the definition of new safety information, see *supra* at note 152.

[217] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85,* codified at 21 U.S.C. § 355(o)(3)(D). For the language with respect to the purposes, see *supra* at note 162.

[218] For example of these three assertions, see the "Postmarketing Requirements under 505(o) section of the FDA's recent approval letter for Myrbetriq (mirabegron). Search the FDA's approved drugs database at "Drugs@FDA: FDA Approved Drug Products", n.d., http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm.

[219] The FDA has stated the results of such determinations in letters to manufacturers at the time of approval. The language in these letters has generally stated, "The new pharmacovigilance system that FDA is required to establish under section 505(k)(3) of the FDCA will not be sufficient to assess this serious risk." To find these letters, one must search the FDA database for newly approved drugs: http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm.

There are no known legal standards indicating what might trigger "the conduct of a safety-related observational epidemiological study by FDA."[220] Further, it is unclear whether a safety-related observational epidemiological study by FDA would be performed within the Mini-Sentinel System as either a protocol-based one-time assessment or a sequential database surveillance assessment, or whether it would be performed under the existing pharmacoepidemiology contract.[221]

3.3.4 Withdrawal

The most stringent regulatory action is withdrawal of a medical product, or removal of its licensure for use. The legal standard for withdrawal of a medical product or a particular indication uses similar broad language to the REMS provisions. Specifically, withdrawal is permitted if new data:

> "show that such drug is unsafe for use under the conditions of use upon the basis of which the application was approved... [or] that there is a lack of substantial evidence that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling thereof."[222]

The first part of the above rationale addresses safety issues whereas the latter addresses efficacy. In the past, varying sources of evidence have supported safety-related withdrawals. Some drugs, e.g., levomethadyl acetate and oral bromfenac sodium, were withdrawn on the basis of risks identified in the spontaneous reporting systems and assessed by creating a case-series analysis (i.e., an in-depth study of particular cases) whereas others, e.g., rofecoxib, were withdrawn on the basis of postmarket clinical trial data.[223]

3.4 Causal Inference – How Necessary for Regulatory Action?

In summary, with the exception of changes to the *Warnings and Precautions* section of a product's label, the law is ambiguous on the strength of causal inference needed to

---

[220] U.S. Department of Health and Human Services, Food and Drug Administration, and Center for Drug Evaluation and Research, "Guidance: Classifying Significant Postmarketing Drug Safety Issues (Draft)."
[221] See *supra* at note 67.
[222] 21 U.S.C. § 355(e). To be clear, this is not a full quotation of the legislative language. Although they are valid grounds for withdrawal, I am excluding issues related to fraud or falsification, or a sponsor's lack of compliance with regard to required reports, inspections, and other assessments.
[223] See Zaina P Qureshi et al., "Market Withdrawal of New Molecular Entities Approved in the United States from 1980 to 2009," *Pharmacoepidemiology and Drug Safety* 20, no. 7 (July 2011): 772–777; Dal Pan, Lindquist, and Gelperin, "Postmarketing Spontaneous Pharmacovigilance Reporting Systems."

support various regulatory actions, leaving much to the discretion of the FDA. While the FDA might prefer to regulate with robust causal evidence to support its action(s), time and ethical constraints may rule out study designs that are likely to generate such evidence (let us assume said designs are well-executed). To that end, Hernán and Robins describe a study design that guarantees robust causal inference: the ideal randomized experiment. It is characterized by "no loss to follow-up, full adherence to the assigned treatment over the duration of the study, a single version of treatment, and double blind assignment."[224] Such conditions are improbable even in a double blind randomized controlled trial.[225] Generally, choosing a study design likely to draw the most robust conclusions of causality comes at a price: longer timeframes and higher costs to generate the evidence.[226]

To that end, randomized controlled trials may be reasonably ruled out as evidence-generating systems from the outset when the outcome is so rare that the costs of such a trial would be prohibitive.[227] Second, they may be ethically questionable when morbidity or mortality associated with the outcome is high and a premium is placed on earlier rather than later regulatory decision-making. Third, the existence of safe substitute products may also favor earlier rather than later regulatory decision-making.[228] Finally, randomized controlled trials may start too late since the Food and Drug Administration Amendments Act stages them as a study design of last resort.[229] At that point, it is difficult to maintain clinical equipoise, which is an ethical requirement for investigator-assigned treatments.[230]

---

[224] Hernán and Robins, *Causal Inference*, 14–15.
[225] Hernán and Hernández-Díaz, "Beyond the Intention-to-treat in Comparative Effectiveness Research."
[226] Randomized controlled trials were discussed earlier in subsection 2.1.3.
[227] See, for example, FDA advisory committee discussion on potential new studies that may clarify conflicting observational studies showing an association between venous thromboembolism and drospirenone-containing contraceptives. The advisory committee discusses the infeasibility of randomized clinical trials because of the rareness of the outcome. See Center for Drug Evaluation and Research, "Reproductive Health Drugs Advisory Committee - Briefing Information for the December 9, 2011 Joint Meeting of the Advisory Committee for Reproductive Health Drugs and the Drug Safety and Risk Management Advisory Committee."
[228] See for example, arguments that the availability of Actos® (pioglitazone) favored regulatory action on Avandia® (rosiglitazone) without necessitating completion of a randomized controlled trial. Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 62.
[229] § 901 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85,* codified at 21 U.S.C. § 355(o)(3)(D)(ii).
[230] Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 187.

In general, the circumstances of the particular exposure-outcome pair, as Hernán and Robins point out, require one to assess whether conclusions of causation are really necessary, or whether prediction (i.e., associational measures) will suffice?[231] Similarly, the IOM has stated:

> "**Passive surveillance, epidemiologic research with administrative databases, and active surveillance can be used to answer many drug safety questions.** When they do not provide definitive answers, they can sometimes provide guidance for the development of further studies or provide sufficient information to narrow the uncertainty about drug-related risks and benefits and **guide regulatory actions and the decisions of patients and providers.** In some instances, full-scale observational studies or clinical trials will be required to answer key questions, particularly if the outcome of interest is common in the patients taking a drug."[232] (emphasis added)

When a randomized controlled trial is ruled out for time or ethical reasons as discussed above **and** when robust findings of causality are still desired, it is possible that observational data **may still** support such findings. Essentially, the observational data must be transformed to emulate a randomized trial.[233] Specifically, observational studies may support causal inference when three criteria are met that allow investigators to conclude that *association is equivalent to causation*. These three criteria are:

> "1. the values of treatment under comparison correspond to well-defined interventions
> 2. the conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on the measured covariates
> 3. the conditional probability of receiving every value of treatment is greater than zero, i.e., positive."[234]

In general, most pharmacoepidemiologic studies easily can satisfy conditions 1 and 3. Condition 2 essentially requires the elimination of bias – selection bias, measurement bias, or confounding bias – that may be responsible for explaining the relationship between the exposure and outcome. Specific discussions of these biases and the ability to eliminate them using data in the Mini-Sentinel System are described next.

---

[231] Hernán and Robins, *Causal Inference*, 38.
[232] Institute of Medicine (IOM), *The Future of Drug Safety: Promoting and Protecting the Health of the Public*, 115.
[233] Hernán and Robins, *Causal Inference*, 26. "For each causal question that we intend to answer using observational data, we will need to carefully describe (i) the randomized experiment that we would like to, but cannot, conduct, and (ii) how the observational study emulates that randomized experiment." See also Miguel A Hernán et al., "Observational Studies Analyzed Like Randomized Experiments: An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease," *Epidemiology (Cambridge, Mass.)* 19, no. 6 (November 2008): 766–779.
[234] Hernán and Robins, *Causal Inference*, 26.

It is sometimes possible to make causal inferences even in the presence of bias when the effect size and direction together are of such a magnitude that the impact of the bias is inconsequential.[235] For example, some drugs are withdrawn on the basis of a case-series analysis[236] and these data do not emulate a randomized experiment. In short, for these circumstances, no amount of bias could explain the effect size[237] observed. Thus, it may be possible to conclude causality with the following criteria:

> "the suspected [adverse event]: i) is rare in the population when the medication is not used, ii) is not a manifestation of the underlying disease, iii) has a strong temporal association with drug administration, and iv) is biologically plausible as a drug reaction or is generally the result of a drug reaction based on other clinical experience."[238]

In the next section, I address potential circumstances that limit causal inference in the Mini-Sentinel System, and develop a qualitative tool by which the FDA might make an initial sufficiency decision: The Mini-Sentinel System Pre-screening Checklist.

---

[235] Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 150.

[236] See *supra* at note 223.

[237] The effect size is the quantitative strength of an association, which is usually a point estimate of the effect. See Strom, "Basic Principles of Clinical Epidemiology Relevant to Pharmacoepidemiologic Studies." Some epidemiologist use the phrase "effect" only to denote casual relationships as opposed to associations. That is not how I use it here. Typically, epidemiologic effect sizes are given in absolute terms as risk differences or rate differences, or are given in relative terms as risk ratios, rate ratios, or odds ratios. See Rothman, Greenland, and Lash, *Modern Epidemiology*, 51–70.

[238] Dal Pan, Lindquist, and Gelperin, "Postmarketing Spontaneous Pharmacovigilance Reporting Systems," 148.

4 MINI-SENTINEL SYSTEM PRE-SCREENING CHECKLIST

As I described in the last section, evidence developed in the Mini-Sentinel System may not be assumed to produce causal inferences. However, causal inference *may still be possible* with sufficient covariate data and elimination of systematic biases. Alternatively, causal inferences may still be possible even in the presence of known but unmeasured bias when the effect size and direction[239] are of such a magnitude that the impact of the bias is inconsequential.[240] In these circumstances, it may be unnecessary (and an inefficient use of public dollars) to more precisely quantify the biases.

This section first summarizes sources of bias using a framework from the epidemiologic literature[241] that was adopted by the Institute of Medicine (IOM).[242] Using this framework, I then address the potential for elimination of these biases in the Mini-Sentinel System. Finally, I address the precision[243] and transportability[244] of statistical inference that can reasonably be supported by Mini-Sentinel System.

The section introduces a qualitative tool – what I call the Mini-Sentinel System Pre-screening Checklist - to address whether the Mini-Sentinel System is likely to be sufficient (or insufficient) *on its face* to generate evidence to resolve a tracked safety issue. If the pre-screening checklist suggests that the Mini-Sentinel System is insufficient,

---

[239] For a definition of effect size, see *supra* at note 237.

[240] Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 150. "Very large relative increases in the background rate, such as the almost 1,000-fold increase in progressive multifocal leukoencephalopathy with natalizumab treatment in patients with multiple sclerosis or Crohn's Disease …, or the greater than ten-fold increase in intussusception seen with rotavirus vaccine, are likely beyond the bounds of anything that can be explained through imbalances on other risk factors for those outcomes, that is confounders. In the setting of large relative risks for an adverse event, designs with quite weak control of confounding, …, might be sufficient for public policy purposes."

[241] Hernán and Robins, *Causal Inference*; Kenneth J Rothman, Sander Greenland, and Timothy L. Lash, "Validity in Epidemiologic Studies," in *Modern Epidemiology*, ed. Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash, Third. (Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008), 128–147.

[242] Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 103–144.

[243] "The statistical precision of a measurement or process is often taken to be the inverse of the variance of the measurements or estimates that the process produces…Precision of estimation can be improved (which is to say, variance can be reduced) by increasing the size of the study." Rothman, Greenland, and Lash, *Modern Epidemiology*, 149.

[244] In epidemiology, transportability has been suggested as a more appropriate term for what is commonly referred to as external validity or generalizability. The IOM has endorsed this term in their most recent report, explaining "the committee thinks that [the term transportability] reflects a nonbinary characteristic better. Different effects can occur in a variety of settings, and study results may be transportable to some populations or settings but not others, so transportability may not be a simple binary property." See Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 118.

then no further analytics are necessary and regulators may have justification to pursue alternative evidence generation mechanisms. If the pre-screening checklist suggests that the Mini-Sentinel System *is sufficient* for evidence generation to support resolution of a tracked safety issue, then quantitative tools may be used to refine the estimation of the Mini-Sentinel System's evidence generation capability, which will be presented in section 6.

Necessary inputs to the Mini-Sentinel System Pre-screening Checklist are: 1) a tracked safety issue that identifies a particular exposure-outcome pair of interest; and 2) a regulator's goal with respect to the strength of causal inference necessary to support regulatory decision-making as was outlined in the previous section.

4.1 Importance of Effect Sizes

Most epidemiologists presume the presence of bias in any observational study, and are particularly sensitive to effect sizes that are marginally different than the null hypothesis.[245] Likewise, several regulators have expressed reluctance to rely on observational studies that demonstrate modest elevated risk (e.g., relative risks less than two-fold).[246] However, even small relative risks can be quite important with high prevalence medications, and in these circumstances, the IOM suggests that "a well-designed and well-conducted postmarketing randomized clinical trial is the best approach for characterizing the risk–benefit profile."[247] An example of such a scenario is the long-acting beta agonists-death tracked safety issue that is currently being investigated via

---

[245] Strom notes "A quantitatively small association may still be causal but it could be created by a subtle error, which would not be apparent in evaluating the study." Strom, "Basic Principles of Clinical Epidemiology Relevant to Pharmacoepidemiologic Studies," 42. See also Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 120–121, 205..."If the estimated relative risks are small, selection bias, confounding, and measurement error may be alternative explanations for associations found in an observational study."

[246] Temple, "Meta-analysis and Epidemiologic Studies in Drug Development and Postmarketing Surveillance"; J Woodcock, "Evidence Vs. Access: Can Twenty-first-century Drug Regulation Refine the Tradeoffs?," *Clinical Pharmacology and Therapeutics* 91, no. 3 (March 2012): 378–380. See also comments of Dr. John Jenkins in Food and Drug Administration and Center for Drug Evaluation and Research, "Transcript of the Joint Meeting of Pulmonary-Allergy Drugs Advisory Committee and the Drug Safety and Risk Management Advisory Committee", March 11, 2010, http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/Pulmonary-AllergyDrugsAdvisoryCommittee/UCM209124.pdf.

[247] Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 205.

FDA-required randomized trials.[248] In general, a pre-screen criterion for use of the Mini-Sentinel System should consider the overall effect size that investigators seek to detect/rule out, and small effect sizes may be a reason for regulators to eliminate consideration of the Mini-Sentinel System as an evidence generation system.

4.2 Sources of Bias

The IOM defines bias as the "difference between the average effect of many hypothetical repetitions of a given study and the true effect in the population being studied."[249] Alternatively, Hernán and Robins define bias as instances "whenever the effect measure (e.g., causal risk ratio or difference) and the corresponding association measure (e.g., associational risk ratio or difference) are not equal."[250] Systematic biases are problematic because they are not represented by confidence intervals surrounding risk estimates. Rather, they shift the mean of confidence intervals in ways that may not be immediately apparent to investigators. Bias may be controlled for in design or analysis, and much of the epidemiology literature is devoted to these controls. Bias can be segregated into three types based on its structure: confounding, selection, and measurement bias, which are each reviewed in turn.

4.2.1 Confounding Bias

Rothman et al. require a confounding variable to meet three conditions:

> "a) the exposure is associated with the confounder,
> b) the outcome is independently associated with the confounder,
> c) temporally, the confounder is NOT affected by either the exposure or the outcome."[251]

In other words, confounding occurs "when the populations compared in a study differ in important predictors of the outcome being studied other than an exposure of interest

---

[248] See transcripts discussing the difficulty in detecting/ruling out a relative risk of <2.0 in observational studies. Food and Drug Administration and Center for Drug Evaluation and Research, "Transcript of the Joint Meeting of Pulmonary-Allergy Drugs Advisory Committee and the Drug Safety and Risk Management Advisory Committee." See also the FDA's announcement on its requirement for randomized controlled trials, including the decision to pursue a composite endpoint because of the rareness of death as an outcome. Food and Drug Administration and Center for Drug Evaluation and Research, "FDA Requires Post-market Safety Trials for Long-Acting Beta Agonists."

[249] Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 116.

[250] Hernán and Robins, *Causal Inference*, 77.

[251] Rothman, Greenland, and Lash, *Modern Epidemiology*, 758.

(such as exposure to a drug)."[252] In short, the presence of an uncontrolled confounder potentially offers an alternative explanation for the cause of the outcome of interest (e.g., adverse events). The presence of confounders in an observational study is virtually guaranteed because treatment assignment is not random. Confounding is more likely to occur when the disease process itself (independent of the therapy assigned) is a risk factor for the outcome of interest. Oft-cited examples include diabetes-cardiovascular risk, and asthma-sudden death. These situations are sometimes referred to as "confounding by indication."[253] This same concept can also be called "confounding by contraindication" or channeling bias, and occurs when a predisposing condition/status of a patient may steer them away from a particular therapy (e.g., perhaps the patient has a known reaction to a competing therapy).

In general, pharmacoepidemiologists are accustomed to inevitable confounding in observational datasets built on secondary data, and consequently, have developed several adjustment techniques. Schneeweiss[254] and others[255] summarize these techniques depending on whether the covariates are measured, unmeasured, or unmeasureable.[256] Measured covariates can be adjusted for with design-based mechanisms (e.g., matching[257], restriction[258]) or analysis-based mechanisms (e.g., stratification[259], inverse

---

[252] Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 117.

[253] Much has been written on confounding by indication. See Ibid., 122; Vandenbroucke and Psaty, "Benefits and Risks of Drug Treatments"; Strom, "Methodologic Challenges to Studying Patient Safety and Comparative Effectiveness"; Schneeweiss and Avorn, "A Review of Uses of Health Care Utilization Databases for Epidemiologic Research on Therapeutics."

[254] Sebastian Schneeweiss, "Sensitivity Analysis and External Adjustment for Unmeasured Confounders in Epidemiologic Database Studies of Therapeutics," *Pharmacoepidemiology and Drug Safety* 15, no. 5 (May 2006): 291–303.

[255] Brookhart et al., "Confounding Control in Healthcare Database Research: Challenges and Potential Approaches"; Alex D McMahon, "Approaches to Combat with Confounding by Indication in Observational Studies of Intended Drug Effects," *Pharmacoepidemiology and Drug Safety* 12, no. 7 (November 2003): 551–558.

[256] Unmeasured data are existing data that are missing or unavailable to the investigator whereas unmeasureable data simply do not exist. In an unrelated policy arena, Donald Rumsfeld famously referred to the former as known unknowns and the latter as unknown unknowns.

[257] Kenneth J Rothman, Sander Greenland, and Timothy L. Lash, "Design Strategies to Improve Study Accuracy," in *Modern Epidemiology*, ed. Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash, Third. (Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008), 168–182.

[258] Bruce M Psaty and David S Siscovick, "Minimizing Bias Due to Confounding by Indication in Comparative Effectiveness Research: The Importance of Restriction," *JAMA: The Journal of the American Medical Association* 304, no. 8 (August 25, 2010): 897–898; S. Schneeweiss et al., "Increasing Levels of Restriction in Pharmacoepidemiologic Database Studies of Elderly and Comparison with Randomized Trial Results," *Medical Care* 45, no. 10 Supl 2 (2007): S131–42.

probability weighting[260], g-estimation methods[261]). Unmeasured covariates may be adjusted for with a subset of internal data[262], external data[263], high-dimensional propensity scores[264], or disease risk scores.[265] Unmeasurable (i.e., unknown) covariates may be dealt with via sensitivity analysis (also known as bias analysis)[266], instrumental variables analysis[267] or self-controlled study designs[268]. The key issue for regulators is to assess the potential for confounding bias with regard to the particular tracked safety issue

---

[259] Sander Greenland and Kenneth J Rothman, "Introduction to Stratified Analysis," in *Modern Epidemiology*, ed. Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash, Third. (Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008), 258–302.

[260] Stephen R Cole and Miguel A Hernán, "Constructing Inverse Probability Weights for Marginal Structural Models," *American Journal of Epidemiology* 168, no. 6 (September 15, 2008): 656–664; Miguel A Hernán and James M Robins, "Estimating Causal Effects from Epidemiological Data," *Journal of Epidemiology and Community Health* 60, no. 7 (July 2006): 578–586.

[261] Sarah L Taubman et al., "Intervening on Risk Factors for Coronary Heart Disease: An Application of the Parametric G-formula," *International Journal of Epidemiology* 38, no. 6 (December 2009): 1599–1611.

[262] Sengwee Toh, Luis A García Rodríguez, and Miguel A Hernán, "Analyzing Partially Missing Confounder Information in Comparative Effectiveness and Safety Research of Therapeutics," *Pharmacoepidemiology and Drug Safety* 21 Suppl 2 (May 2012): 13–20.

[263] Til Stürmer et al., "Adjustments for Unmeasured Confounders in Pharmacoepidemiologic Database Studies Using External Information," *Medical Care* 45, no. 10 Supl 2 (October 2007): S158–165; Sebastian Schneeweiss et al., "Adjusting for Unmeasured Confounders in Pharmacoepidemiologic Claims Data Using External Information: The Example of COX2 Inhibitors and Myocardial Infarction," *Epidemiology (Cambridge, Mass.)* 16, no. 1 (January 2005): 17–24.

[264] Rassen and Schneeweiss, "Using High-dimensional Propensity Scores to Automate Confounding Control in a Distributed Medical Product Safety Surveillance System"; Rassen et al., "Privacy-maintaining Propensity Score-based Pooling of Multiple Databases Applied to a Study of Biologics"; S. Schneeweiss et al., "High-dimensional Propensity Score Adjustment in Studies of Treatment Effects Using Health Care Claims Data," *Epidemiology (Cambridge, Mass.)* 20, no. 4 (2009): 512–522.

[265] Glynn, Gagne, and Schneeweiss, "Role of Disease Risk Scores in Comparative Effectiveness Research with Emerging Therapies."

[266] Timothy L Lash et al., "Methods to Apply Probabilistic Bias Analysis to Summary Estimates of Association," *Pharmacoepidemiology and Drug Safety* 19, no. 6 (June 2010): 638–644; Onyebuchi A Arah, Yasutaka Chiba, and Sander Greenland, "Bias Formulas for External Adjustment and Sensitivity Analysis of Unmeasured Confounders," *Annals of Epidemiology* 18, no. 8 (August 2008): 637–646; Lawrence C McCandless, Paul Gustafson, and Adrian R Levy, "A Sensitivity Analysis Using Information About Measured Confounders Yielded Improved Uncertainty Assessments for Unmeasured Confounding," *Journal of Clinical Epidemiology* 61, no. 3 (March 2008): 247–255; Sander Greenland, "Multiple-bias Modelling for Analysis of Observational Data," *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168, no. 2 (2005): 267–306; Timothy L Lash and Aliza K Fink, "Semi-automated Sensitivity Analysis to Assess Systematic Errors in Observational Data," *Epidemiology (Cambridge, Mass.)* 14, no. 4 (July 2003): 451–458.

[267] M Alan Brookhart, Jeremy A Rassen, and Sebastian Schneeweiss, "Instrumental Variable Methods in Comparative Safety and Effectiveness Research," *Pharmacoepidemiology and Drug Safety* 19, no. 6 (June 2010): 537–554; Jeremy A Rassen et al., "Instrumental Variables I: Instrumental Variables Exploit Natural Variation in Nonexperimental Data to Estimate Causal Relationships," *Journal of Clinical Epidemiology* 62, no. 12 (December 2009): 1226–1232; M Alan Brookhart et al., "Evaluating Short-term Drug Effects Using a Physician-specific Prescribing Preference as an Instrumental Variable," *Epidemiology (Cambridge, Mass.)* 17, no. 3 (May 2006): 268–275; S Greenland, "An Introduction to Instrumental Variables for Epidemiologists," *International Journal of Epidemiology* 29, no. 4 (August 2000): 722–729.

[268] Self-controlled study designs will be explained in more detail in section 5.1

of interest in the Mini-Sentinel System, and particularly to identify whether important confounding covariates are measured, unmeasured, or unmeasurable.

4.2.1.1 Limited Covariate Data in the Mini-Sentinel System

As a primarily claims-based system with some laboratory results, many commonly sought confounding covariates[269] – e.g., smoking status, alcohol use, body mass index, socio-economic status - are simply uncoded in the Mini-Sentinel System. Some of these uncoded covariates may be measured via medical chart confirmation, but these data are occasionally missing in medical charts as well.[270] When singular covariates of importance are not identified, high-dimensional propensity scores and disease risk scores can potentially account for these covariates. These scores aggregate many covariates into a single summary measure. A high-dimensional propensity score is a summary measure of a patient's "propensity" or likelihood of being exposed to the medical product of interest and a disease risk score is a summary of the patient's likelihood of experiencing the disease when they are unexposed to the medical product of interest. Theoretically, these scores can be used to match similar patients and thus eliminate the bias created by confounding covariates. Also, creation of these summary scores limits sharing of patient-level data among data partners. For these reasons, the automated use of such techniques in the Mini-Sentinel System is growing.[271]

In general, absent a self-controlled design or other technique to mimic randomization (e.g., instrumental variables), observational studies performed in the Mini-Sentinel System will retain some degree of residual confounding. As of yet, there has been little routine implementation of sensitivity/bias analyses in the Mini-Sentinel System to estimate the strength of bias necessary to eliminate findings of an effect size. These sensitivity analyses would propose the existence of a bias at some level (e.g., the prevalence of a confounder in both the treatment group and the comparison group) and then test how the presence of this bias affects the risk estimate. In the future, regulators

---

[269] For more examples of such covariates, see Schneeweiss et al., "Adjusting for Unmeasured Confounders in Pharmacoepidemiologic Claims Data Using External Information."
[270] See, for example, discussion on the difficulty in covariate capture in Dore et al., "A Cohort Study of Acute Pancreatitis in Relation to Exenatide Use."
[271] Rassen and Schneeweiss, "Using High-dimensional Propensity Scores to Automate Confounding Control in a Distributed Medical Product Safety Surveillance System."

may make use of these sensitivity/bias analyses to determine whether potential confounding bias in the Mini-Sentinel System is tolerable. Such determinations would also require the regulator to estimate the effect sizes of interest.

4.2.2 Selection Bias

Selection bias refers to how people were selected into the analysis, and results when "the exposure affects selection in the study or analysis and selection is associated with the outcome of interest."[272] Hernán and Robins emphasize that both the exposure and outcome must be associated with a *common effect,* or something that follows the exposure and outcome.[273] The common effect is typically a censoring condition, or a condition that causes people to be excluded from a study. Examples of selection bias include differential loss to follow-up, censoring due to missing data, healthy user bias, self-selection/volunteer bias, and prevalent user bias. Evidence generation in the Mini-Sentinel System is more vulnerable to the first two forms of selection bias noted above. Differential loss to follow-up occurs when persons in the two groups being compared stop contributing information (i.e., drop out) at different rates. This is problematic because no one knows what would have occurred had they kept contributing information. Censoring due to missing data can just be seen as another form of dropout, i.e. the investigator censors or omits data because it is incomplete. If this missing data is not random (i.e., presumably distributed in the two groups proportional to their size), this is a form of selection bias. Unlike confounding bias, which is presumed eliminated via randomization, any statistical inference can suffer from selection biases. Additionally, issues such as differential loss-to-follow-up are difficult to predict in prospective designs. However, these biases may be corrected in data analysis using inverse probability weighting and g-estimation methods.[274]

With regard to other forms of selection bias, the healthy user bias generally occurs with therapeutics that are considered "preventative" such as statins, and the concern is that users of such therapies would also engage in other forms of health-seeking behavior

---

[272] Institute of Medicine (IOM), *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*, 117.
[273] Hernán and Robins, *Causal Inference*, 95–108.
[274] See *supra* at notes 260-261.

and thus, be markedly different than non-users.[275] This concern is generally mitigated in the design by choice of comparator group, and specifically, avoiding comparisons between users and non-users.[276] Second, self-selection/volunteer bias will not apply in database studies because individual consent is not required. Finally, it is now routine for investigators to adopt "incident user" designs[277] to correct for biases created by prevalent users.[278]

4.2.2.1 Churn/Database Turnover in the Mini-Sentinel System

Differential loss-to-follow-up in the Mini-Sentinel System is typically related to changes in insurance coverage and death. For example, inclusion criteria typically requires a subject of study (i.e., a patient) to have continuous insurance coverage during the time the subject is "at risk" of developing the adverse event. If some aspect of being exposed to a particular medical product causes a loss/change in insurance coverage and the outcome of interest is also associated with a loss/change in insurance coverage, then selection bias may have occurred. These types of insurance coverage changes may be less likely due to the "guaranteed issue" and "community rating" provisions of the Patient Protection and Affordable Care Act of 2010.[279] These provisions prevent denial of coverage as a result of pre-existing conditions. However, increased competition among insurers may still generate significant churn in health plans, which could lead to differential loss-to-follow-up. Similarly, if patients die during the at-risk period at notably different rates, these deaths create differential loss-to-follow-up problems. Unfortunately, it may be difficult to anticipate the degree to which selection bias may affect evidence generation for particular tracked safety issues in the Mini-Sentinel System. However, long follow-up times (i.e., long "at risk" windows following exposure for outcomes such

---

[275] William H. Shrank, Amanda R. Patrick, and M. Alan Brookhart, "Healthy User and Related Biases in Observational Studies of Preventive Interventions: A Primer for Physicians," *Journal of General Internal Medicine* 26, no. 5 (May 2011): 546–550.
[276] Choice of comparison groups will be discussed at greater length herein in section 5.1.
[277] Wayne A Ray, "Evaluating Medication Effects Outside of Clinical Trials: New-user Designs," *American Journal of Epidemiology* 158, no. 9 (November 1, 2003): 915–920.
[278] Goodarz Danaei, Mohammad Tavakkoli, and Miguel A Hernán, "Bias in Observational Studies of Prevalent Users: Lessons for Comparative Effectiveness Research from a Meta-analysis of Statins," *American Journal of Epidemiology* 175, no. 4 (February 15, 2012): 250–262.
[279] *Patient Protection and Affordable Care Act, P.L. 111-148*, 2010.

as cancer) are likely to be difficult to achieve. Gagne et al. also refer to these concepts as the onset and duration of the risk window.[280]

4.2.3 Measurement Bias

Measurement bias (also known as information bias) occurs when "the association between treatment and outcome is weakened or strengthened as a result of the process by which the study data are measured."[281] Measurement bias with respect to discrete variables is often referred to as misclassification. Misclassification can occur when measuring exposures, outcomes, and covariates, and can occur for a variety of reasons in database systems.[282] I focus on misclassification rather than measurement bias because the Mini-Sentinel System is built around count-based data[283], which require discrete categorizations. Covariate measurement may be discrete or continuous, but generally continuous covariates require regression models, which are not easily achieved without pooling data.[284] However, as mentioned in the section 4.2.1.1, high-dimensional propensity scores or disease risk scores may circumvent this limitation.

Misclassification is further described as independent or dependent, and as non-differential or differential. Independent misclassification typically refers to situations when misclassification of one variable (e.g., exposures) is independent of misclassification in other variables (e.g., outcomes). Non-differential misclassification refers to situations when the treatment and comparator groups have identical patterns of measurement error (i.e., the sensitivity and specificity of the classification are the same) with respect to a particular variable (e.g., exposure). There are many corrective techniques to deal with misclassification that is independent and non-differential in observational data.[285] For the purposes of the Mini-Sentinel System Pre-screening

---

[280] Gagne et al., "Design Considerations in an Active Medical Product Safety Monitoring System."
[281] Hernán and Robins, *Causal Inference*, 109.
[282] For a comprehensive summary, see Jessica Chubak, Gaia Pocobelli, and Noel S Weiss, "Tradeoffs Between Accuracy Measures for Electronic Health Care Data Algorithms," *Journal of Clinical Epidemiology* 65, no. 3 (March 2012): 343–349.e2; Schneeweiss and Avorn, "A Review of Uses of Health Care Utilization Databases for Epidemiologic Research on Therapeutics."
[283] Curtis et al., "Design Considerations, Architecture, and Use of the Mini-Sentinel Distributed Data System."
[284] The Mini-Sentinel System is designed to limit pooling data. See *supra* at 142.
[285] J. P. Mullooly, "Misclassification Model for Person-time Analysis of Automated Medical Care Databases," *American Journal of Epidemiology* 144, no. 8 (1996): 782–792; H. Brenner and O. Gefeller, "Use of the Positive Predictive Value to Correct for Disease Misclassification in Epidemiologic Studies,"

Checklist, it is more important to assess whether particular tracked safety issues are more vulnerable to misclassification, and the ability of the Mini-Sentinel System to adjust for these biases. Quantitative methods to adjust for misclassification bias are discussed at length in Section 8.

4.2.3.1 Exposure Misclassification in the Mini-Sentinel System

Exposure to a therapeutic is not often measured, but assumed via proxy variables. When using claims data, there is an assumption that a medical product that has been dispensed is a medical product that has been used therapeutically in accordance with the dispensing instructions. For self-administered therapeutics, these assumptions may be quite strong as they assume strict adherence to the prescribing regimen. Absent additional primary data collection mechanisms such as body-based sensors or mobile phone-based applications for self-reporting, there are few ways to collect validation data to estimate the potential degree of exposure misclassification. The Mini-Sentinel System is not designed to enable such primary data collection from patients due to its status as public health activity.[286] Generally, there are fewer adherence concerns for infusions or injections that are administered in a healthcare setting (i.e., sometimes termed medically-attended exposures).

Assuming adherence when it is not present creates false positive exposures, but false negative exposures are also possible. This type of exposure misclassification may arise when patients obtain therapeutics outside of their insurance coverage. Recent studies have documented this type of exposure misclassification due to restrictive formularies[287], and due to therapeutics that are available in both prescription and over-the-counter status.[288]

American Journal of Epidemiology 138, no. 11 (1993): 1007–1015; M. S. Green, "Use of Predictive Value to Adjust Relative Risk Estimates Biased by Misclassification of Outcome Status," *American Journal of Epidemiology* 117, no. 1 (1983): 98–105.

[286] See *supra* at note 19.

[287] John-Michael Gamble et al., "Restrictive Drug Coverage Policies Can Induce Substantial Drug Exposure Misclassification in Pharmacoepidemiologic Studies," *Clinical Therapeutics* 34, no. 6 (June 2012): 1379–1386.e3.

[288] Joseph A C Delaney et al., "Demographic, Medical, and Behavioral Characteristics Associated with over the Counter Non-steroidal Anti-inflammatory Drug Use in a Population-based Cohort: Results from the Multi-Ethnic Study of Atherosclerosis," *Pharmacoepidemiology and Drug Safety* 20, no. 1 (January 2011): 83–89.

In summary, in terms of the Mini-Sentinel System Pre-screening Checklist, there may be exposure misclassification when investigating 1) therapeutics with predicted or known poor adherence patterns (e.g. anti-epileptic medications[289]), 2) therapeutics that are prescribed "as needed" such as migraine medications, 3) therapeutics that are not routinely included in commercial insurance formularies, and 4) therapeutics that have both prescription and over-the counter status (e.g., omeprazole).

4.2.3.2 Outcome Misclassification in the Mini-Sentinel System

Outcome misclassification is a major concern in the Mini-Sentinel System, and a focus of research.[290] Systematic validation of algorithms to detect particular outcomes has just begun.[291] Generally, these validation studies are estimating positive predictive values, which are only informative with respect to false positives. Estimation of false negatives, or missed outcomes, is difficult in database systems, although Chubak et al. suggest that use of narrow identification algorithms are likely to increase false negatives.[292]

Outcome misclassification is more likely to occur when diagnostic definitions of particular outcomes are ambiguous, uncertain, or evolving (e.g., psychiatric disorders) and diagnoses are made in the absence of laboratory values or other objective criteria. Error-prone diagnostic tests also contribute to outcome misclassification. Additionally, outcome misclassification may occur for diseases with slow disease progression because the "onset" date is unclear (e.g., cancer may be detected at a late stage so the precise "onset" date of the cancer may be impossible to know). Outcome misclassification may create problems for identifying the targeted outcome (i.e., the adverse event), but also for defining the initial cohort. Often, cohorts are defined with respect to the presence/absence

---

[289] Avani C Modi, Joseph R Rausch, and Tracy A Glauser, "Patterns of Nonadherence to Antiepileptic Drug Therapy in Children with Newly Diagnosed Epilepsy," *JAMA: The Journal of the American Medical Association* 305, no. 16 (April 27, 2011): 1669–1676.

[290] R. M. Carnahan, "Mini-Sentinel's Systematic Reviews of Validated Methods for Identifying Health Outcomes Using Administrative Data: Summary of Findings and Suggestions for Future Research," *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 90–99.

[291] Sarah L Cutrona et al., "Validation of Acute Myocardial Infarction in the Food and Drug Administration's Mini-Sentinel Program," *Pharmacoepidemiology and Drug Safety* (June 29, 2012), http://www.ncbi.nlm.nih.gov/pubmed/22745038.

[292] Chubak, Pocobelli, and Weiss, "Tradeoffs Between Accuracy Measures for Electronic Health Care Data Algorithms."

of pre-existing diagnoses. If these baseline outcomes are mismeasured, then the cohort itself is misclassified.

In general, use of *post hoc* corrective procedures to deal with misclassification requires it to be non-differential, which is a strong and untestable assumption in the Mini-Sentinel System. In fact, therapeutics other than vaccines are given to people with active disease processes, and additional follow-up medical visits and/or diagnostics tests are likely to follow from exposure to new therapeutics as part of the normal course of care. These follow-up visits and tests create differential opportunities to detect adverse events among the exposed, thereby creating differential misclassification. The use of active comparators as opposed to non-users may mitigate this possibility of differential misclassification. Complete medical chart confirmation is the only sure way to correct outcome misclassification, but these procedures may be very costly and time-consuming.

For the purposes of the Mini-Sentinel System Pre-Screening Checklist, there is a higher likelihood of outcome misclassification when investigating outcomes that are do not have clear onsets, or stable and repeatable diagnostic criteria over a variety of medical settings (e.g., specialty practices as compared to general practitioners, as well as emergency departments as compared to outpatient clinical settings).

4.2.3.3 Covariate Misclassification in the Mini-Sentinel System

As was discussed earlier, concerns regarding covariates tend to focus on whether they are measured at all, rather than on how accurately they are measured. If important covariates include concomitant exposures or co-morbidities, then these types of covariates are subject to the same misclassification problems described above. Other covariates tend to be related to demographics, or current health status. Measures of general health status often are at greater risk for misclassification. For example, although obesity has a medically specific definition, it is difficult to ascertain when body mass index is not explicitly coded or annotated in a medical record. Also, discretization of continuous covariates to enable count-based analyses results in loss of information (e.g., a smoker/non-smoker classification may lump together very heavy smokers with occasional smokers). Validation procedures may solve covariate misclassification issues,

but these biases may be subsumed by concerns about unmeasured or unmeasurable covariates that were discussed in subsection 4.2.1.

For the purposes of the Mini-Sentinel System Pre-screening Checklist, regulators should strongly consider which covariates are relevant to the tracked safety issue, and what is known about their classification in database systems.

4.3 Issues of Precision or Sample Size

Precision[293] – or the inverse of the variance - is a function of sample size, and an advantage of administrative datasets is the capacity for very large sample sizes. To establish greater precision, one may extend the study period or take other measures to increase the sample size. In section 6, I will use mathematical models to more finely address planning for sample size considerations in the Mini-Sentinel System, particularly with respect to sequential database surveillance. However, without engaging in more complex analyses, it is useful to simply consider whether exposures in the Mini-Sentinel System are likely to generate sample sizes that would meet a regulator's requirement for precision (i.e., the statistical power) when trying to establish postmarket evidence. Thus, I review circumstances in which sample size is difficult to achieve in the Mini-Sentinel System.

4.3.1 Route of Administration and Healthcare Setting of Exposure

To start, one should consider both the primary route of administration and healthcare setting in which the therapeutic is administered, and then determine if it is well captured in the Mini-Sentinel System. The Mini-Sentinel System is more efficient at capturing outpatient/"clinic" exposures and less efficient in capturing inpatient or emergency department exposures. That is, exposures in the latter two settings may not be recorded as individual line items, but aggregated as part of a diagnosis-related group code. It can be difficult to identify dates of exposure based on these data.

The Mini-Sentinel System is also more efficient at capturing self-administered exposures (e.g., oral medication or self-injectables) than at capturing infusions and injections primarily administered in the "clinic"/outpatient or inpatient setting. Infusions

[293] For a definition of precision, see *supra* at note 243.

and injections in these settings are treated as procedures, and may be recorded with Current Procedural Terminology (CPT) or Healthcare Common Procedure Coding System (HCPCS) codes. There are significant time lags associated with producing new codes to capture newly-approved infusions or injections, which are often the exposures of interest.[294] For reimbursement purposes, temporary codes are relied on but these temporary codes can create significant instability in data repurposed for secondary use.

In section 9, I analyzed data from the Mini-Sentinel System with respect to new molecular entities approved in the years 2004-2006. I eliminated certain new molecular entities from the dataset because they were typically delivered in an inpatient setting, and thus, generally less reliably captured in the Mini-Sentinel System. Of the 78 new molecular entities approved in those years, 37 were eliminated (i.e., 47%).

4.3.2 Low Exposure Prevalence Therapeutics

Also, by drawing on data held by commercial insurers, the Mini-Sentinel System is designed to be nationally representative and capture broad patterns in medical care. Even in a large system, orphan/rare diseases and their accompanying therapeutics will have low exposure prevalence that may be challenging to study. For example, in Mini-Sentinel System data accessed in support of this dissertation, orphan-designated drugs had less than 1500 new users over a 5-year period. One should consider whether exposure-based registries would be required to generate appropriate sample sizes for certain tracked safety issues.

Additionally, low exposure prevalence can result from a crowded market or availability of many substitute products, poor (i.e., high-tiered) placement in formularies resulting in high co-payments, or non-preferred status in clinical guidelines. In the Mini-Sentinel System data accessed for this dissertation, two drugs in the same class (i.e., competitors) – one with preferred status in clinical guidelines and one without such a status – had more than a tenfold difference in new users over a five year period.

---

[294] See generally American Medical Association, "CPT® Process - How a Code Becomes a Code", n.d., http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt/cpt-process-faq/code-becomes-cpt.page; Centers for Medicare and Medicaid Services, "Healthcare Common Procedural Coding System (HCPCS) Public Meetings", August 3, 2012, http://www.cms.gov/Medicare/Coding/MedHCPCSGenInfo/HCPCSPublicMeetings.html.

4.3.3 Medicare Populations

Medicare - the publicly-funded national insurance system for adults over 65, those with end-stage renal disease, and the permanently disabled - fundamentally changed on January 1, 2006. Effective that date, the Medicare Modernization Act of 2003 created Medicare Part D drug coverage plans, which allowed private insurers to offer prescription drug coverage benefits.[295] Eligible persons could obtain stand-alone prescription drug coverage or could be enrolled in Medicare Advantage plans with prescription drug benefits. Although many of the details of this transition are beyond the scope of this dissertation, the net result is that the Mini-Sentinel System is a less than ideal setting to evaluate exposure-outcome pairs that primarily affect Medicare populations because they may not have complete coverage (i.e., drug **and** medical benefits coverage) in a single data source.[296] New research shows this same potential in the elderly who are also veterans.[297] In general, it is likely that individuals without complete coverage would be excluded from pharmacoepidemiologic studies.

Additionally, the elderly that do have complete coverage under one insurer may be unique and not representative of the elderly population generally, leading to transportability issues, which will be discussed in subsection 4.4. The FDA's collaboration with other federal partners like the Centers for Medicare and Medicaid Services may alleviate some of these concerns. However, if an elderly person's coverage is scattered among several insurers, then complete capture is undoubtedly problematic.

4.3.4 Long Follow-up Times

If the tracked safety issue requires a long follow-up time (i.e., there is a long latency period before the event is biologically expected to occur following exposure), then the Mini-Sentinel System may have too much "churn" to allow for these long follow-up times, which contributes to small sample size and selection bias as discussed previously

---

[295] *Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Public Law 108-173*, 2003.
[296] For more details on Medicare Part D coverage patterns, see Gerald F. Riley, Jesse M. Levy, and Melissa A. Montgomery, "Adverse Selection In The Medicare Prescription Drug Program," *Health Affairs* 28, no. 6 (December 2009): 1826–37; Amy J. Davidoff et al., "Lessons Learned: Who Didn't Enroll In Medicare Drug Coverage In 2006, And Why?," *Health Affairs* 29, no. 6 (June 2010): 1255–63.
[297] See Amal N Trivedi et al., "Duplicate Federal Payments for Dual Enrollees in Medicare Advantage Plans and the Veterans Affairs Health Care System," *JAMA: The Journal of the American Medical Association* 308, no. 1 (July 4, 2012): 67–72.

in section 4.2.2.1. This problem may become exacerbated by the implementation of the Patient Protection and Affordable Care Act[298], which intends to increase competition among health insurers for enrollees. For surveillance purposes, increases in churn effectively reduce sample sizes.

4.4 Issues related to Transportability

Transportability describes how well the inference can be applied (i.e., transported) to other populations.[299] Recall that the Mini-Sentinel System is comprised of databases of commercially-insured persons. If a tracked safety issue affects an uninsured population, then an alternative evidence generation system may be necessary. The Mini-Sentinel System is generally composed of commercially insured individuals with medical and drug benefits coverage under the age of 65. Therefore, the transportability of results to other populations is uncertain.

4.5 Mini-Sentinel System Pre-Screening Checklist

In summary, every evidence generation system has strengths and weaknesses, and the particular circumstances of some tracked safety issues may favor some systems over others. Herein, I have sought to review circumstances that may be unfavorable to using the Mini-Sentinel System as an evidence generation system and compress these weaknesses into the Mini-Sentinel System Pre-screening Checklist in Table 2. This checklist is intended as a qualitative aid to prompt thoughtful analysis on whether the Mini-Sentinel System is likely insufficient *on its face* to evaluate a particular tracked safety issue. Key inputs include the tracked safety issue being investigated and the regulators' estimation of the strength of causal inference necessary to support regulatory decision-making.

If regulators proceed through this checklist without eliminating the Mini-Sentinel System as an evidence generation system, then I presume they have found it likely to be **sufficient**, per the statute[300], to resolve the tracked safety issue. The FDA may also wish to establish some similar procedures to decide whether its other postmarket evidence

---

[298] *Patient Protection and Affordable Care Act, P.L. 111-148.*
[299] For a definition of transportability, see *supra* at note 244.
[300] See *supra* at note 217.

generation systems – e.g., the spontaneous reporting systems - are also deemed sufficient, but I leave that to others.

MINI-SENTINEL SYSTEM PRE-SCREENING CHECKLIST
Bias Issues (General) • Is there a need to detect/rule out small effect sizes (i.e., relative risks less than 2)? To mitigate: • Is a self-controlled design possible?
Bias Issues (Confounding Bias) • Is there likely to be confounding by indication/confounding by contraindication/channeling bias? • Are known important confounders unmeasured or unmeasurable? To mitigate: • Is a self-controlled design possible? • Do instrumental variables (e.g., required formularies) exist? • Is medical chart confirmation in a subset of the study population feasible?
Bias Issues (Selection Bias) • Are changes in insurance coverage potentially associated with the exposure AND outcome of interest? • Are there other censoring conditions (e.g., death) that are associated with the exposure and outcome of interest? To mitigate: • Is *post hoc* analysis likely to account for selection biases? How sensitive are the results to such biases?
Bias Issues (Measurement Bias) • Do the exposures have predicted or known poor adherence patterns? • Are the exposures prescribed "as needed"? • Are the exposures not consistently included in commercial insurance formularies, or are there other conditions that would cause patients to purchase out-of-pocket? • Are the exposures ones for which prescription and over-the-counter equivalents exist? • Are the outcomes related to diagnoses that are evolving, ambiguous or uncertain? • Do the outcomes have an unclear onset? • What is the existing validation data regarding sensitivity and specificity of the algorithm in the Mini-Sentinel System? • Are the outcomes likely to be diagnosed in a variety of practice settings or by a variety of practitioners? • Are important covariates continuous variables? Are measures of these variables unsettled? To mitigate: • Is medical chart confirmation in a subset of the study population feasible?

Precision/Sample Size Issues • Is the exposure primarily expected to occur in an inpatient or emergency department setting? • Is the exposure via intravenous infusion or medically attended injection? To mitigate: • Has the intravenous infusion or medically attended injection been available for 18 months-2 years (i.e., has enough time elapsed that procedure codes should be stable)? • Is low exposure prevalence likely due to either the rareness of the disease being treated, the availability of substitute therapies, suboptimal insurance coverage (e.g., tiering), or non-preferred status with respect to clinical guidelines? • Is the population affected likely to have different insurers for medical benefits and drugs? Is the exposure in question likely covered under Medicare Part D? • Is a long follow-up time required (i.e., is there an induction period of several months or years)?
Generalizability/Transportability/External Validity • Is the tracked safety issue likely to substantially affect an uninsured or underinsured population?

Table 2. The Mini-Sentinel System Pre-Screening Checklist

If the Mini-Sentinel System is likely to be **sufficient** *on its face* to resolve the tracked safety issue, then I presume the FDA's next decision (shown as Decision 5 in Figure 5) regards whether to perform retrospective or prospective sequential analysis in the Mini-Sentinel System. This calculation is often a function of sample size at the time of initiation of the analysis, and that approximate sample size can be ascertained via modular programs as described in subsection 2.2.2.1. I focus the rest of this dissertation on prospective sequential methods because the conduct of protocol-based one-time assessments (described in subsection 2.2.2.3) is very similar to traditional retrospective pharmacoepidemiologic studies, which have been well-studied. However, the desire to avoid pooling data in the Mini-Sentinel System when performing these types of studies does pose some new challenges in this area.[301]

[301] See *supra* at note 142.

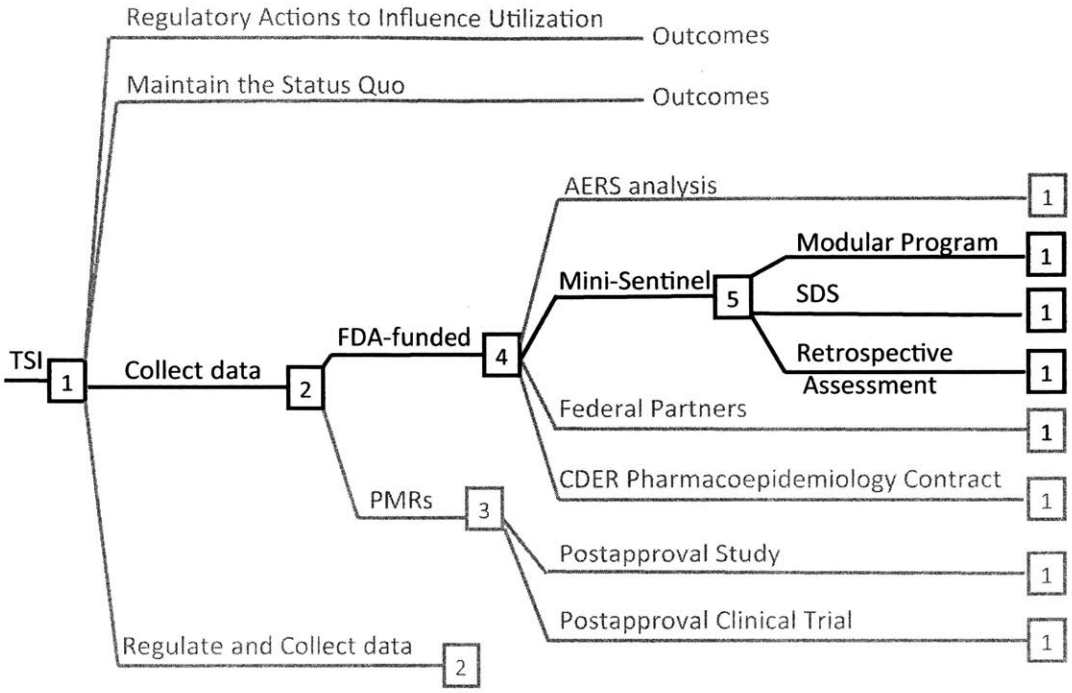

Figure 5. FDA Decision Process AFTER the Mini-Sentinel System is Deemed Sufficient
Abbreviations: TSI, tracked safety issue; FDA, Food and Drug Administration; PMR, postmarket requirement; CDER, Center for Drug Evaluation and Research; AERS, Adverse Event Reporting System; SDS, sequential database surveillance.

THIS PAGE INTENTIONALLY LEFT BLANK

5 PROSPECTIVE SEQUENTIAL DATABASE SURVEILLANCE

At this point in regulatory decision-making process, I assume that: 1) a tracked safety issue has been initially evaluated, 2) the Mini-Sentinel System has been deemed sufficient for regulatory decision-making using qualitative tools such as the Mini-Sentinel System Pre-screening Checklist, and 3) regulators are pursuing sequential prospective methods for the tracked safety issue (i.e., a pre-specified exposure-outcome pair) using the Mini-Sentinel System.

Briefly, I review the conduct of a prospective sequential database surveillance (SDS) evaluation. First, one prospectively gathers data from multiple databases (e.g., population-based health data) to monitor the incidence rate of an exposure-outcome pair under surveillance. One then compares the observed incidence rate to an expected rate, which is calculated based on either a concurrent-, historical-, or self-controlled group. Comparisons are made at regular intervals as data accrue using sequential statistical tests with pre-specified signaling thresholds. If the test statistic exceeds the threshold, then a statistical signal of excess risk is identified, the hypothesized exposure-outcome association is strengthened, and the null hypothesis of no excess risk is rejected. This signal is ordinarily followed by confirmatory assessments and review to validate or refute the finding. SDS analyses depend on amassing sufficient exposed person-time (i.e., sample size) to reach a stopping point, either by rejecting the null (i.e., detecting a safety signal) or ending surveillance (i.e., failing to signal).[302]

Beyond feasibility, another important consideration is the context of safety surveillance, and specifically, what is known, if anything, about the tracked safety issue. SDS may be undertaken in circumstances when there is little expectation that a safety problem exists, but surveillance is performed for reassurance. It may also be undertaken when data from spontaneous reporting systems or underpowered pre-licensure data suggest the possibility of a safety signal. These considerations are unique to the post-licensure environment when individuals outside of the observed population are affected by the speed and confidence (i.e., statistical power) with which a safety signal is detected or ruled out.

---

[302] J. C. Maro and J. S. Brown, "Impact of Exposure Accrual on Sequential Postmarket Evaluations: a Simulation Study," *Pharmacoepidemiology and Drug Safety* 20, no. 11 (2011): 1184–1191.

Of prospective sequential surveillance methods, the Food and Drug Administration (FDA) either can perform formal prospective SDS methods as described above or repeatedly execute modular programs[303] at set time intervals. Recall that Modular Program 3 is a computer program for safety signal assessment that can be used to calculate crude (i.e., mostly unadjusted) associational measures of medical product-associated adverse event rates. Data in Modular Program 3 can be subjected to *post hoc* statistical testing, but there is no ability to automatically control for multiple hypothesis tests and there are no "stopping boundaries." There is also no identified comparator group and very limited confounding control (i.e., stratification by age, sex, site only). I will return to a consideration of these two options after reviewing the state of the science with regard to prospective sequential surveillance methods next.

5.1 Epidemiological Design Considerations

Often, epidemiologists approach new research/surveillance questions by considering various epidemiologic designs along scientific, public health, ethical, and practical axes. With systems like the Mini-Sentinel System in mind, Gagne et al.[304] and others[305] have written extensively on considerations that should inform the epidemiologic design in an "active monitoring" context. Important factors include the exposure persistence (e.g., is it a continuing exposure like a statin or a brief one like an antibiotic), the onset and duration of the risk window period (e.g., is it biologically plausible that the exposure caused the outcome 1-10 days post-exposure or 30-365 days post-exposure), the strength of confounding, and the timing of the onset of the outcome of interest (e.g., short for allergic reactions, quite long for cancers and other chronic diseases that include undiagnosed subclinical activity).[306] All told, consideration of these factors leads epidemiologists to choose either between-person study designs or within-person study designs, and sometimes both are employed for reassurance.[307]

---

[303] Modular Programs are explained herein in 2.2.2.1.
[304] Gagne et al., "Design Considerations in an Active Medical Product Safety Monitoring System."
[305] D. L. McClure et al., "Comparison of Epidemiologic Methods for Active Surveillance of Vaccine Safety," *Vaccine* 26, no. 26 (2008): 3341–3345.
[306] Gagne et al., "Design Considerations in an Active Medical Product Safety Monitoring System."
[307] See, for example, a summary of vaccine safety studies in Yih et al., "Active Surveillance for Adverse Events: The Experience of the Vaccine Safety Datalink Project." Nearly all employ both designs and designate one as the primary.

Between-person study designs compare information contributed by different persons whereas within-person study designs compare information contributed by the same person during periods of exposure and non-exposure. Functionally, these designs dictate the choice of comparator group. In a between-person design, the comparator group can be a historical cohort of similarly situated individuals, or a concurrent cohort matched on either exposure (i.e., presence or absence of medical product use) or outcome status (i.e., presence or absence of adverse event of concern). Typically, these designs may sample data from the entire cohort of individuals (e.g., a full cohort study) or a smaller subset of the cohort (e.g., a case-control study).[308] Gagne et al. argue that case-control studies provide no meaningful gains in an administrative database setting when a pre-specified exposure-outcome pair exists.[309]

In a within-person design or self-controlled design, the comparator group is the treatment group, except the individuals are sampled during a different time period. In a self-controlled design, the individuals may be sampled based on their exposure status (e.g., a self-controlled risk interval design[310]) or outcome status (e.g., a self-controlled case series[311]). Self-controlled designs are a useful technique to mitigate unmeasured and unmeasureable confounding[312], but there are limited opportunities to employ such designs because of their many assumptions. As Gagne et al. point out:

> "When the key assumptions of self-controlled designs are fulfilled (i.e., lack of within-person, time-varying confounding; abrupt HOI [health outcome of interest] onset; and transient exposure), within-person comparisons are preferred because they inherently avoid confounding by fixed factors."[313]

5.2 Sequential Statistical Methods

It is important to note that the choice of epidemiologic design and comparator group limit the selection of available sequential statistical methods to analyze the data collected.

---

[308] For more on case-control studies and cohort studies in epidemiology, see Rothman, Greenland, and Lash, *Modern Epidemiology*, Chapter 7 and 8.
[309] Gagne et al., "Design Considerations in an Active Medical Product Safety Monitoring System."
[310] For examples of a self-controlled risk interval designs, see Alison Tse et al., "Signal Identification and Evaluation for Risk of Febrile Seizures in Children Following Trivalent Inactivated Influenza Vaccine in the Vaccine Safety Datalink Project, 2010-2011," *Vaccine* 30, no. 11 (March 2, 2012): 2024–2031; Lee et al., "H1N1 and Seasonal Influenza Vaccine Safety in the Vaccine Safety Datalink Project."
[311] For an example of a self-controlled case series design, see S. K. Greene et al., "Near Real-Time Surveillance for Influenza Vaccine Safety: Proof-of-Concept in the Vaccine Safety Datalink Project," *American Journal of Epidemiology* (2009).
[312] For more on confounding, see section 4.2.1.
[313] Gagne et al., "Design Considerations in an Active Medical Product Safety Monitoring System."

Sequential Statistical Models		Epidemiologic Designs: Within-Person Comparisons: Self Controls: Self-Controlled Case Series (matched on outcome)	Epidemiologic Designs: Within-Person Comparisons: Self Controls: Self-Controlled Risk Interval (matched on exposure)	Epidemiologic Designs: Between-Person Comparisons: Historical Controls: Cohort Design (matched on exposure)*	Epidemiologic Designs: Between-Person Comparisons: Concurrent Controls: Cohort Design (matched on exposure)*
Continuous	Poisson MaxSPRT			X	
Continuous	Conditional Poisson MaxSPRT			X	
Continuous	Binomial MaxSPRT**	X	X		X
Continuous	Exact Sequential Analysis**				X
Group Sequential	Group Sequential Likelihood Ratio Test**	X	X		X
Group Sequential	Conditional Sequential Sampling Procedure (CSSP)				X
Group Sequential	Propensity Score-Enabled CSSP				X
Group Sequential	Group Sequential Estimating Equations***			X	X
Group Sequential	Group Sequential Lan Demets***			X	X

Table 3. Compatibility of Epidemiologic Designs and Sequential Statistical Methods
*It is possible to use between-person comparisons and match on outcome. This would be a case-control design, which generally is not used in this context.[314]
**It is not applicable to a continuous exposure setting for which the number of exposed days will vary by each patient (e.g., most drugs). This is a quite limiting feature.
***It is generally not feasible when the outcome is rare or very rare.[315]
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test; CSSP, Conditional Sequential Sampling Procedure

While the choice of epidemiologic design and comparator group affect data collection (i.e., how we sample the database of individuals), the choice of sequential statistical

[314] Ibid.
[315] For definitions of rare or very rare outcomes, see The Council for International Organizations of Medical Sciences (CIOMS) Working Group III, *Guidelines for Preparing Core Clinical Safety Information on Drugs*.

method largely affects data analysis. Table 3 presents the compatibility of epidemiologic designs with sequential statistical methods, which are explained in detail in this section.

Generally, different methods can be compared with respect to their treatment of type I error and its distribution over multiple hypothesis tests, type II error/statistical power, the time to detect a signal (i.e., what I will later describe as the median sample size), and the maximum sample size. Sequential statistical methods also require that investigators set a stopping boundary, or a way to interrupt and end surveillance through signaling. The shape of this boundary dictates the likelihood of signaling at various interim hypothesis tests and determines some of the tradeoffs between power (i.e., sample size) and the timeliness of signal detection. Others[316] have reviewed sequential statistical methods that can be employed to perform SDS, which will be briefly summarized here.

A broad and important classification of the methods regards the frequency of multiple hypothesis testing. Continuous hypothesis testing methods perform hypothesis tests with the arrival of each observation. Group sequential testing methods specify the number of interim hypothesis tests based either on how information accrues in exposed-time or calendar-time increments. There are generally fewer interim tests than under a continuous hypothesis testing regimen. Group sequential clinical trials, which employ the same statistical models, commonly only allow 1-3 interim tests. In general, more frequent testing (i.e., continuous) performs better on timeliness by minimizing sample size, but does less well with regard to type I and type II errors.

Often, the choice of a continuous or group sequential statistical method is defined by logistic feasibility rather than epidemiologic choices. First, the frequency of testing should be conducive to the way in which data arrive. Specifically, hypothesis tests only need to be performed as often as new data are expected to arrive. Second, in order to keep with the assumptions of the underlying statistical models, one must understand how much

---

[316] A. J. Cook et al., "Statistical Approaches to Group Sequential Monitoring of Postmarket Safety Surveillance Data: Current State of the Art for Use in the Mini-Sentinel Pilot," *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 72–81; J. C. Nelson et al., "Challenges in the Design and Analysis of Sequentially Monitored Postmarket Safety Surveillance Evaluations Using Electronic Observational Health Care Data," *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 62–71; Jennifer C Nelson et al., "Methods for Observational Post-licensure Medical Product Safety Surveillance," *Statistical Methods in Medical Research* (December 2, 2011), http://www.ncbi.nlm.nih.gov/pubmed/22138688; Martin Kulldorff, "Sequential Statistical Methods for Prospective Postmarketing Safety Surveillance," in *Pharmacoepidemiology*, ed. Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy, Fifth. (John Wiley & Sons, 2011), 852–867.

new information arrives in each download of new data. For continuous sequential methods that perform a new hypothesis test on *each new observation,* if several observations were to arrive simultaneously, then there would be a mismatch between the statistical model assumptions and real-world surveillance. Consequently, as Cook et al. describe: "[the] use of continuously-designed [statistical] thresholds when testing is actually performed [less frequently] means that higher-than-necessary thresholds are used. This yields a type I error rate that is lower than desired and suboptimal statistical power."[317] Thus, the overall type I error threshold is set too conservatively. Further, "the magnitude of the conservatism increases with the amount of new data received between discrete testing points."[318]

5.2.1 Continuous Sequential Testing Methods – SPRT Adaptations

In the initial proof-of-principle analysis within the Vaccine Safety Datalink[319], statistical analysis was performed using Wald's Sequential Probability Ratio Test (SPRT)[320] to analyze vaccinated cohorts and compare them to historical data on a weekly basis.[321] The SPRT is a continuous hypothesis testing method as described above. Acceptance or rejection of the null hypothesis of no excess risk is dependent on the value of a test statistic, which is posed as a log likelihood ratio. The SPRT's stopping boundaries use a "flat threshold," meaning that the critical value for the test statistic is constant across all hypothesis tests. Although this threshold is mathematically simple and easy-to-understand, others have noted that a flat threshold may yield more false positives early in monitoring when there is less data, and lower power (i.e., higher type II error) at later points.[322] One way to overcome this limitation of a flat spending boundary while

---

[317] Cook et al., "Statistical Approaches to Group Sequential Monitoring of Postmarket Safety Surveillance Data: Current State of the Art for Use in the Mini-Sentinel Pilot."
[318] Nelson et al., "Methods for Observational Post-licensure Medical Product Safety Surveillance."
[319] The Vaccine Safety Datalink is a precursor to the Mini-Sentinel System and is described by Baggs et al., "The Vaccine Safety Datalink: a Model for Monitoring Immunization Safety."
[320] A. Wald, "Sequential Tests of Statistical Hypotheses," *The Annals of Mathematical Statistics* 16, no. 2 (1945): pp. 117–186.
[321] R. L. Davis et al., "Active Surveillance of Vaccine Safety: a System to Detect Early Signs of Adverse Events," *Epidemiology (Cambridge, Mass.)* 16, no. 3 (2005): 336–341.
[322] Nelson et al., "Challenges in the Design and Analysis of Sequentially Monitored Postmarket Safety Surveillance Evaluations Using Electronic Observational Health Care Data."

still performing continuous testing is to *delay* hypothesis testing until a certain number of events or exposures have accrued.[323] Such a delay prevents early signaling on little data.

Adjustment for confounding is done by stratifying the historical count data according to a limited set of covariates (e.g., age, sex, site), and then performing a logistic regression to calculate a risk-adjusted probability of the outcome in question.[324] The log likelihood ratio is the product of the likelihoods in the various strata. This approach limits the total number of confounders that can be considered to prevent multiple strata that are uninformative or too small.

Wald's SPRT requires a *simple* alternative hypothesis (e.g., a specific risk estimate such as an incidence rate ratio of 5). Kulldorff et al. have shown that the method is highly dependent on the selecting the correct alternative hypothesis (i.e., if an excess risk exists, then this presumption implies knowing the approximate value of that elevated risk, which is improbable).[325] Subsequent analyses in the Vaccine Safety Datalink utilized an adaptation of that method, Kulldorff's maximized sequential probability ratio test (MaxSPRT).[326] The MaxSPRT uses a composite alternative hypothesis as opposed to Wald's simple alternative. Initially, two variants of the MaxSPRT were posed: a binomial and a Poisson variant.

5.2.1.1 Binomial Variants

The binomial variant has been used for concurrent-[327] and self-controlled[328] analyses. Like Wald's SPRT, the binomial MaxSPRT model typically uses a flat threshold although Kulldorff indicates that it can support other types of sequential statistical

---

[323] Martin Kulldorff and Ivair Silva, "Continuous Sequential Analysis with Delayed Start", Unpublished Manuscript, 2012.
[324] S H Steiner, R J Cook, and V T Farewell, "Risk-adjusted Monitoring of Binary Surgical Outcomes," *Medical Decision Making: An International Journal of the Society for Medical Decision Making* 21, no. 3 (June 2001): 163–169.
[325] M. Kulldorff et al., "A Maximized Sequential Probability Ratio Test for Drug and Vaccine Safety Surveillance," *Seq Anal* 30, no. 1 (2011): 58–78.
[326] Ibid.
[327] T. A. Lieu et al., "Real-time Vaccine Safety Surveillance for the Early Detection of Adverse Events," *Medical Care* 45, no. 10 Supl 2 (2007): S89–95.
[328] Tse et al., "Signal Identification and Evaluation for Risk of Febrile Seizures in Children Following Trivalent Inactivated Influenza Vaccine in the Vaccine Safety Datalink Project, 2010-2011"; Lee et al., "H1N1 and Seasonal Influenza Vaccine Safety in the Vaccine Safety Datalink Project"; Greene et al., "Near Real-Time Surveillance for Influenza Vaccine Safety: Proof-of-Concept in the Vaccine Safety Datalink Project."

boundaries.[329] Confounding is controlled through matching. Early implementations of the binomial variant required a fixed matching ratio of the treatment group to the comparator group, which became difficult to implement in practice[330]. Specifically, investigators had to tradeoff more stringent matching criteria (i.e., better confounding control) against loss of information that resulted from the inability to find a match.

Subsequent improvements to the method relaxed this restriction and allowed the number of matched controls to vary by making use of an exact binomial test. The Exact Sequential Analysis[331], as it is now known, also can accommodate multiple sequential stopping boundaries. It is important to note that neither the binomial MaxSPRT model nor the Exact Sequential Analysis can be utilized with continuous exposures because of the inability to control the ratio of exposed/unexposed persons at each hypothesis test. Practically, it means that these tests cannot be used with most drugs and therapeutic biologics. As noted before, this is a major limitation in the Mini-Sentinel System.

5.2.1.2 Poisson Variants

The Poisson MaxSPRT variant is used with historical control groups, assumes a flat threshold, and assumes that the historical comparison rate is known (i.e., calculated from a large sample of the historical cohort). Confounding control is via stratification and then regression modeling as described above. The Poisson MaxSPRT model has been used in extensively in vaccine safety surveillance[332] and has been piloted for use in drug surveillance[333]. This sequential statistical test was later adapted into the conditional

---

[329] Kulldorff, "Sequential Statistical Methods for Prospective Postmarketing Safety Surveillance."
[330] Lieu et al., "Real-time Vaccine Safety Surveillance for the Early Detection of Adverse Events."
[331] J. Gee et al., "Monitoring the Safety of Quadrivalent Human Papillomavirus Vaccine: Findings from the Vaccine Safety Datalink," *Vaccine* 29, no. 46 (2011): 8279–8284.
[332] Lieu et al., "Real-time Vaccine Safety Surveillance for the Early Detection of Adverse Events"; E. A. Belongia et al., "Real-Time Surveillance to Assess Risk of Intussusception and Other Adverse Events After Pentavalent, Bovine-Derived Rotavirus Vaccine," *The Pediatric Infectious Disease Journal* 29, no. 1 (2010): 1–5; N. P. Klein et al., "Measles-mumps-rubella-varicella Combination Vaccine and the Risk of Febrile Seizures," *Pediatrics* 126, no. 1 (2010): e1–8; W. K. Yih et al., "An Assessment of the Safety of Adolescent and Adult Tetanus-diphtheria-acellular Pertussis (Tdap) Vaccine, Using Active Surveillance for Adverse Events in the Vaccine Safety Datalink," *Vaccine* 27, no. 32 (2009): 4257–4262; Gee et al., "Monitoring the Safety of Quadrivalent Human Papillomavirus Vaccine: Findings from the Vaccine Safety Datalink."
[333] J. S. Brown et al., "Early Adverse Drug Event Signal Detection Within Population-based Health Networks Using Sequential Methods: Key Methodologic Considerations," *Pharmacoepidemiology and Drug Safety* 18, no. 3 (2009): 226–234; J. S. Brown et al., "Early Detection of Adverse Drug Events Within

MaxSPRT (CMaxSPRT)[334] model to account for uncertainty in expected event rate in the historical comparison group. The CMaxSPRT model has been used for influenza vaccine surveillance.[335]

5.2.2 Group Sequential Testing Methods

Group sequential methods have a well-established statistical literature that was developed primarily in a randomized controlled trial setting.[336] Like continuous testing methods, these methods also allow a statistically valid way to stop a clinical trial early as a result of evidence of excess harm or demonstrated benefit. That is, it would be unethical to continue to allow the unexposed group to remain unexposed if the harmful or beneficial effect were proven statistically at an interim test. Typically, only a limited number of interim tests are performed in a clinical trial because of concerns related to loss of power. Additionally, group sequential clinical trials often were based on efficacy endpoints, and so approaches to deal with rare safety endpoints (e.g., Type B adverse reactions) are underdeveloped. Very recently, these techniques have been proposed and simulated in an observational safety surveillance context[337], although no actual surveillance activities have yet been completed with these methods.

5.2.2.1 Lan-Demets Group Sequential Approach

The Lan and Demets statistical model is widely used in clinical trials.[338] Lan and Demets developed a general statistical sequential boundary function for any

---

Population-based Health Networks: Application of Sequential Testing Methods," *Pharmacoepidemiology and Drug Safety* 16, no. 12 (2007): 1275–1284.

[334] L. Li and M. Kulldorff, "A Conditional Maximized Sequential Probability Ratio Test for Pharmacovigilance," *Statistics in Medicine* 29, no. 2 (2010): 284–295.

[335] Tse et al., "Signal Identification and Evaluation for Risk of Febrile Seizures in Children Following Trivalent Inactivated Influenza Vaccine in the Vaccine Safety Datalink Project, 2010-2011"; Lee et al., "H1N1 and Seasonal Influenza Vaccine Safety in the Vaccine Safety Datalink Project."

[336] For summaries, see John Whitehead, *The Design and Analysis of Sequential Clinical Trials* (Wiley, 1997); Christopher Jennison and Bruce W. Turnbull, *Group Sequential Methods with Applications to Clinical Trials* (Boca Raton: Chapman & Hall/CRC, 2000).

[337] Cook et al., "Statistical Approaches to Group Sequential Monitoring of Postmarket Safety Surveillance Data: Current State of the Art for Use in the Mini-Sentinel Pilot"; Nelson et al., "Methods for Observational Post-licensure Medical Product Safety Surveillance."

[338] D L DeMets and K K Lan, "Interim Analysis: The Alpha Spending Function Approach," *Statistics in Medicine* 13, no. 13–14 (July 15, 1994): 1341–1352; discussion 1353–1356; D L Demets, "Group Sequential Procedures: Calendar Versus Information Time," *Statistics in Medicine* 8, no. 10 (October

asymptotically normal test statistic, which means it performs well with frequent observations. However, in practice, for sequential surveillance with rare events, the asymptotic properties of the boundary fail to hold.[339] Consequently, such an approach has limited applicability in safety surveillance for Type B adverse events, but may still prove quite worthwhile for Type A or Type C adverse events.[340]

5.2.2.2 Group Sequential Likelihood Ratio Test

Cook et al. extended the binomial MaxSPRT model as described in subsection 5.2.1.1 and adapted it to accommodate situations in which multiple observations are likely with each arrival of new data.[341] Like the Lan-Demets method, this approach – the group sequential likelihood ratio test (GS LRT) - is potentially well adapted to Type A or Type C adverse events. As in other group sequential methods, a hypothesis test is not performed at the arrival of each new observation of an outcome, but rather based on the total exposure time accrued between hypothesis tests. Similar to the binomial MaxSPRT model, the GS LRT employs a fixed matching ratio, but accommodates multiple sequential stopping boundaries. However, its matching requirements create problems for use of the approach with continuous exposures because of the inability to keep the *same* ratio of person time contributed (i.e., exposed person-time/unexposed person-time) at each hypothesis test.

5.2.2.3 Conditional Sequential Sampling Procedure

The conditional sequential sampling procedure (CSSP)[342] is a group sequential method created to accommodate continuous exposures and concurrent comparison groups with active comparators as opposed to non-users. The choice of the sequential stopping boundary is flexible, and confounding control is performed via stratification on

---

1989): 1191–1198; K. K. Gordon Lan and David L. Demets, "Discrete Sequential Boundaries for Clinical Trials," *Biometrika* 70, no. 3 (December 1, 1983): 659–663.

[339] Cook et al., "Statistical Approaches to Group Sequential Monitoring of Postmarket Safety Surveillance Data: Current State of the Art for Use in the Mini-Sentinel Pilot."

[340] For more on Types A, B and C adverse reactions, see Meyboom, Lindquist, and Egberts, "An ABC of Drug-related Problems."

[341] Cook et al., "Statistical Approaches to Group Sequential Monitoring of Postmarket Safety Surveillance Data: Current State of the Art for Use in the Mini-Sentinel Pilot."

[342] L. Li, "A Conditional Sequential Sampling Procedure for Drug Safety Surveillance," *Statistics in Medicine* 28, no. 25 (2009): 3124–3138.

categorical confounders. The CSSP is efficient for evaluating rare events, but becomes less useful with very frequent testing or stratification on many confounders.[343] Subsequent improvements to the procedure include enabling it to handle propensity scores to enhance confounding control.[344]

5.2.2.4 Group Sequential Estimating Equations

Cook et al. developed the group sequential estimating equation approach, which is a regression-based group sequential test analyzed with a score test statistic.[345] It can handle continuous exposures and multiple continuous confounders, but it relies on significant information at the first analysis to estimate the parameters of the regression model. As noted by Nelson et al., approaches that delay hypothesis testing until a sufficient amount of information has accrued may solve these problems.[346]

Fireman et al. are using a regression-based approach with a Cox proportional hazards model in the pilot sequential database surveillance activity of the Mini-Sentinel System.[347] Confounding is being controlled through propensity-score matching.

5.3 Formal Sequential Database Surveillance Methods Compared to Modular Programs

Sequential database surveillance methods as discussed herein require specification of epidemiologic designs, comparator groups, and sequential statistical models, whereas repeated execution of modular programs does not require such specification. Modular program outputs are stratified incidence rates, which when compared, can produce mostly unadjusted associational measures. If desired, these measures can be subject to hypothesis testing although there are no formal means to control for multiple hypothesis

[343] Cook et al., "Statistical Approaches to Group Sequential Monitoring of Postmarket Safety Surveillance Data: Current State of the Art for Use in the Mini-Sentinel Pilot."
[344] Lingling Li et al., "A Propensity Score-Enhanced Sequential Analytic Method for Comparative Drug Safety Surveillance," *Statistics in Biosciences* 3, no. 1 (2011): 45–62.
[345] Cook et al., "Statistical Approaches to Group Sequential Monitoring of Postmarket Safety Surveillance Data: Current State of the Art for Use in the Mini-Sentinel Pilot."
[346] Nelson et al., "Challenges in the Design and Analysis of Sequentially Monitored Postmarket Safety Surveillance Evaluations Using Electronic Observational Health Care Data."
[347] Fireman et al., "A Protocol for Active Surveillance of Acute Myocardial Infarction in Association with the Use of a New Antidiabetic Pharmaceutical Agent."

tests. When might one prefer repeated execution of modular programs to a formal sequential statistical method?

A key determination for each tracked safety issue being considered is twofold: whether there is enough information to suggest a clinically relevant control group and *whether more precise quantification of the safety signal relative to that control group is necessary*. The first point speaks to the case when little is known about what patients might use as a substitute product, or if a suitable substitute product even exists. This situation may present itself for medical products with potentially diverse patient populations, or medical products that are monitored soon after the time of approval. In these circumstances, undertaking prospective sequential database surveillance may be premature.

Second, circumstances in which further quantification may be unnecessary include 1) when the background rate of the outcome occurring spontaneously in the clinical population of concern is very rare or near zero, 2) when the outcome of concern is biologically and temporally clearly drug-induced. In such cases, little is likely to be gained from quantification using formal sequential prospective surveillance methods. In the past, these tracked safety issues were resolved with evidence from case series analyses.[348]

5.4 Lessons Learned

I now turn to lessons learned in prospective sequential database surveillance. An early concern regarding these studies was the potential for false positive signals that might overwhelm regulators and manufacturers performing post-signal investigations. Yih et al.[349] report on the early Vaccine Safety Datalink experience investigating thirty designated vaccine-outcome combinations, which resulted in ten statistical signals of excess risk while performing SDS analyses. However, following further investigation, only *one* of the initial statistical signals was confirmed to be a true association, and thus a true signal of a serious risk. I have already discussed the sources of bias in observational studies that may lead to false positive conclusions. In general, nearly all of the false

[348] Dal Pan, Lindquist, and Gelperin, "Postmarketing Spontaneous Pharmacovigilance Reporting Systems."
[349] W. K. Yih et al., "Active Surveillance for Adverse Events: The Experience of the Vaccine Safety Datalink Project," *Pediatrics* 127 Suppl 1 (May, 2011), S54.

signals were due to confounding bias or measurement bias, while a few appeared to be transient early signaling which has since been corrected with improved statistical designs.

Another concern that has emerged is the potential for prolonged surveillance activities resulting from a lack of adoption of the medical product. In these instances, there is literally an inability to reach a stopping point resulting from the slow arrival of new information. An analogous concern in prospective postmarket clinical trials is under-enrollment, which results in underpowered trials. Generally, sequential analyses require managing two time-scales: calendar time and information time. Attention to calendar time is important because of potential excess harms to the population that may result from delayed detection of a safety problem. Attention to information time – how sample size is accrued – is important for statistical considerations. In that sense, the balance between timeliness and sample size is more challenging in prospective sequential database surveillance than in sequential clinical trials because the need to minimize calendar time while resolving a tracked safety issue has a greater priority.

For these reasons, decisions to proceed with prospective sequential database surveillance will, in part, hinge on whether accrual of information (i.e., sample size) occurs within a calendar time frame appropriate for regulatory decision-making. To wit, the advantage of such surveillance occurs when information accumulates rapidly enough to provide regulators an earlier opportunity for regulatory intervention than would be possible via another evidence generation system.

Next, I develop a tool – the Sequential Database Surveillance Simulator - to allow regulators or public health investigators to explore quantitative assessments of sufficiency of the Mini-Sentinel System under varying conditions of uncertainty. The simulator is intended to be a learning tool and quantitative aid to decision-makers that must simultaneously manage information time and calendar time while performing prospective sequential database studies. The simulator allows regulators/investigators to explore the many potential surveillance scenarios they could face. With such modeling and simulation tools, regulators may more precisely deploy evidence generation systems like the Mini-Sentinel System, and further refine their assessments of its sufficiency for evidence generation.

THIS PAGE INTENTIONALLY LEFT BLANK

6 SEQUENTIAL DATABASE SURVEILLANCE SIMULATOR

At this point in the Food and Drug Administration (FDA)'s decision algorithm, I assume that the FDA is proceeding with sequential database surveillance and is seeking to refine its understanding of the potential surveillance scenarios that it may encounter for a particular tracked safety issue of interest. In systems like the Mini-Sentinel System, sequential database surveillance is performed in an observational setting where the regulator/public health investigator has no control over information accrual. However, in light of new safety information suggesting medical product-associated harm, the efficient use of the Mini-Sentinel System requires a more refined understanding of whether accrual of information (i.e., sample size) occurs within a calendar time frame appropriate for regulatory decision-making. In other words, evaluating the Mini-Sentinel System's evidence generation capabilities necessitates the translation of information time into calendar time.

In this section and the following two sections, I perform this translation by modeling and then simulating sequential database surveillance scenarios in the Mini-Sentinel System via the Sequential Database Surveillance Simulator. I developed this quantitative tool with the intention to aid the regulator/public health investigator in the initial planning stages of surveillance when pursuing a tracked safety issue of interest. By demonstrating how surveillance may unfold given various sets of initial circumstances, the tool is designed to allow the regulator/public health investigator to explore the performance limitations and capabilities of sequential database surveillance *virtually* and in a *low-cost* way. That is, in this planning stage, there is no need to "learn-by-doing" while expending public health resources. In general, this tool is not intended to be strictly predictive or to forecast exactly how sequential database surveillance of a particular tracked safety issue will occur. Rather, it more akin to a "management flight simulator,"[350] which allows the regulator/public health investigator to explore different potential paths to manage sequential database surveillance activities.

---

[350] Sterman has pioneered the development of management flight simulators for business operations and described them thusly, "Virtual worlds for learning and training are commonplace in the military, in pilot training, in power plant operations, and in many other real time tasks where human operators interact with complex technical systems." See John Sterman, *Business Dynamics: Systems Thinking and Modeling for a Complex World* (Boston: Irwin/McGraw-Hill, 2000), 35.

The Sequential Database Surveillance Simulator is compromised of three interlinked sub-models: the information time sub-model, the calendar time sub-model, and the analysis sub-model. I walk through an explanation of these sub-models in this section. I presume the user of the Sequential Database Surveillance Simulator is a regulator or public health investigator, but I forthwith refer to the "user" to be more general.

As a note, the current version of the Sequential Database Surveillance Simulator accommodates two specific sequential statistical models that have been frequently used in prior vaccine safety surveillance.[351] I began this process using these models because they are well-established in this still developing field. However, this simulator could be built out to accommodate other models, such as the group sequential models reviewed in subsection 5.2.2. An important aspect of future work will be to increase representation of group sequential models to better under their comparative performance characteristics.

6.1 Information Time Sub-Model: Sample Size Calculations

6.1.1 Information Time Sub-Model Inputs

The user begins with the following inputs to support the information time sub-model: 1) the sequential statistical model to be used for analysis; and 2) the expected incidence rate of the outcome of interest in the comparison population, i.e., the background rate. First, the sequential statistical model specifies the type of hypothesis testing that will be performed (e.g., continuous or group sequential), and what types of sequential statistical boundaries can be accommodated. Additionally, each sequential statistical model has its own set of parameters that need specification. That is, each model uses particular test statistics, assigns some quantities as known and others as random, etc. Second, the background rate is the incidence rate of the outcome expected under the null hypothesis, i.e. that there is no excess risk in the treatment group.

The information time sub-model helps the user to perform sample size calculations considering a range of effect sizes[352], and to get a general idea of how much information is required to make a particular finding. Typical sample size calculations for non-

[351] Yih et al., "Active Surveillance for Adverse Events: The Experience of the Vaccine Safety Datalink Project."
[352] For a definition of effect size, refer to *supra* at note 237.

sequential statistical methods require investigators to calculate the relationship between the pre-specified upper limit for accepting false positive results (i.e., type I error), the statistical power to detect a particular effect size (i.e., type II error), and the sample size. Sample size calculations for sequential statistical models are similar except they incorporate the ability to interrupt surveillance by rejecting the null hypothesis at a point *earlier* than the prescribed end of surveillance. Thus, in sequential database surveillance, there are *two* sample sizes to consider: one is the sample size needed to reject the null hypothesis (i.e., the time to detect a signal) and the other is the maximum sample size (i.e., the maximum length of surveillance). More modest (i.e., numerically smaller) true effect sizes require larger samples. Additionally, sequential statistical models require specification of the sequential stopping boundary[353], which relates to how type I error is apportioned among multiple hypothesis tests and also limits the statistical power that can be achieved at any interim testing point. Therefore, one must consider six variables in the sample size calculation: (a) the sequential stopping boundary; (b) the overall type I error across all hypothesis tests; (c) the true effect size; (d) the statistical power; (e) the maximum sample size; and (f) the time to detect a safety signal. These last three quantities are random variables that are easiest to understand with a brief example that will also allow the reader to observe sequential database surveillance scenarios.

Let us use the Mini-Sentinel System pilot project as our tracked safety issue of interest. In the pilot project, investigators wished to detect a 1.33 incidence rate ratio of acute myocardial infarctions among new users of saxagliptin as compared with new users of other oral anti-diabetic agents with 80% power and 5% overall type I error.[354] The investigators specified the background incidence rate in the comparator group to be nine acute myocardial infarctions per 1000 person years among diabetics. While the investigators ultimately chose a Cox proportional hazards model with 10 interim tests (i.e., a group sequential statistical model), let us explore how those same parameters would play out in the two continuous sequential statistical models supported in the simulator: the Poisson Maximized Sequential Probability Ratio Test (MaxSPRT) and binomial MaxSPRT models.

---

[353] For a discussion on sequential statistical boundary types, see subsection 5.2

[354] Fireman et al., "A Protocol for Active Surveillance of Acute Myocardial Infarction in Association with the Use of a New Antidiabetic Pharmaceutical Agent."

6.1.2 Poisson MaxSPRT Model

Recall that a Poisson MaxSPRT model supports a cohort design with a historical comparator group, and that the historical comparator group is defined by the incidence rate of the outcome of interest expected in that group (i.e., the background rate). I simulate data on treatment and comparator populations in accordance with an assigned true effect size, and then analyze these data with the assigned sequential statistical model (i.e., the Poisson MaxSPRT).

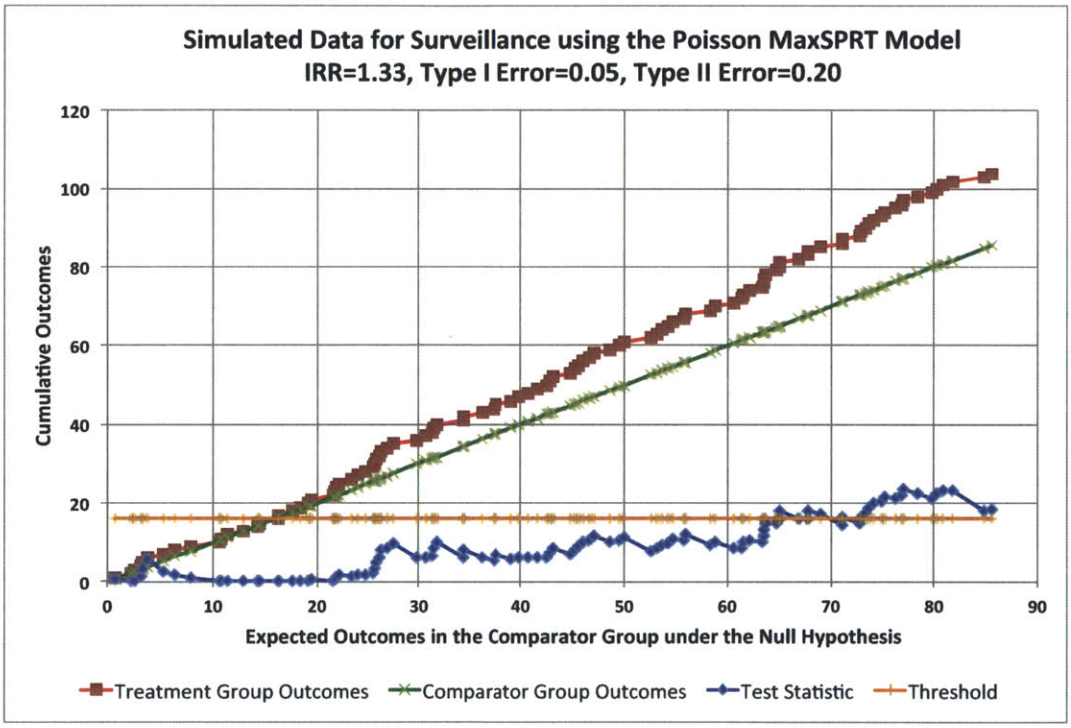

Figure 6. Example Sequential Database Surveillance Scenario Analyzed using the Poisson MaxSPRT Model
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test. IRR, incidence rate ratio.

First, I simulate an effect size equivalent to a 1.33 incidence rate ratio, and power my surveillance to limit overall type I error to 0.05, and type II error to 0.20 (i.e., statistical power is 0.80).[355] I use a flat sequential stopping boundary consistent with the default for

[355] In the Poisson MaxSPRT model, one must specify the upper limit on the number of outcomes expected in the comparator group under the null hypothesis. This value represents the maximum sample size, which will be explained in greater detail in this section. Practically, it is more sensical to set this upper limit based on the statistical power that is desired, which in this case is 0.80, and then calculate the upper limit from that threshold. The upper limit that satisfies the condition of at least 80% statistical power to detect a 1.33 incidence rate ratio is 111.75 outcomes.

the MaxSPRT models, and similar to the sequential stopping boundary chosen for the Cox proportional hazards model in the Mini-Sentinel System pilot.[356] In other words, the null hypothesis is rejected when the test statistic exceeds a set threshold that does not change over the course of surveillance (i.e., the "flat" characterization, which is easy to see by observing the shape of the "Threshold" line in Figure 6).

One instantiation of this simulation produces the scenario depicted in Figure 6. In this particular instantiation, the number of outcomes in the treatment group separates enough from the comparator group when the test statistic crosses the threshold value at 65 outcomes, and the sequential statistical model correctly detects a signal of excess risk. Using simple algebra and the background rate of 9 acute myocardial infarctions per 1000 person-years, 65 outcomes are equivalent to 7,222 person-years. Then, I repeat this analysis 100,000 times, and display the results as a distribution in Figure 7.

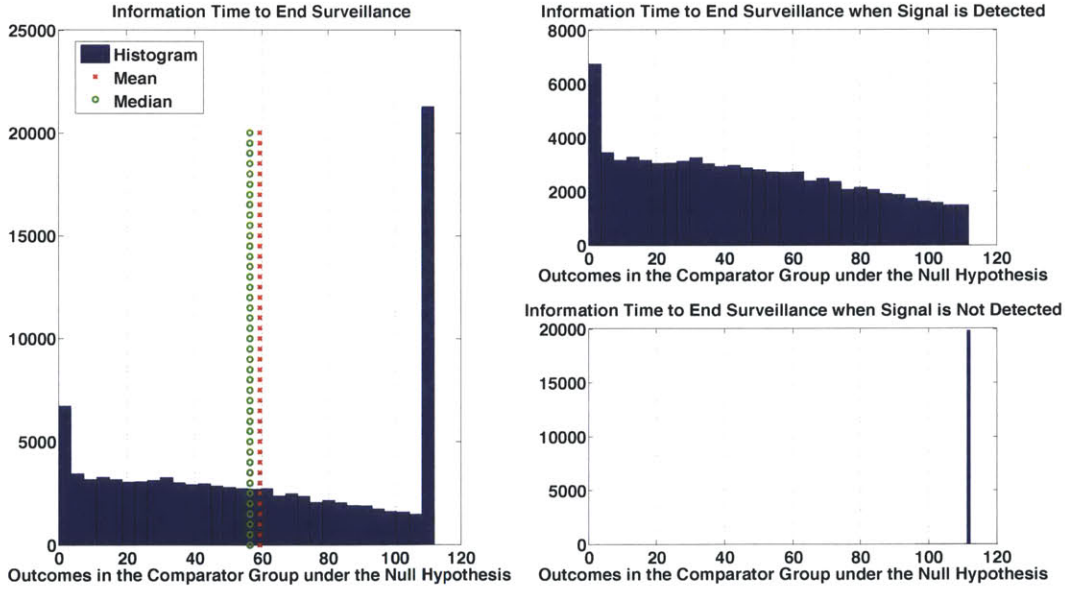

Figure 7. Distribution of the Information Time to End Surveillance for 100,000 Simulations Analyzed using the Poisson MaxSPRT Model
Parameters: Incidence Rate Ratio=1.33, overall type I error=0.05, statistical power=0.80. The left panel represents the overall (i.e., unconditional) distribution. The right upper panel represents the conditional distribution when a signal was (correctly) detected. The lower right panel represents the conditional distribution when a signal (incorrectly) failed to be detected.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test.

[356] Fireman et al., "A Protocol for Active Surveillance of Acute Myocardial Infarction in Association with the Use of a New Antidiabetic Pharmaceutical Agent."

In Figure 7, the left panel is a distribution of the information time required to end surveillance, and includes the full 100,000 instances of simulation. On the right panel, I break this histogram into two circumstances: the upper portion represents the instances when a signal was detected correctly (i.e., the null hypothesis was rejected) and the lower portion represents the instances when a signal was missed (i.e., the null hypothesis incorrectly failed to be rejected). These distributions provide us with three important values.

First, the statistical power is simply the percentage of time that the statistical model correctly rejects the null hypothesis of no excess risk (i.e., the total frequency counts in the upper right panel of Figure 7). In this case, the surveillance was powered to limit type II error to ≤ 0.20 when the effect size was equal to an incidence rate ratio of 1.33. Second, the maximum sample size, which represents the stopping boundary when one fails to reject the null, takes on one value when using the Poisson MaxSPRT model (i.e., shown in the lower right panel of Figure 7) and is a consequence of having reached the maximum expected number of outcomes under the null hypothesis. In this case, the maximum number of outcomes is 111.75, which means that the maximum sample size is 12,416.67 person-years. Last, the information time required to detect a safety signal (i.e., the other stopping boundary) is represented by the median of the information time until surveillance ends (i.e., the median of the left panel of Figure 7), irrespective of whether a signal is detected. In this case, the median is 56.4 outcomes, which translates into 6,269 person-years.

I choose the unconditional median as a summary statistic to represent the distribution in the left panel of Figure 7 for several reasons. First, the unconditional median assures an accurate comparison of the time-to-signal for different systems with different statistical power. I only use this statistic when the statistical power is at least 50%, ensuring the median reflects a time when a signal was detected. If I had used the median *conditioned* on when a signal was detected (i.e., the median of the upper portion of the right panel of Figure 7), then the amount of information that contributes to the distribution varies based on the statistical power, which generates misleading comparisons. Second, the unconditional mean is clearly influenced by outliers that become more prominent as statistical power decreases. Henceforth, I refer to this

summary statistic as the *median sample size*, which is analogous to the average or expected sample size in the group sequential trials literature.[357] Generally, smaller median sample sizes are preferred because less information is needed to detect a safety problem.

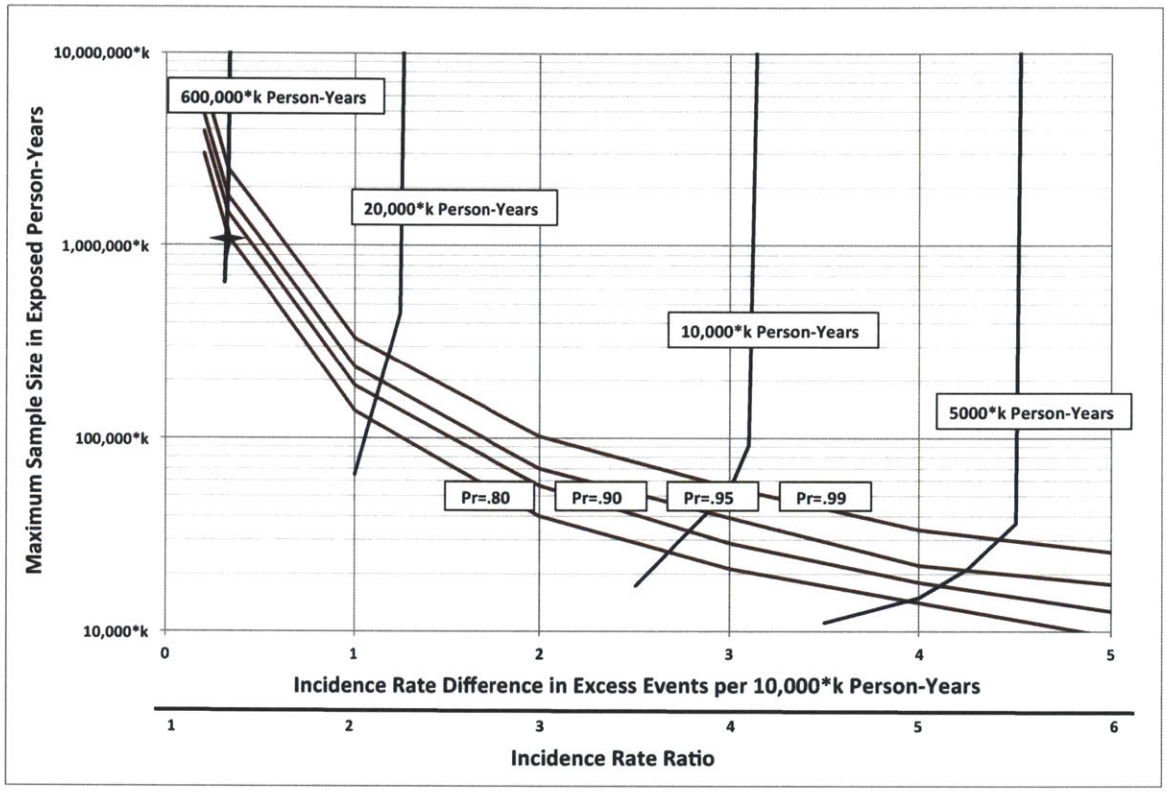

Figure 8. Relationship between Statistical Power, Median Sample Size, Maximum Sample Size, and True Effect Size Analyzed using the Poisson MaxSPRT Model
Statistical power isolines travel from northwest to southeast. Median sample size isolines travel from southwest to northeast. Overall type I error set to 0.05.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test; Pr, power.

Let us return to the relationships among the six variables in the sample size calculation: (a) the sequential stopping boundary; (b) the overall type I error across all hypothesis tests; (c) the true effect size; (d) the statistical power; (e) the maximum sample size; and (f) the median sample size. To evaluate the relationships among the six quantities described above, I perform the same simulations that allowed the creation of

[357] Demets, "Group Sequential Procedures"; Scott S Emerson, John M Kittelson, and Daniel L Gillen, "Frequentist Evaluation of Group Sequential Clinical Trial Designs," *Statistics in Medicine* 26, no. 28 (December 10, 2007): 5047–5080; J M Kittelson and S S Emerson, "A Unifying Family of Group Sequential Test Designs," *Biometrics* 55, no. 3 (September 1999): 874–882.

Figure 7 and I repeat this process for a range of effect sizes of interest (e.g., incidence rate ratios of 1.2 to 10). For each effect size, I use the same number of simulations: 100,000. I display an abbreviated version of my results compactly in Figure 8.

In Figure 8, the independent variables – in the *mathematical*, not the *statistical* sense - are the sequential stopping boundary, the overall type I error, the true effect size, and the maximum sample size. The flat sequential stopping boundary and the overall pre-specified type I error (0.05) are constant throughout Figure 8. The other two independent variables – the true effect sizes and the maximum sample sizes - do vary over the figure. The true effect sizes, illustrated along the x-axis, are given in two scales. The upper scale is an absolute risk measure. The lower scale is defined using the equivalent relative risk measure. The y-axis is the maximum sample size, i.e., the stopping boundary for surveillance when one fails to reject the null. The maximum sample size is shown in person-years (as opposed to expected outcomes in the comparator group under the null hypothesis) to be more descriptive and explanatory to the user.[358] In summary, these four independent variables are fixed to evaluate their effect on statistical power and median sample size.

The dependent variables – in the *mathematical*, not the *statistical* sense – are the statistical power and the median sample size. In Figure 8, the effects of the independent variables on the dependent variables are depicted in two sets of isolines. Statistical power is depicted in the first set of isolines that travels from northwest to southeast. These isolines are downward sloping because the same statistical power can be attained with smaller sample sizes when greater true effect sizes exist. Statistical power is higher as the maximum sample size increases because there are more opportunities to detect a signal. Median sample size is depicted in the second set of isolines that travels from southwest to northeast. For a given effect size, there are minimal increases in the median sample size by increasing the maximum sample size. Vertically asymptotic behavior dominates as statistical power approaches unity. The values of the median sample size isolines become smaller as the true effect size increases because smaller sample sizes will signal under conditions of greater risk.

[358] Note that this formulation necessitates specification of the incidence rate expected under the null hypothesis. However, both expected outcomes in the comparator group under the null hypothesis and their equivalent person-years are information time measures.

These figures are shown with an arbitrarily chosen reference outcome of interest of 1/10,000 person-years. Therefore, to use the same figures universally, the user must calculate an appropriate "*k*" scaling constant to account for differences between the problem-specific outcome frequency and the reference outcome frequency (i.e., when k=1, the problem-specific outcome occurs in the control group at 1 outcome/10,000 person-years). A user simply should divide the reference outcome by the user-specified outcome to generate the *k* value.

Briefly, I walk through how to read Figure 8 and again employ the example used in the Mini-Sentinel pilot project. Recall that the intent was to detect an incidence rate ratio of 1.33 with 80% power. Figure 8 is marked with a star at this point. However, to interpret the maximum and median sample sizes, it is necessary to first calculate the *k* scaling constant by dividing the reference outcome (i.e., 1 outcome/10,000 person-years) by the problem-specific outcome of interest (i.e., 9 outcomes/1000 person-years), yielding a *k* scaling constant of 0.0111. As the reader would expect from the previously simulated distributions shown in Figure 7, the maximum sample size is 12,416.67 person-years (~1.1M*k=.0111) and the median sample size is 6,269 person-years (~600,000*k=.0111).

The objective of Figure 8 is to show the user the realm of other possibilities for surveillance, and to allow them to draw comparisons. Depending on whether the maximum sample size and median sample size are believed to be feasible given the data available, the user can employ this figure to gain an understanding of the tradeoffs that occur by moving away from the baseline scenario (i.e., the starred point). For example, if the user believes that these baseline sample sizes are easily attainable, then one might consider detecting a 1.33 incidence rate ratio with higher power (e.g., 90%). That northward jump in the statistical power isoline would increase maximum sample size but might only modestly increase median sample size because of the verticalness of the median sample size isolines. Practically, that means there may be a small price to pay regarding losses in the timeliness of signal detection when seeking higher statistical power. On the other hand, if the user believes these baseline sample sizes are not attainable, then the user can consider reducing power, or detecting a numerically larger effect size (e.g., an incidence rate ratio of 2). Numerically larger effect sizes tend to

accompany less vertical portions of the median sample size isolines indicating real tradeoffs start to occur between statistical power and the timeliness of signal detection.

Now, let us investigate the same scenario with the other supported model in the simulator: the binomial MaxSPRT model. [359]

6.1.3 Binomial MaxSPRT Model

Recall that the binomial MaxSPRT model supports concurrent and self-controlled comparator groups, but requires the user to specify the matching ratio between the treatment group and comparator group. Using a 1:1 matching ratio, I simulate data on the treatment and comparator populations in accordance with an assigned true effect size, and then analyze these data with the binomial MaxSPRT model using a flat sequential stopping boundary. Again, I first simulate an effect size equivalent to a 1.33 incidence rate ratio, and power my surveillance to limit overall type I error to 0.05, and type II error to 0.20 (i.e., statistical power is 0.80).

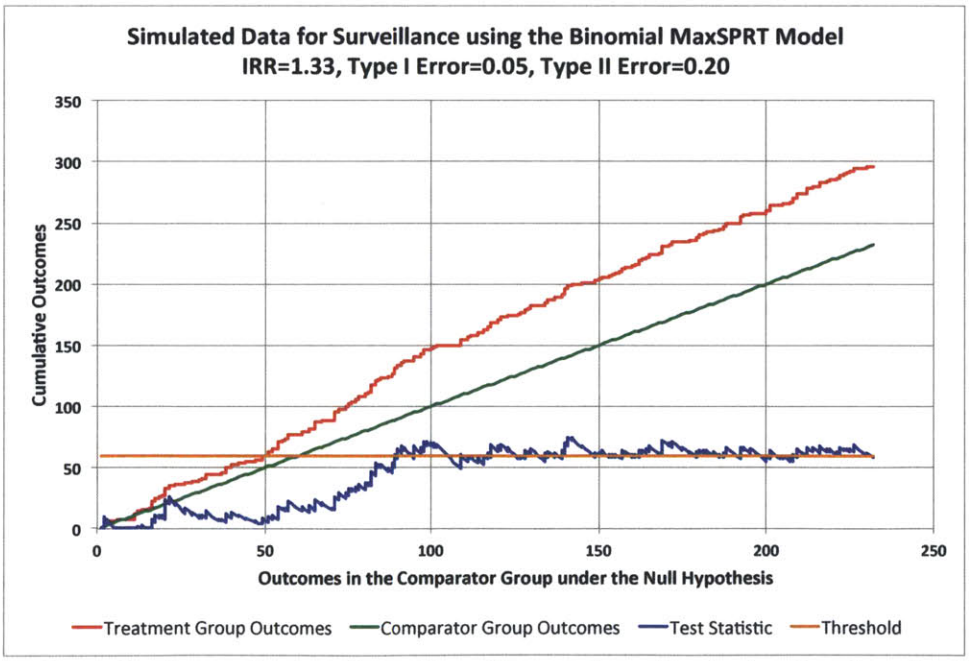

Figure 9. Example Sequential Database Surveillance Scenario Analyzed using the Binomial MaxSPRT Model
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test; IRR, incidence rate ratio.

[359] I show the binomial MaxSPRT model to familiarize the reader with this model since I will use it extensively and the simulator supports it. However, the binomial MaxSPRT model would most likely have been ruled out for use in the Mini-Sentinel System pilot project because it does not accommodate continuous exposures. See subsection 5.2.1.1.

As before in Figure 9, I show one instantiation of the sequential database surveillance scenario. In this particular instantiation, the number of outcomes in the treatment group separates enough from the comparator group when the test statistic crosses the flat threshold value at 89 outcomes (or 9,888 person-years), detecting a signal of excess risk. Then, I repeat this analysis 100,000 times, and display the results as a distribution in Figure 10.

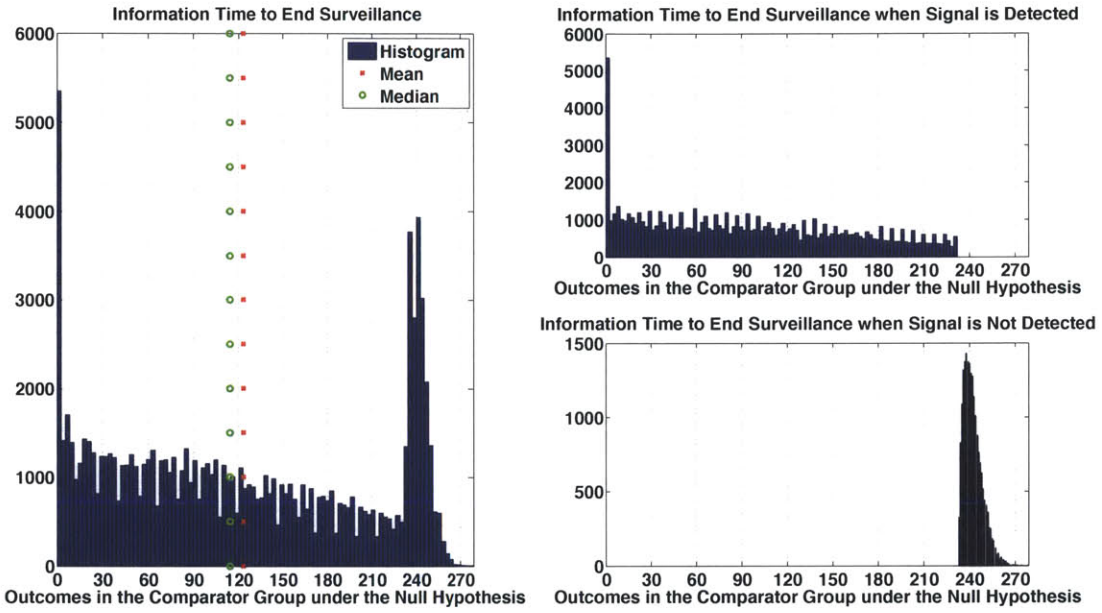

Figure 10. Distribution of the Information Time to End Surveillance for 100,000 Simulations Analyzed Using the Binomial MaxSPRT Model
Parameters: Incidence Rate Ratio=1.33, overall type I error=0.05, statistical power=0.80, matching ratio=1:1. The left panel represents the overall (i.e., unconditional) distribution. The right upper panel represents the conditional distribution when a signal was (correctly) detected. The lower right panel represents the conditional distribution when a signal was (incorrectly) not detected.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test.

Again, as in Figure 7 and here in Figure 10, the left panel is a distribution of the information time required to end surveillance, and includes the full 100,000 instances of simulation. On the right panel, I break this histogram into two circumstances: the upper portion represents the instances when a signal was correctly detected (i.e., the null hypothesis was rejected) and the lower portion represents the instances when a signal was missed (i.e., the null hypothesis incorrectly failed to be rejected). As before, I use the unconditional median of the left panel to describe the median sample size: 114 outcomes or 12,666 person-years. However, note that in the lower right panel of Figure 10, the

maximum sample size is now a *distribution* as opposed to a singular value as it was in the Poisson MaxSPRT model. If I chose the median of this distribution, which is conditioned on failing to detect a signal, then I will be inappropriately comparing systems with different power. Therefore, to appropriately represent the maximum sample size, I re-perform the simulations again, except that I set the effect size to be equal to the effect size assumed in the null hypothesis (i.e., in other words, I create a situation when the null hypothesis is true). These results are shown in Figure 11.

In Figure 11, the unconditional median of the leftmost panel is 265 outcomes or 29,444.4 person-years.[360] In the upper right panel of this figure, one sees instances in which the null hypothesis is rejected inappropriately (i.e., false positives occur in accordance with the set type I error margin).

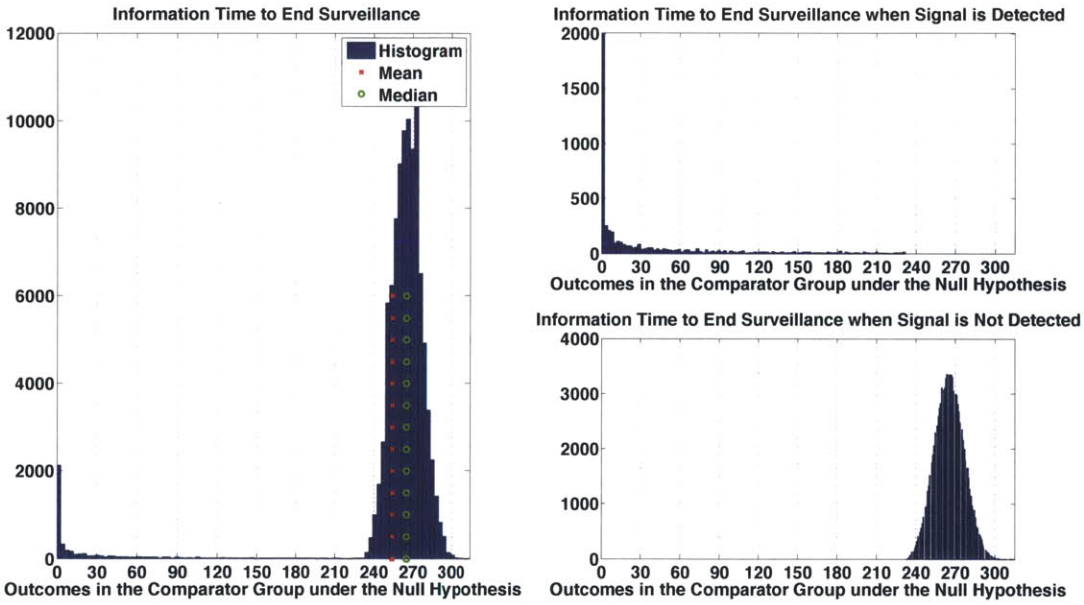

Figure 11. Distribution of the Information Time to End Surveillance for 100,000 Simulations Analyzed Using the Binomial MaxSPRT Model
Parameters: Incidence Rate Ratio=1, overall type I error=0.05, statistical power=0.80, matching ratio=1:1. The left panel represents the overall (i.e., unconditional) distribution. The right upper panel represents the conditional distribution when a signal was (incorrectly) detected. The lower right panel represents the conditional distribution when a signal was (correctly) not detected.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test.

[360] In Figure 11, I could also use the mean or median conditioned on the times when a signal is not detected. The amount of contributing information would be the same as long as the number of simulations and the overall type I error was kept the same. However, for consistency, I use the unconditional median.

Again, I repeat this process for a range of effect sizes of interest, holding the number of simulations constant at 100,000. I display an abbreviated version of my results compactly in Figure 12, which again contains a starred mark for the intersection of an incidence rate ratio=1.33 and statistical power=0.80, the intended effect size to be detected in the Mini-Sentinel System pilot project. As was evident from Figure 10, and again is shown here in Figure 12, the median sample size is 12,666 person-years (~1.14M*k=0.0111). Similarly, as was shown in Figure 11 and again here in Figure 12, the maximum sample size is 29,444.4 person-years (~2.65M* k=0.0111).

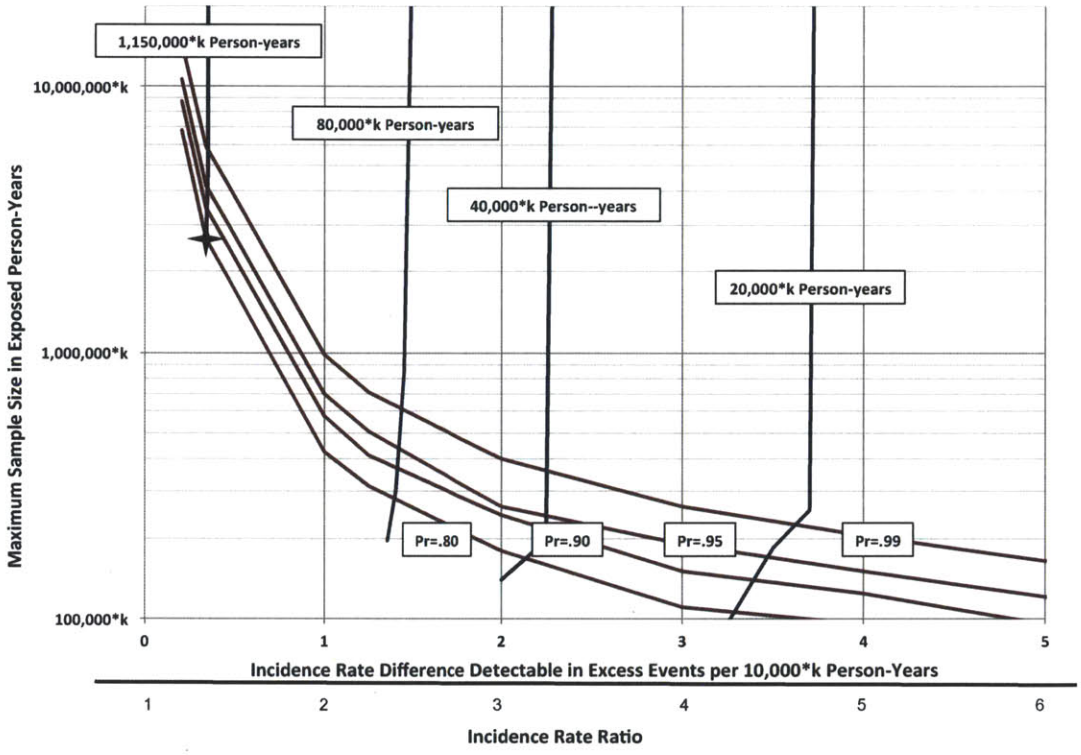

Figure 12. Relationship between Statistical Power, Median Sample Size, Maximum Sample Size, and True Effect Size Analyzed using the Binomial MaxSPRT Model
Statistical power isolines travel from northwest to southeast. Median sample size isolines travel from southwest to northeast. Overall type I error set to 0.05.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test; Pr, power.

6.1.4 Importance of Comparison

Why is it important for the user to take into account the comparisons available in Figure 8 and Figure 12? The most important point to consider is that moderate decreases in statistical power (i.e., jumps in the isolines traveling from northwest to southeast) may or may not decrease the median sample size depending on whether one is operating in the

nearly vertical portions of the median sample size isolines. That is, some of the time, decreasing power (i.e., increasing type II error and the chance of a missed signal) results in faster detection of a signal (i.e., smaller median sample sizes). Is this tradeoff worth it? *Should a user take on the bigger risk of missing the signal if it means they can find it faster?*

These tradeoffs have more concrete meaning by translating the information time concepts into calendar time. A user may value the decrease in median sample size (i.e., the quicker detection time) differently depending on whether the quicker detection occurs in 1 month or 12 months. These considerations are also influenced by the user's prior perceptions of whether an excess risk is likely or not. The user's prior perception is sometimes referred to as a Bayesian prior belief, so named for the statistician Thomas Bayes who is responsible for the subfield of Bayesian statistics and inference that allows one to incorporate existing knowledge or belief into analysis.[361]

For example, if a user believes there is a significant possibility of excess risk to be detected, then waiting an additional 11 months to detect such a risk may be unacceptable because real harm occurs in the greater population (i.e. external to the one under observation) during that additional detection time. In summary, information time is an important first step to understanding sequential database surveillance scenarios, but public health policymakers must consider how information time translates into calendar time to *meaningfully* assess sample size considerations. In the simulator, this step is performed with the Calendar Time Sub-Model, explained next.

6.2 Calendar Time Sub-Model

The calendar time sub-model estimates adoption and utilization of the medical product under surveillance using the databases proposed for inclusion in sequential database surveillance. This sub-model generates estimates of information time as a function of calendar time.

---

[361] An overview of the differences between Bayesian and frequentist statistical inference is beyond the scope of this dissertation. Generally, the FDA informally operates in a "Bayesian" framework as it continually updates and incorporates new knowledge regarding particular tracked safety issues. For information on formal use of Bayesian methods in healthcare decision-making and and policy, see David J. Spiegelhalter, "Incorporating Bayesian Ideas into Health-care Evaluation," *Statistical Science* 19, no. 1 (2004): 156–174.

6.2.1 Inputs to the Calendar Time Sub-Model: Surveillance-specifics

The user begins with the following surveillance-specific inputs to the calendar time sub-model: the exposure-outcome pair to be evaluated, an epidemiologic design, and the corresponding sequential statistical model to be used for analysis. The sequential statistical models should be the same ones investigated in the information time sub-model.

6.2.1.1 Exposure-Outcome Pair / Tracked Safety Issue

First, the exposure-outcome pair is presumably the tracked safety issue of interest. The description of the exposure-outcome pair must include specific details about when an exposed person is considered "at risk" to develop the outcome of interest. These details include: 1) specification of the onset of the "at risk" period; and 2) specification of the duration of the "at risk" period. The onset and duration are driven by pharmacokinetic and biological parameters related to the medical product, i.e., when could the medical product reasonably cause the outcome in question. For example, anaphylaxis (i.e., a severe allergic reaction) may reasonably be plausible immediately after administration of a medical product whereas an acute myocardial infarction might require more time to elapse before a medical product could have plausibly caused it. The period before the onset of the "at risk" period is also referred to as the induction period or the latency period. Once the onset of the "at risk" period begins, the user will also have to specify the duration of the "at risk" period, which is referred to as the risk window. This period typically extends some time after the last administration of the particular medical product because the medical product is still believed to be "active" in a person's body.

6.2.1.2 Epidemiologic Design

An epidemiologic design indicates the way the population of interest and the comparison population will be sampled for statistical inference. Epidemiologic designs are discussed herein in subsection 5.1. For planning purposes, the important point is that historical and self-controlled designs are one-group study designs[362] (i.e., they require

[362] Rothman, Greenland, and Lash, *Modern Epidemiology*, 758

continuous observations of only one population) and the user is only concerned with modeling one subpopulation. Concurrent-controlled designs are two-group study designs, necessitating additional modeling. When the user specifies the epidemiologic design to be modeled (and most specifically, the comparison population), it is also necessary to specify the expected incidence rate of the outcome in the comparison population, which is also referred to as the background rate. The background rate is the incidence rate of the outcome expected under the null hypothesis (i.e., that there is no excess risk in the treatment group).

For self-controlled designs[363], two additional quantities must be specified: the "washout" period and the comparison window. The former is a period of time when the person is neither exposed nor unexposed and during that time, the person contributes no information to the surveillance. The latter is a period when a person contributes "unexposed" time. Unexposed time refers to time when the person is presumed to be unexposed to the medical product and either occurs prior to the administration of a medical product or after a medical product has cleared their system. Typically, there are concerns related to confounding by indication/contraindication[364] if a "pre-exposure" comparison period is used, and so "post-exposure" comparison periods are preferred.[365] Figure 13 shows an example diagram of these time periods, and the index event is the time of initial exposure. To be clear, it is possible to have a washout period equal to zero. If a "pre-exposure" comparison window is planned, then no washout period is necessary.

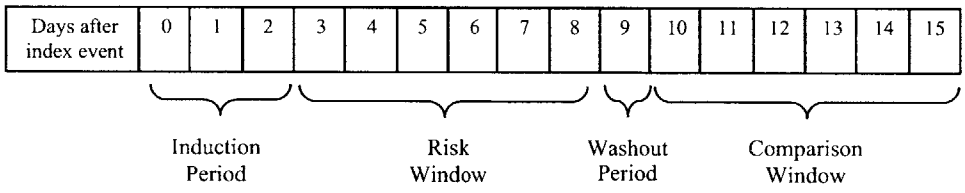

Figure 13. Example Diagram of Time Periods in Self-Controlled Epidemiologic Designs
The index event is typically the date of exposure.

[363] For more on self-controlled designs, see section 5.1.
[364] For more on confounding by indication/contraindication, see section 4.2.1.
[365] Specifically, if a person experiences the outcome of interest (i.e., adverse event) during a "pre-exposure" comparison window, then that person may be more or less likely to then be subsequently exposed to the medical product of interest. This change in their likelihood of being exposed introduces a bias.

6.2.1.3 Sequential Statistical Model

The sequential statistical model specifies the type of hypothesis testing that will be performed (e.g., continuous or group sequential), and what types of sequential statistical boundaries can be accommodated. The sequential statistical model in the information time sub-model should correspond to the sequential statistical model used here. Table 3 in section 5.2 indicates the compatibility between epidemiologic designs and sequential statistical models.

6.2.2 Inputs to the Calendar Time Sub-Model: Database-specifics

The following inputs are necessary for *each database* that may contribute to sequential database surveillance: a) the size of the subpopulation of interest; b) an estimated mathematical function that describes the adoption/uptake pattern; c) the refresh delay time; d) the processing delay time; and e) any exposure and outcome misclassification estimates. The first two inputs allow the user to model how medical product adoption evolves in calendar time. The remaining three inputs – the refresh delay time, the processing delay time, and exposure and outcome misclassification estimates – allow the user to model how exposures and outcomes appear to the user conducting sequential database surveillance. Recall that validated exposure and outcome data are unavailable to the user at the time of surveillance. Instead, the user must rely primarily on electronic claims data that arrives with some lag time. Modeling these database delays mimics the *near-real time*[366] aspects of surveillance.

6.2.2.1 Subpopulation Size

The size of the subpopulation of interest (e.g., persons over 18 with a diagnosis of diabetes) can be ascertained in the Mini-Sentinel System using modular programs (i.e., database queries) while planning surveillance. In particular, summary tables or modular programs 1, 2, and 4 can be used for this purpose.[367] This subpopulation is the pool of potential adopters of the medical product.

---

[366] For discussion of near-real time surveillance, see *supra* at note 21.

[367] See explanation of summary tables and modular programs in subsection 2.2.2.1.

6.2.2.2 Estimated Adoption Functions

The estimated adoption function for the medical product is perhaps the most substantial uncertainty in the sub-model, and is the subject of more analysis in section 9. Certain products, like vaccines, which are routinely administered for enrollment in school, childcare, etc., may be less complicated to model. However, adoption of new drugs is influenced by many more factors (e.g., formulary policy, co-payments, the availability of substitute therapies, treatment guidelines, etc.), and is considerably more complex. An adoption function may be based on historical data on similar products, early adoption data on the product in question, or no data at all.

6.2.2.3 Refresh Delay Time

The refresh delay time is the frequency with which a participating data partner in the Mini-Sentinel System renews their dataset and makes it available for analysis. Essentially, as a participant in the Mini-Sentinel System, data partners agree to update their data on some periodic basis. For most partners, this is monthly or quarterly, but can be as long as annually. In a similar, precursor system - the Vaccine Safety Datalink - refreshes are performed weekly.[368]

6.2.2.4 Processing Delay Time

The processing delay time, or the claims lag time, is the time that elapses between when an exposure or outcome occurs, and when it is recorded and available for analysis in the Mini-Sentinel System. Different data streams have different processing delay times even if they originate from the same data partner. Generally, exposure data (specifically, pharmacy dispensing data) are available sooner than outcome data. Also, outcome data may have differing lag times based on their origin (e.g., ambulatory encounter data may become available more quickly than inpatient data).[369] These incoming datastreams may be modeled explicitly, and the additional modeling efforts are likely worthwhile when calendar time for surveillance is very short (e.g., influenza vaccination surveillance).

---

[368] Baggs et al., "The Vaccine Safety Datalink: a Model for Monitoring Immunization Safety."
[369] S. K. Greene et al., "Near Real-time Vaccine Safety Surveillance with Partially Accrued Data," *Pharmacoepidemiology and Drug Safety* 20, no. 6 (2011): 583–590.

Else, one can use the maximal processing delay across all incoming datastreams as the input parameter. Allowing this "data settling" period also may mitigate issues associated with latent data correction, and potential instability in sequential database surveillance.

6.2.2.5 Misclassification Parameters

The misclassification estimates - data on the sensitivity[370] and specificity[371] of the exposure and outcome classifications - allow the investigator to model noisy data. For the calendar time sub-model, I focus on the sensitivity and positive predictive value[372] of **outcomes only**. As was discussed earlier in subsection 4.2.3.1, validation data for exposures is essentially unavailable because of the secondary data collection mechanism. One could do a sensitivity analysis and speculate on exposure sensitivity and specificity. I focus on the positive predictive value of the outcomes (as opposed to specificity) because it would be implausible with this data collection mechanism to validate "true negative" cases of the outcomes. That is, it is unlikely that resources would be expended to perform medical chart validation on exposed individuals who are not electronically identified as having experienced the outcome.

6.2.3 Modeling Database-Specific Inputs to Surveillance

6.2.3.1 Modeling Exposures

Together, the inputs above parameterize a customizable delay differential equation model (sometimes referred to as a "compartment model") to estimate adoption and utilization of the medical product under surveillance. Figure 14 shows the general model that can be adapted for the particular circumstances of the exposure-outcome pair being monitored.[373]

---

[370] Sensitivity is equal to the number of true positive cases/(true positive cases + false negative cases).
[371] Specificity is equal to the number of true negative cases/(false positive cases + true negative cases).
[372] Positive predictive value is the number of true positive cases/(true positive cases + false positive cases).
[373] Note, in Figure 14, I do not show the general compartmental model to include a pre-exposure control period, which is possible but not preferred in self-controlled designs. See *supra* at note 365.

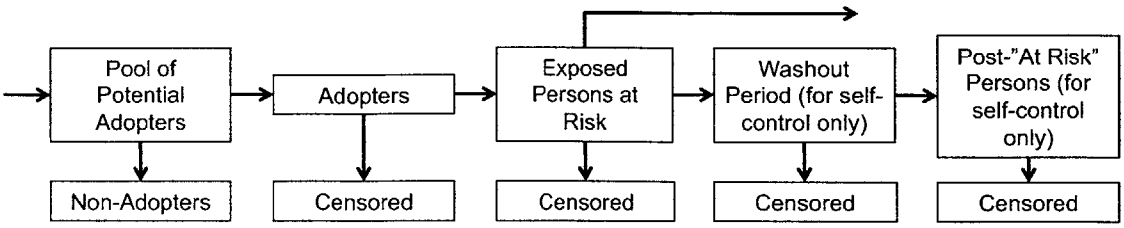

Figure 14. General Compartmental Model of Adoption and Utilization of Medical Products

It begins with the subpopulation identified as the pool of adopters. From there, they may adopt according to some adoption function or exit the system as non-adopters. Once a person has adopted, an induction period/latency period elapses and the exposed person may become "at risk" for the medical-product associated health outcome of interest (i.e., in the risk window for the purposes of surveillance). Once an exposed person becomes "at risk", that person may leave the category for three reasons: 1) that person is censored (i.e., removed from the data analysis); 2) the risk window elapses without being censored; or 3) the surveillance ends.

For point/discrete exposures such as vaccinations, the time spent in the risk window is fixed and entirely defined by the time of initial exposure. However, for continuous exposures (e.g., most drugs and therapeutic biologics), the time spent in the risk window will vary based on patterns of patient adherence, tolerance of side effects, etc. Thus, in addition to modeling adoption, one must also monitor continued utilization of the medical product. Future versions of the simulator will be able to incorporate data on utilization patterns derived from the Mini-Sentinel System. Modular programs to create those datasets are still being tested.

As a note, compartmental models are only designed to handle population averages. If there is significant variation anticipated in continued medical product utilization, an agent-based model that can accommodate more heterogeneity may be more appropriate.

For concurrently-controlled epidemiologic designs when the concurrent control is defined with an active comparator (i.e., two separate populations exposed to two separate medical products need to be monitored), a compartmental model needs to be estimated for both groups, and then additional assumptions are necessary with respect to matching.

When the design is self-controlled and the comparison period of "unexposed time" occurs post-exposure to the medical product, it is necessary to track time after the "at risk" period. In some instances, a "washout" period occurs between the "at risk" window and the comparison window. Following the washout period, or directly following the "at risk" period if the "washout" period does not exist, a person contributes time to the comparison window.

For any sequential database surveillance problem, a person can be censored for a variety of reasons, but one common issue is the loss of insurance coverage benefits, which is regarded as a loss-to-follow-up. Other censoring criteria are likely specific to the exposure-outcome pair being evaluated.

Once the general delay differential equation model has been customized to the particulars of the exposure-outcome pair (i.e., programmed to account for surveillance-specific inputs listed in 6.2.1), each database will make unique contributions to adoption and utilization. That is, the same set of general delay differential equations will be solved with the database-specific inputs listed in 6.2.2, making it a linear system of delay differential equations with one stratum for each database contributing information. Solving the system of delay differential equations produces a pattern of exposure (i.e., information time or sample size contributions) in calendar time.

6.2.3.2 Modeling Outcomes

Given a pattern of exposures as a function of person-time, then the calendar-time sub-model generates outcomes for each database based on the incidence rate under the null hypothesis, input effect sizes, and misclassification estimates. The first two variables are the same across databases but the misclassification estimates are particular to the database being considered.

An example of the output of the calendar time sub-model is shown as a table shell in Table 4 for one particular effect size. This table is repeatedly populated for the range of effect sizes of interest. It is also populated for each database being evaluated for use in surveillance.

	Database i									
		Incidence Rate Ratio=Y								
Time (months)	Cumulative Exposure (person-months)	True Positive Outcomes			False Negative Outcomes			False Positive Outcomes		
		1	...	*Nsim*	1	...	*Nsim*	1	...	*Nsim*
0	0	0	...	0	0	...	0	0	...	0
1	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
T-End	XXX	XX	...	XX	XX	...	XX	XX	...	XX

Table 4. Example Blank Output of the Calendar Time Sub-Model for an Effect Size of Interest
Abbreviations: Nsim, number of simulations; T-end, Time at the end of surveillance.

6.3 Analysis Sub-Model

The analysis sub-model allows the user to perform sequential database surveillance on simulated data, produce tabular or graphical results, and account for prior knowledge regarding the tracked safety issue when exploring sequential database surveillance.

6.3.1.1 Decision Analysis with Uncertainty

There are many parameters that a user can vary in sequential database surveillance that may lead to substantial changes in the statistical power, median sample size, or maximum sample size. Each set of parameters can be thought of as a unique configuration for surveillance. In this version of the simulator, there is some basic decision support to choose the optimal configuration among the remaining candidates. For these scenarios, a user needs to indicate the outcomes or measures on which a particular configuration is deemed to be "best." When there are multiple measures, these decisions can be performed with a multi-attribute utility function.[374] These decisions are examined under conditions of uncertainty, particularly related to the uncertainty of the true effect size. In general, one should consider at least two scenarios in all configurations: when an elevated risk exists and when it does not. Figure 15 is a basic schematic of decision tree that can be used to choose an optimal configuration.

[374] Ralph L. Keeney, "Utility Functions for Multiattributed Consequences," *Management Science* 18, no. 5, Theory Series, Part 1 (1972): 276–287.

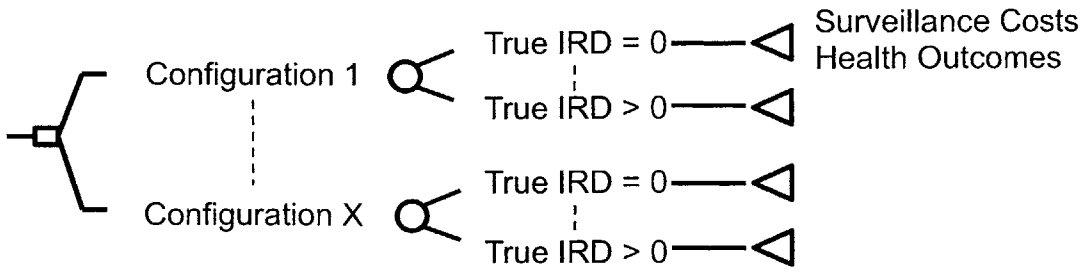

Figure 15. Basic Schematic of Decision Analysis under Uncertainty

6.4 Intention of the Simulator

In short, the Sequential Database Surveillance Simulator uses, in sequence, the information time sub-model, the calendar time sub-model, and the analysis sub-model to allow the user to experiment virtually with various sequential database surveillance scenarios. The goal of such a simulator is to be a learning tool for the user so they may gain intuition about the range of potential scenarios that could occur given certain problem-specific input parameters. The simulator is not meant to be a forecasting tool. Rather, it is meant to be a quantitative decision tool that sheds light on the evidence-generation capabilities of the Mini-Sentinel System when using the system to perform sequential database surveillance.

Next I show how to use the Sequential Database Surveillance Simulator with a fully worked example.

THIS PAGE INTENTIONALLY LEFT BLANK

7 VACCINE EXAMPLE

I plan surveillance for a single-dose, newly available live attenuated childhood vaccine that is being evaluated for an elevated risk of idiopathic thrombocytopenic purpura (ITP). ITP is bleeding disorder in which the immune system impairs the body's ability to perform normal blood clotting. Intracranial and gastrointestinal bleeding are severe complications. ITP is known to occur after many types of infections, including numerous vaccine-preventable diseases.[375] Biologically, vaccines induce an immune response, and it is theoretically possible that abnormal responses could trigger ITP.

I chose this example because exposure is discrete, the risk window is finite, and the adoption pattern is simple (i.e., children receive routine vaccinations during well visits). Throughout this example, I will make choices to tailor this example and simulated surveillance to mirror previously completed vaccine surveillance activities.[376] Accordingly, I will plan for two concurrently performed epidemiologic designs and accompanying analyses, which has been common practice.

7.1 Model Inputs

7.1.1 Information Time Sub-Model

I will use both the Poisson and binomial MaxSPRT sequential statistical models. I assume ITP in infants is expected to occur at a rate of 2 outcomes per 100,000 person-years in a clinically relevant comparison group (i.e., the background rate).[377]

7.1.2 Calendar Time Sub-Model

7.1.2.1 Surveillance-Specific Parameters

Table 5 lists relevant user-specified surveillance parameters and database-specific parameters. The primary design will be a cohort design with a historical comparison cohort of infants exposed to other vaccines. The secondary design will be a self-

[375] Sean T O'Leary et al., "The Risk of Immune Thrombocytopenic Purpura After Vaccination in Children and Adolescents," *Pediatrics* 129, no. 2 (February 2012): 248–255.
[376] See summary in Yih et al., "Active Surveillance for Adverse Events: The Experience of the Vaccine Safety Datalink Project."
[377] Deirdra R Terrell et al., "Determining a Definite Diagnosis of Primary Immune Thrombocytopenia by Medical Record Review," *American Journal of Hematology* 87, no. 9 (September 2012): 843–847.

controlled risk interval design. The primary design will be analyzed using the Poisson MaxSPRT model and the secondary design will be analyzed using the binomial MaxSPRT model. Based on the incubation period of the live virus, I assume that it is only biologically plausible that ITP could be vaccine-associated if it occurs between 1-42 days post-vaccination, which is the risk window. Therefore, both the primary and secondary designs employ a 1-day induction period (i.e., 0-1 days post-vaccination) and then a 41-day risk window (i.e., 1-42 days post-vaccination). The secondary design requires two additional values: a washout period of 1 day (i.e., 42-43 days post-vaccination) and a 41-day comparison window (i.e., 43-84 days post-vaccination). Each dose of vaccination contributes 41 person-days to the analysis, resulting in a background rate of ~1 case of ITP for every 445,427 doses.

This event is very rare and therefore, it is possible to signal on very little accrued data.[378] Therefore, I require a minimum of four outcomes to signal using the Poisson MaxSPRT.[379] I use a flat sequential statistical boundary, which is conventional for the MaxSPRT analyses and set overall type I error to 0.05. I chose this particular combination of epidemiological designs and supporting sequential statistical analyses to mimic prior analyses.

7.1.2.2 Database-Specific Parameters

To familiarize the reader with the planning process, I begin by modeling just one database: the aggregate Mini-Sentinel System Distributed Database. Using a dataset current through 2010, I used the mean enrollment data across the years 2008-2009 to generate a cohort of ~564,000 0-1 year olds, which I use as the size of my subpopulation of interest. I used these years because all data partners contributed data in the dataset available to me. Assuming approximately 4 million children are born in the US annually[380], this represents ~14% of the US population of this age group.

I model the adoption function linearly (i.e., adoption is coincident with one-year well-visits and children are assumed to be equally likely to be born on any day of the year),

[378] For evidence of this phenomenon, refer back to Figure 7 and note the higher frequency of earlier signaling in the leftmost panel.
[379] See *supra* at note 323.
[380] Centers for Disease Control and Prevention, "National Vital Statistics System", n.d., http://www.cdc.gov/nchs/nvss.htm.

and allow for a 5% probability of non-adoption (i.e., vaccine refusers). I assume a one-month refresh delay (i.e., I receive new data monthly in the Mini-Sentinel System[381]) and a two-month processing delay. In the base case, I assume no known or estimated misclassification.

Surveillance Parameters		
	Primary Design	**Secondary Design**
Effect Size of Interest	8 **excess** outcomes/ 100,000 person-years	8 **excess** outcomes/ 100,000 person-years
Comparator Outcome Rate (ρ)	2 outcomes/ 100,000 person-years	2 outcomes/ 100,000 person-years
Induction Period (δ_1)	0-1 days post-vaccination	
Risk Window (δ_2)	1-42 days post-vaccination	
Washout Period (δ_3)	N/A	42-43 days post-vaccination
Comparison Window (δ_4)	N/A	43-84 days post-vaccination
Statistical Model	Poisson MaxSPRT with minimum of 4 outcomes	Binomial MaxSPRT
Matching Ratio	N/A	1:1
Type I Error	0.05	
Database Parameters		
Subpopulation Size (M)	564,000 0-1 year olds	
Probability of Adoption (φ)	0.95	
Database Delay	2 months	
Refresh Delay	1 month	

Table 5. User-Specified Surveillance and Database Parameters in the Simulated Vaccine Example
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test

7.2 Performing Sample Size Calculations

7.2.1 Primary Design: Cohort Design with the Poisson MaxSPRT Model

The initial surveillance goal will be to detect an incidence rate ratio of 5 (i.e., an incidence rate difference of 8 excess outcomes/100,000 person-years) with 90% power. Using the information sub-model, I produce a compact display of my sample size choices

[381] Here, I am treating the entire Mini-Sentinel System as though it were one giant database to illustrate how the simulator works. In reality, each participating component database would have a different refresh time.

using the Poisson MaxSPRT model with a minimum of 4 outcomes, shown as Figure 16. To properly read Figure 16, I calculate my k scaling constant as 5 (i.e., [(1 outcome /10,000 person-years) / (2 outcomes / 100,000 person-years)]). I find the intersection of an incidence rate ratio of 5 and 90% power, marked with a star in Figure 16. The maximum sample size required is 75,000 person-years (15,000*(k=5) person-years on the y-axis), or 668,140 doses. The median sample size is just to the left of the 7000*k person-years isoline (i.e., 35,000 person-years when scaling by k). To generate a more precise median sample size, I perform a simulation as described in the previous section and find the unconditional median of the information time until the end of surveillance. The median sample size is 36,658 person-years (i.e., 326,567 doses).[382]

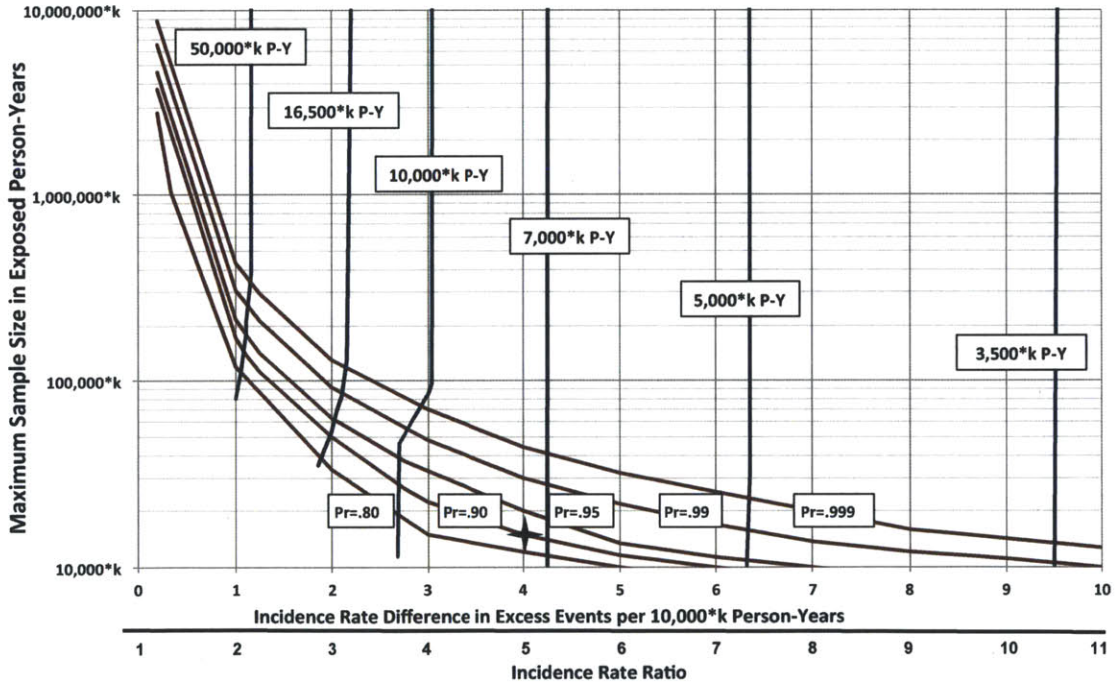

Figure 16. Relationship between Statistical Power, Median Sample Size, Maximum Sample Size, and True Effect Size Analyzed using the Poisson MaxSPRT Model with a minimum of four events
Statistical power isolines travel from northwest to southeast. Median sample size isolines travel from southwest to northeast. Overall type I error set to 0.05. The star represents the starting point of the example.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test; P-Y, person-years; Pr, power.

[382] Every dose contributes 41 person-days to the analysis, allowing for a conversion from events/person-year to events/dose.

7.2.2 Secondary Design: Self-Controlled Design with the Binomial MaxSPRT Model

Similarly, I produce Figure 17 for use with the binomial MaxSPRT model. The k scaling constant remains 5 and the effect size of interest remains 8 excess outcomes per 100,000 person-years or a fivefold incidence rate ratio. Using Figure 17, the maximum sample size required is 625,000 person-years (125,000*(k=5) person-years), and the median sample size is between 10,000*k and 20,000*k person-years. Because information time is based on the number of outcomes and only integer-valued outcomes can occur, this estimation is as precise as I can be using the binomial MaxSPRT model and information time.[383]

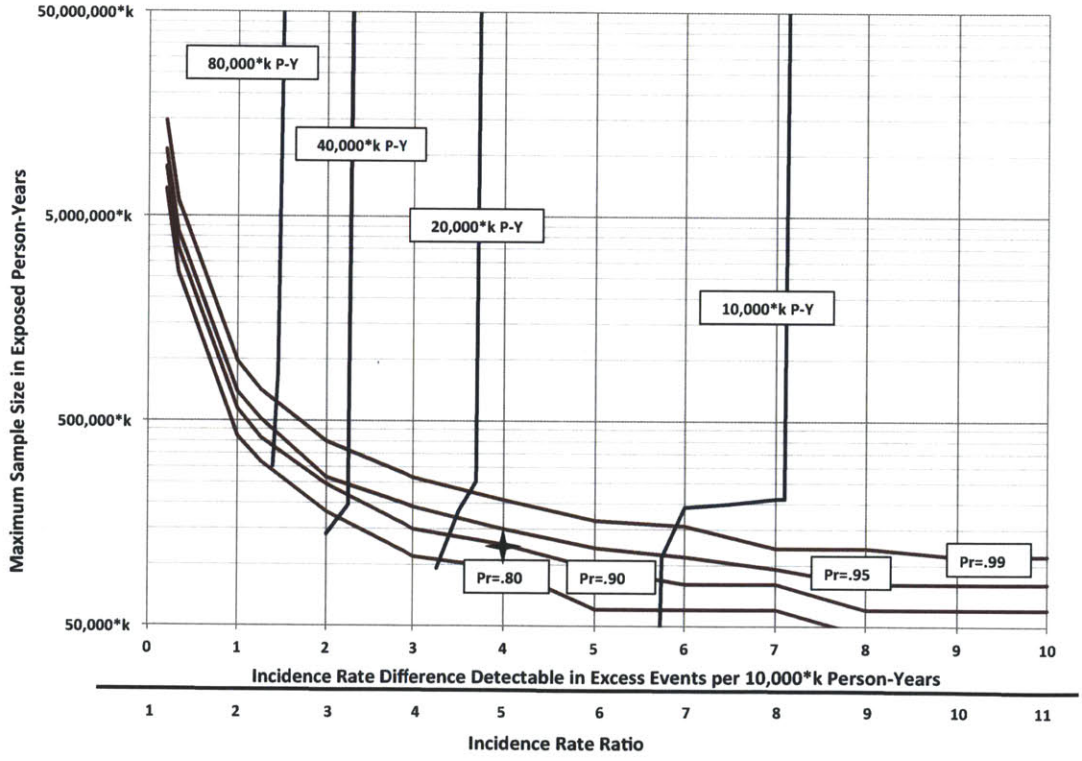

Figure 17. Relationship between Statistical Power, Median Sample Size, Maximum Sample Size, and True Effect Size Analyzed using the Binomial MaxSPRT Model
Statistical power isolines travel from northwest to southeast. Median sample size isolines travel from southwest to northeast. Overall type I error set to 0.05. The star represents the starting point of the example.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test; P-Y, person-years; Pr, power.

[383] The reference outcome of interest is 1 outcome / 10,000 person-years. Because the binomial model requires the end of surveillance to occur coincident with the arrival of an outcome, the median sample size can *only* occur at discrete values of exposed person-years.

7.3 Estimating Database-specific Contributions to Surveillance

Using the calendar time sub-model, I model sets of delay differential equations to describe potential adoption patterns. The adoption function must be specified for each epidemiologic design. Once the set of delay differential equations are solved to generate adoption patterns, I use these patterns to simulate a range of potential effect sizes. For these simulations, I perform 10,000 repetitions per effect size, which is recommended when analysis is exploratory. It is always possible to perform more repetitions later with a narrower set of surveillance configurations that are most important to the user.

As mentioned previously, I chose a linear adoption pattern to model adoption of new routine childhood vaccinations that will be explained in detail next. Later, I obtained data from the Mini-Sentinel System to validate this choice and that data is described in Appendix A.

7.3.1 Primary Design: Cohort Design with the Poisson MaxSPRT Model

For the cohort design, the general model as shown in Figure 14 can be simplified into one compartment: "exposed persons at risk," E_i, for databases i=1, 2, ... m, specified by equation (1):

$$\frac{dE_i}{dt} = s_1\theta\varphi_i M_i(t-\delta_1) - \rho E_i(t) - s_2 s_1\theta\varphi_i M_i(t-\delta_1-\delta_2) \quad (1)$$

M_i is the subpopulation size in database i. This formulation assumes a linear adoption function coincident with annual well visits and is reflective of the rate at which data arrive (θ = 12 months^{-1}). It also allows for a proportion of these children to be non-adopters (φ_i=0.95). The input flow to E_i, (i.e., "exposed persons at risk" category) is proportional to the input flow to the "adoption category" and is delayed by the induction period (δ_1). The proportionality constant is the survival probability $s_1 = exp(-\rho\delta_1)$ when an adopter "survives" the induction period (δ_1) without being censored for experiencing the outcome of interest (i.e., ITP) at the background rate (ρ). This formulation results from treating the occurrence of ITP as a Poisson process that occurs over the induction period (δ_1). The outflow from E_i occurs by either 1) being censored if one experiences the outcome according to the background rate ($-\rho E_i$); or 2) completing the time in the risk

window (δ_2) according to the proportionality constant $s_2 = exp(-\rho\delta_2)$. The reader can compare the first term of equation 1 with the third term and note they are only different based on the survivability constant and the added delay. I assume that exit and entry from the system (e.g., perhaps due to a change in health coverage or other loss to follow-up) is constant and equal, which is reasonable given the very short period of follow-up. The equation can also be written in vector form, as shown in equation (2) although all multiplication is element-wise.

$$\frac{d\mathbf{E}}{dt} = s_1\theta\boldsymbol{\varphi}\mathbf{M}(t-\delta_1) - \rho\mathbf{E}(t) - s_2 s_1\theta\boldsymbol{\varphi}\mathbf{M}(t-\delta_1-\delta_2) \quad (2)$$

7.3.2 Secondary Design: Self-Controlled Design with the Binomial MaxSPRT Model

The self-controlled design requires a slightly different adaptation of the general model as shown in equation (3).

$$\frac{dSC_i}{dt} = s_3 s_1\theta\varphi_i M_i(t-(\delta_1+\delta_2+\delta_3+\delta_4)) \quad (3)$$

In addition to the induction period (δ_1) and risk window (δ_2), the self-controlled design contains a washout period (δ_3) and a comparison window (δ_4). While the adoption function is the same (i.e., $\theta\varphi_i M_i$), persons that contribute to the analysis are not censored for experiencing the outcome of interest (i.e., ITP) during the risk window or comparison window. However, they may still be censored for experiencing the outcome during the induction period or the washout period. Consequently, survivability constants are used to account for loss-to-follow-up that occurs during these respective periods (i.e., $s_1 = exp(-\rho\delta_1)$ and $s_3 = exp(-\rho\delta_3)$). Essentially, to contribute to the analysis, the person must have completed both the risk window and the comparison window, and these persons can be simplified into a single compartment, the study completions (SC_i) for databases i=1, 2, ... m. Therefore, the input to the SC_i is the adoption inflow modified by survivability constants and delayed by the various time constants that each person must complete in

order to complete the study. Equation (3) can also be written compactly in vector notation as shown in equation (4) although element-wise multiplication operations are required.

$$\frac{d\mathbf{SC}}{dt} = s_3 s_1 \theta \boldsymbol{\varphi} \mathbf{M}(t - (\delta_1 + \delta_2 + \delta_3 + \delta_4)) \tag{4}$$

7.3.3 Modification by Database-Specific Delay and Misclassification Parameters

Once a pattern of information accrual in calendar time is estimated, the pattern must be altered to reflect when the data become accessible for analysis by incorporating the refresh delay time, processing delay time and outcome misclassification parameters. Additionally, it is at this point when the information accrual (i.e., the pattern of exposures in calendar time) is used to simulate outcomes for a pre-specified set of effect sizes. A sample of the dataset is shown in Table 6. This table is repeatedly populated for the range of effect sizes of interest. The reader should note that neither exposures nor outcomes occur until Month 3, reflecting the two-month processing delay assumption. Also, the data are only updated monthly, reflecting the one-month refresh delay.

	Database 1 (Mini-Sentinel Distributed Database)								
		Incidence Rate Ratio=5							
Time (months)	Cumulative Exposure (person-months)	True Positive Outcomes			False Negative Outcomes			False Positive Outcomes		
		1	...	*Nsim*	1	...	*Nsim*	1	...	*Nsim*
0	0	0	...	0	0	...	0	0	...	0
1	0	0	...	0	0	...	0	0	...	0
2	0	0	...	0	0	...	0	0	...	0
3	20,562	0	...	0	0	...	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...
100	5,851,056	46	...	44	0	...	0	0	...	0
101	5,911,201	48	...	44	0	...	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...
149	8,798,136	63	...	65	0	...	0	0	...	0
150	8,858,281	63	...	65	0	...	0	0	...	0

Table 6. Partial Dataset for the Simulated Vaccine Example with a Fivefold Incidence Rate Ratio and No Misclassification
Abbreviations: Nsim, number of simulations

7.4 Analysis Sub-Model

7.4.1 Performance Measure: Calendar Time to End Surveillance

I take the simulated data produced by the calendar time sub-model and perform sequential database surveillance on these data, thereby syncing my information time sub-model and calendar-time sub-model. Recall that I have specified what the true effect sizes are, and therefore, I am simply exploring the ability of my surveillance configuration (i.e., composed of particular databases) to detect these pre-specified effect sizes. Now, I observe the ***distributions of the calendar time to end surveillance*** under both my primary and secondary designs, which may be more meaningful to users who are managing the surveillance process. In other words, while information time is useful and necessary to understand statistical performance, it is the **calendar time** that helps the user estimate the impact on public health that the potential excess risk could have.

Initially, I look at how surveillance occurs when the true effect size is equal to the effect size I wish to detect (i.e., incidence rate ratio of 5). These results are in Figure 18. In this figure, the primary design/analysis (i.e., cohort design with a Poisson MaxSPRT model) is shown in the left panel and the secondary design/analysis (i.e., self-controlled design with a Binomial MaxSPRT model) is shown in the right panel. The reader will note that the statistical power is a bit higher in secondary design/analysis. This occurs because of the desire to achieve ≥ 0.90 statistical power while requiring integer-valued outcomes.

Also, the median sample size is higher in the secondary design/analysis. This occurs for two reasons. The primary reason is that only individuals who experience the outcome contribute information to the binomial model. That is, the binomial model is indifferent to individuals exposed to the medical product who do not experience the outcome of interest – these individuals are non-informative. In contrast, the Poisson model incorporates information from individuals who experience the outcome and individuals who do not. Because the Poisson model makes greater use of the information available, it is able to detect the differences between the treatment group and the comparison group more quickly. The secondary reason that the binomial model has a higher median sample size is that individuals have to contribute 84 days before their information is available to the user as compared with 42 days in the primary design.

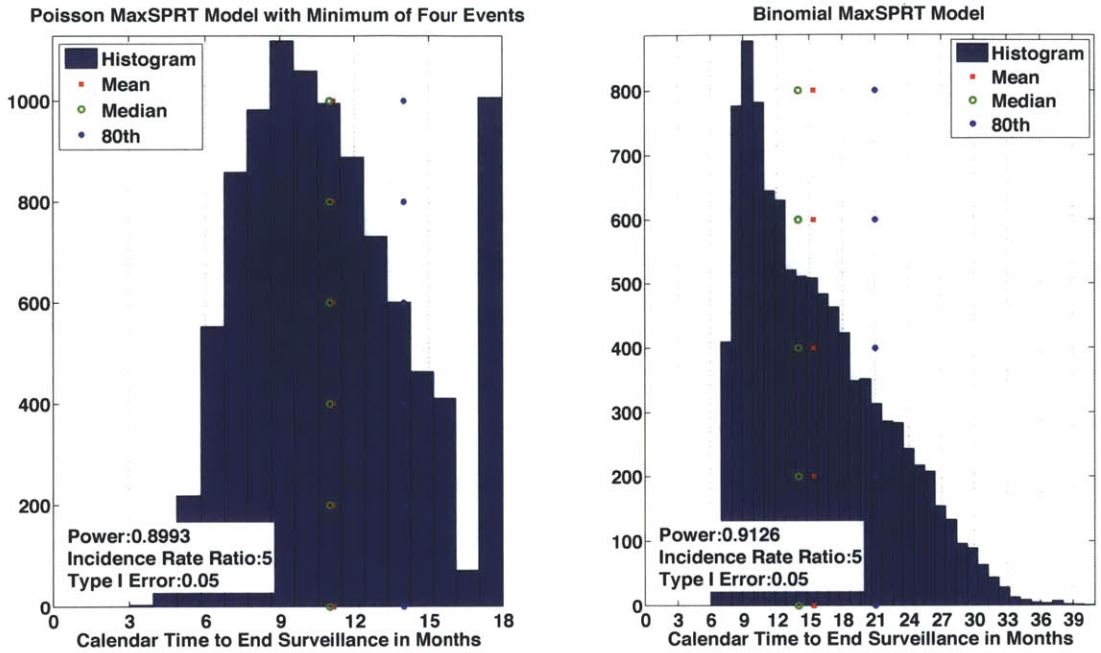

Figure 18. Distribution of Calendar Time to End Surveillance in Months in the Vaccine Example
The left panel is the primary design and analysis (i.e., cohort design with the Poisson MaxSPRT model with a minimum of four events) and the right panel is the secondary analysis (i.e., self-controlled design with the binomial MaxSPRT model). Nsim=10,000. All other parameters are as shown in Table 5.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test.

Recall that these distributions are the unconditional distributions that show the calendar time until surveillance ends, irrespective of whether a signal was detected. However, particularly with the primary design/analysis in the left panel of Figure 18, the user can also observe the rightmost bar of the distribution (i.e., the 18 month bar), which represents the instances when the surveillance configuration has incorrectly failed to reject the null hypothesis. The frequency count of this bar is the type II error. This is harder to observe in the secondary design/analysis when the instances of the failure to signal are a distribution rather than a singular value.

Figures like these can be produced for every true effect size that a user wishes to evaluate, and the shape of these distributions is similar when the surveillance configuration is powered correctly to detect the effect size. These figures start to look quite different when the null hypothesis is true (i.e., the pre-specified true effect size is an incidence rate ratio of 1), or when the surveillance configuration is underpowered. Refer back to Figure 11 in the previous section for an example of this circumstance.

Next, to evaluate how surveillance performs over a range of effect sizes, I extract key points in these distributions and present them in Table 7. In a simpler version of this table, I could show only the two sample sizes of interest: median sample size and maximum sample size. Recall that the median sample size is the summary statistic that I chose to represent the time-to-detect-a-signal. A user might wish to use a more conservative measure for planning, e.g., the 80th percentile of the distribution. Such a measure would be more conservative because it would plan the end of surveillance with an 80% probability rather than a 50% probability (i.e., the median). In any case, these tables are customizable to the user's desires and can include a number of user-specified summary statistics to describe these distributions.

True Risks		Signal (%)		Mean (months)		Median (months)		80th (months)		95th (months)	
IRD	IRR	Poi	Bin	Poi	Bin	Poi	Bin	Poi	Bin	Poi	Bin
0	1.0	0.05	0.05	17.8	69.3	18	70	18	81.5	18	94
1	1.5	0.14	0.16	17.3	52.9	18	55	18	65	18	75
2	2.0	0.28	0.32	16.5	41.5	18	44	18	54	18	62
4	3.0	0.56	0.64	14.7	27.2	16	28	18	38	18	45
8	5.0	0.90	0.91	11.2	15.4	11	14	14	21	18	27
18	10.0	0.999	0.993	7.3	9.2	7	8	9	11	11	14

Table 7. Descriptive Statistics of the Calendar Time to End Surveillance for the Primary and Secondary Design/Analyses over a Range of True Risks
Nsim=10,000. Shading indicates when the maximum sample size is reached and the null hypothesis is not rejected. Signal is the percent of time the null hypothesis is rejected. Incidence Rate Difference (IRD) is given in events per 100,000 person-years.
Abbreviations: IRD, incidence rate difference; IRR, incidence rate ratio; Poi, Poisson MaxSPRT Model; Bin, Binomial MaxSPRT Model.

The user can learn much about the performance of sequential database surveillance from Table 7. First, as expected, when the true effect size is numerically smaller than an incidence rate ratio of 5, the surveillance is underpowered. Similarly, when the true effect size is greater than the effect size that I set out to detect, power is greater. The gray shading indicates places in the distribution when the signal was not detected, which only correctly occurs when the incidence rate ratio is equal to 1 (i.e., the null hypothesis is true).

Second, the user should note that while the primary and secondary designs have similar median sample sizes when surveillance is powered appropriately (i.e., when the

incidence rate ratio ≥ 5), these sample sizes start to diverge quickly when surveillance is underpowered and the secondary design/analysis (i.e., the binomial MaxSPRT model) has notably higher median sample sizes. This higher median sample size in the binomial MaxSPRT model occurs because the maximum sample size is a random variable in this model, and the median is pulled increasingly rightward in the distribution with decreasing statistical power. By contrast, with the Poisson MaxSPRT model, because the maximum sample size is a fixed value, the median sample size has a more limited range (i.e., it will be capped at the maximum sample size of 18 months).

Third, the maximum sample sizes - equivalent to the median sample sizes when the incidence rate ratio is 1 - diverge significantly between the two designs (i.e., compare 18 months to 70 months). Practically, this divergence means that if there were no excess risk in the environment, the cohort design reaches the maximum sample size (i.e., the stopping point when one fails to reject the null hypothesis) at 18 months whereas the self-controlled design reaches the maximum sample size at 70 months. To a planner concerned with generating findings within a particular calendar period of time, these calendar time differences are substantial, particularly if there is strong prior assumption that there is no excess risk in the environment.

Fourth, the much larger maximum sample size required when using the secondary design/analysis is particularly an issue when a signal is missed but a true risk exists (i.e., type II error) because surveillance will have ended by failing to reject the null and avoidable excess events will continue to occur until some other evidence generation mechanism uncovers the true risk. That is, suppose the true effect size in the environment is equivalent to an incidence rate ratio of 2. The primary design/analysis with the current surveillance configuration will miss this result 72% percent of the time and the secondary design/analysis only performs slightly better by missing it 68% of the time. However, the primary design/analysis would have (incorrectly) declared surveillance over at a median of 18 months while failing to detect the risk whereas the secondary design/analysis would require a median of 44 months to reach the same (incorrect) result.

7.4.2 Performance Measures: Nationally Projected Excess Events

Calendar time is important for planning (i.e., when will surveillance end), but it is also important to get a sense of the avoidable excess events that might occur while surveillance is ongoing. To do this, it is important to project nationally because the population affected by a true excess risk in the environment extends beyond the observed population for the purposes of surveillance. If I assume that the rough overall size of the 0-1 year old cohort in the United States is 4 million children[384] and 95% of them receive their routine vaccinations, then I can assume 316,667 are vaccinated monthly. Given a vaccination rate in doses/month, a true effect size in excess events/doses, and the distribution of months to end surveillance, I can perform an algebraic transformation on Figure 18 to reflect the distribution of excess events that may occur if the effect size is equal to an incidence rate ratio of 5, shown in Figure 19.

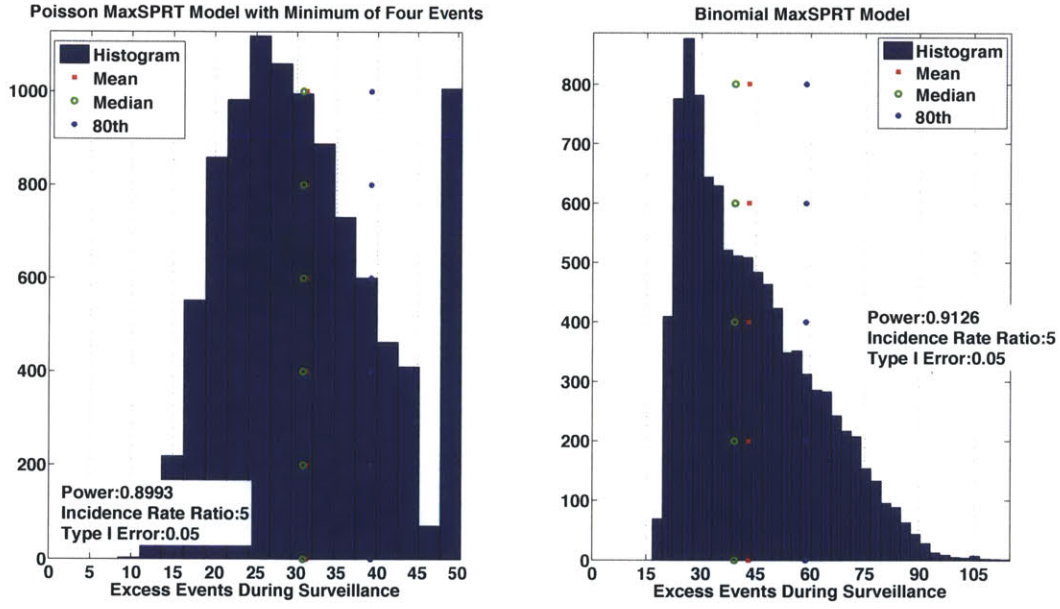

Figure 19. Distribution of Excess Events that Occur During Surveillance in the Vaccine Example
The left panel is the primary design/analysis (i.e., cohort design with the Poisson MaxSPRT model with a minimum of four events) and the right panel is the secondary design/analysis (i.e., self-controlled design with the binomial MaxSPRT model). Nsim=10,000. All parameters are as stated in Table 5.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test.

[384] Centers for Disease Control and Prevention, "National Vital Statistics System."

It is important to note that when the surveillance incorrectly fails to reject the null, the excess events calculated are the **minimum** that might occur. The actual number of excess events that will occur is dependent on how quickly the true effect size can be detected via other mechanisms. Likewise, I can produce a transformation of Table 7 to be in excess events, shown as Table 8. Again, the excess events when the null hypothesis fails to be rejected are shown as being the **minimum** possible.

True Risks		Signal (%)		Mean (events)		Median (events)		80th (events)		95th (events)	
IRD	IRR	Poi	Bin	Poi	Bin	Poi	Bin	Poi	Bin	Poi	Bin
0	1.0	0.05	0.05	0	0	0	0	0	0	0	0
1	1.5	0.14	0.16	>6.1	>18.4	>6.3	>19.3	>6.3	>22.8	>6.3	>26.3
2	2.0	0.28	0.32	>11.6	>29.1	>12.6	>30.8	>12.6	>37.8	>12.6	>43.4
4	3.0	0.56	0.64	20.5	37.6	22.4	37.8	>25.2	>53.3	>25.2	>63.1
8	5.0	0.90	0.91	31.2	43.0	30.8	39.2	39.2	58.9	>50.5	>75.7
18	10.0	0.999	0.993	45.8	57.5	44.1	50.5	56.8	69.4	69.4	88.3

Table 8. Descriptive Statistics of the Excess Events During Surveillance for the Primary and Secondary Analyses over a Range of True Risks
Nsim=10,000. Shading indicates when the maximum sample size is reached and the null hypothesis is not rejected. Signal is the percent of time the null hypothesis is rejected. Incidence Rate Difference (IRD) is given in events per 100,000 person-years.
Abbreviations: IRD, incidence rate difference; IRR, incidence rate ratio; Poi, Poisson MaxSPRT Model; Bin, Binomial MaxSPRT Model.

7.4.3 Incorporating the User's Prior Assumptions Regarding Excess Risk

As the reader may have guessed, the interpretation of tables such as Table 7 and Table 8 may be different depending on the user's prior assumptions/beliefs (i.e., hypotheses) regarding the likelihood of the true effect size in the environment. That is, when planning surveillance and anticipating both its calendar time until completion and potential excess harm in the environment that occurs while an excess risk is being detected, the user may be influenced by the strength and quantity of the existing supporting data that suggested the potential for a true excess risk in the first place (i.e., the culmination of the signal detection phase discussed in section 2.1.1). While most users may be hesitant to numerically quantify their prior assumptions, this simulator provides a way to do so. It weights the outcomes shown in Table 7 and Table 8 according to the user-specified likelihood of those true effect sizes actually occurring.

Here, a user specifies the probability distribution of the true effect sizes they believe to be possible by assigning a likelihood over the range of true effect sizes being considered, known as a Bayesian prior probability. There are many different ways to create such a distribution and there is a significant literature about how to choose a prior that is beyond the scope of this research[385], but I prefer the "lump-and-smear" and I will use that with this example. The "lump" is a lump of probability on the null hypothesis, and the "smear" spreads the remaining probability over the range of alternative effect sizes. Some authors have constructed the smear with a truncated normal distribution whereas others use a uniform distribution over an appropriate range. Of course, in all cases, the final probability distribution function must sum to one.

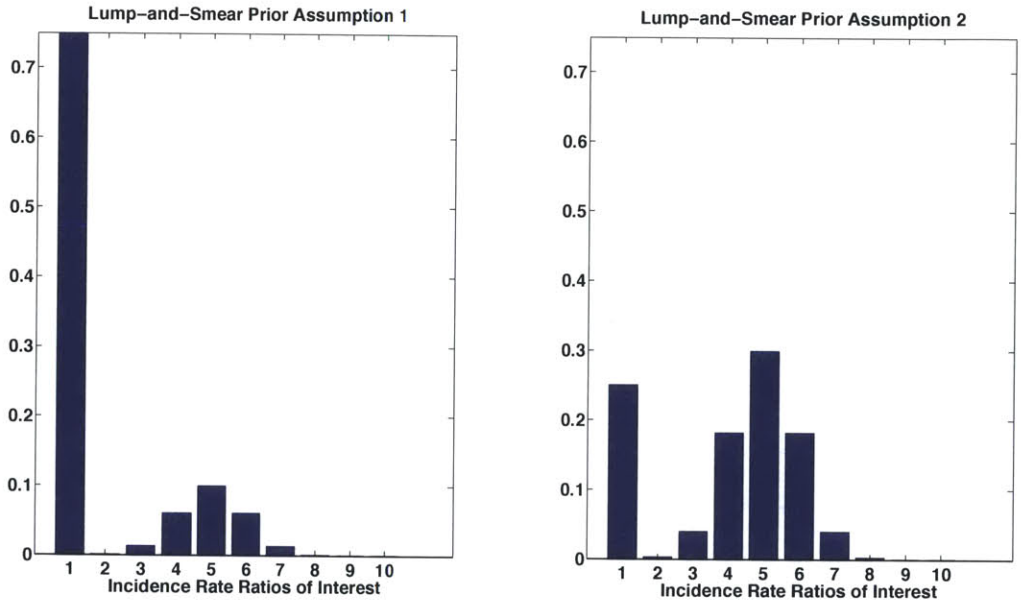

Figure 20. Two Example Prior Distributions regarding Likelihood of True Effect Size

I begin with two lump-and-smear prior assumptions of the true effect size with the following values: 1) the lump is equal to 0.75 and the smear is a truncated normal with a mean equivalent to the incidence rate ratio that I have powered the surveillance to detect (i.e., 5), a sigma of 1.0, and truncation bounds of 1 and 10; and 2) the lump is equal to

[385] See D J Spiegelhalter et al., "Bayesian Methods in Health Technology Assessment: a Review," *Health Technology Assessment (Winchester, England)* 4, no. 38 (2000): 1–130.

0.25 and all else the same. These two prior assumptions are shown respectively in Figure 20. The leftmost panel represents the prior assumption when the user has a strong feeling that there is little likelihood of a true excess risk. This circumstance might occur if the user had an obligation to monitor a set of adverse events for each new vaccination, but lacked any data to support a potential association. The rightmost panel represents the prior assumption when the user has a strong feeling that there is an excess risk in the environment. This circumstance might occur if data from spontaneous reporting systems prompted the creation of the tracked safety issue. Again, the user can create many potential prior distributions to explore, but I assume the user is, in part, relying on whatever existing data suggested a tracked safety issue was worth evaluating in the first place.

To make use of these prior distributions, I sample them using an importance sampling scheme[386] and then, using the sampled effect size, I run the simulator. I do this with 10,000 samples. From these numbers, I can produce a new weighted distribution of the time to end surveillance and the excess events to end surveillance for both the primary and secondary analyses with the weighting reflecting the prior assumptions regarding the likelihood of a true excess risk.

7.4.3.1 Prior Assumption with a Lower Likelihood Assigned to a True Excess Risk

Figure 21 shows the distribution of the time to end surveillance in the circumstance that the prior assumption that assigned a lower likelihood to a true excess risk existing was true (i.e., lump-and-smear prior 1 or the leftmost graph of Figure 20). The user should note the significant differences in the time to end surveillance for the primary v. secondary design are consistent with instances of no excess risk shown in Table 7 on page 137 (i.e., when the incidence rate ratio is equivalent to 1).

[386] See Reuven Y. Rubinstein and Dirk P. Kroese, *Simulation and the Monte Carlo Method*, vol. 2 (Hoboken, N.J.: John Wiley & Sons, 2008), 131–140.

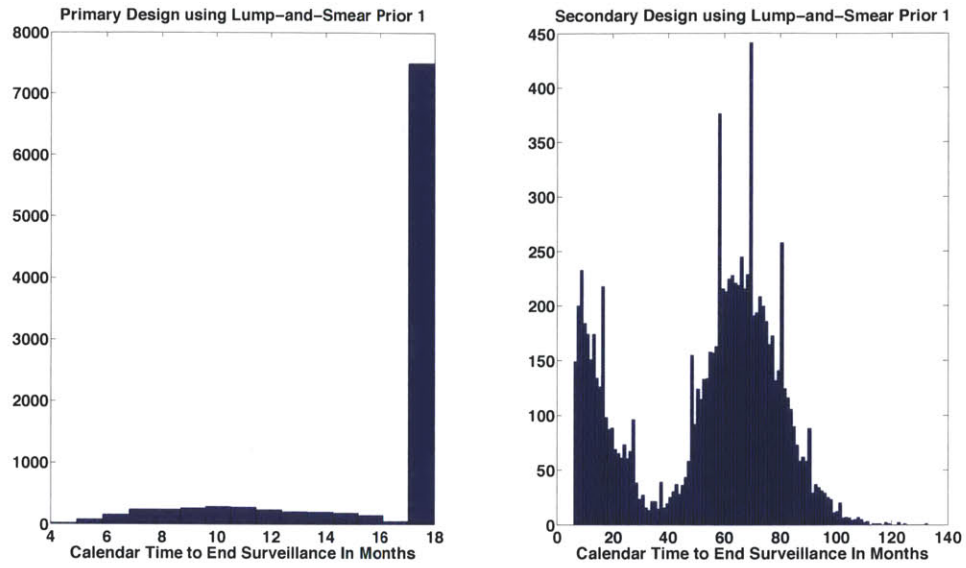

Figure 21. Distribution of the Calendar Time to End Surveillance in Months when Applying Lump-and-Smear Prior Assumption 1 in the Vaccine Example
The left panel is the primary design/analysis (i.e., cohort design with the Poisson MaxSPRT model with a minimum of four events) and the right panel is the secondary design/analysis (i.e., self-controlled design with the binomial MaxSPRT model). Nsim=10,000. All parameters are as stated in Table 5.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test

Figure 22 shows the distribution of excess events that occur during surveillance in the circumstance when the prior assumption that assigned a lower likelihood to a true excess risk existing was true (i.e., lump-and-smear prior assumption 1). As should be expected, when no excess risk exists, there are no excess events that occur during surveillance. Since the user expects this situation 75% of the time in lump-and-smear prior assumption 1, the user would really be looking to understand the distribution of excess events that could occur based on their smaller assigned probability that a true excess risk exists. Optimization for surveillance criteria could take a form that relates to minimizing these potential excess events below a particular ceiling with a particular probability.

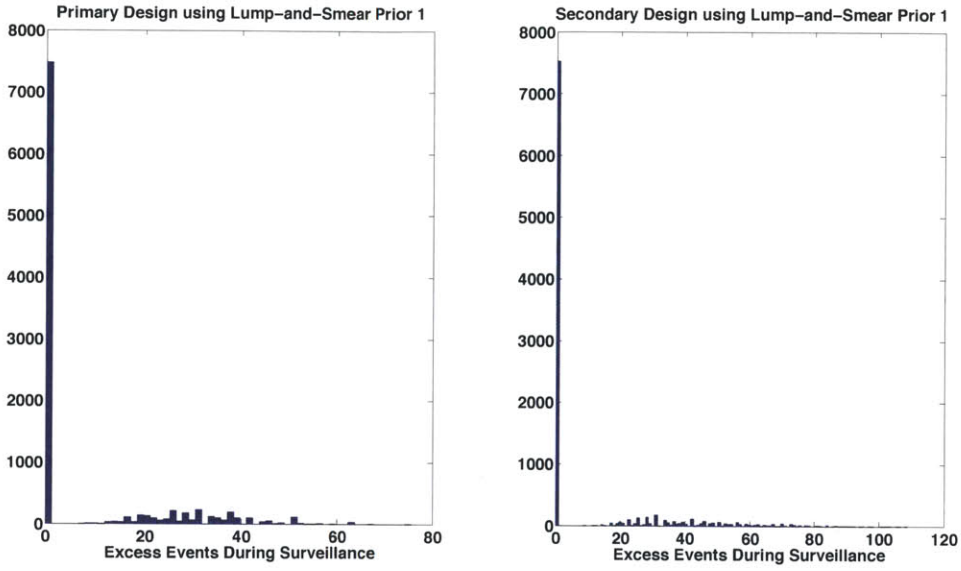

Figure 22. Distribution of Excess Events that Occur During Surveillance when Applying Lump-and-Smear Prior Assumption 1 in the Vaccine Example
The left panel is the primary design/analysis (i.e., cohort design with the Poisson MaxSPRT model with a minimum of four events) and the right panel is the secondary design/analysis (i.e., self-controlled design with the binomial MaxSPRT model). Nsim=10,000. All parameters are as stated in Table 5.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test.

7.4.3.2 Prior Assumption with a Higher Likelihood Assigned to a True Excess Risk

Figure 23 and Figure 24 show the same distributions with the second lump-and-smear prior assumption that places considerable likelihood on the existence of an excess risk. As expected, in Figure 23, the time to end surveillance is faster than in Figure 21 reflecting the increased likelihood of a true excess risk. Figure 24 also reflects the greater probability of excess events that occur during surveillance as a result of the belief that a true excess risk exists.

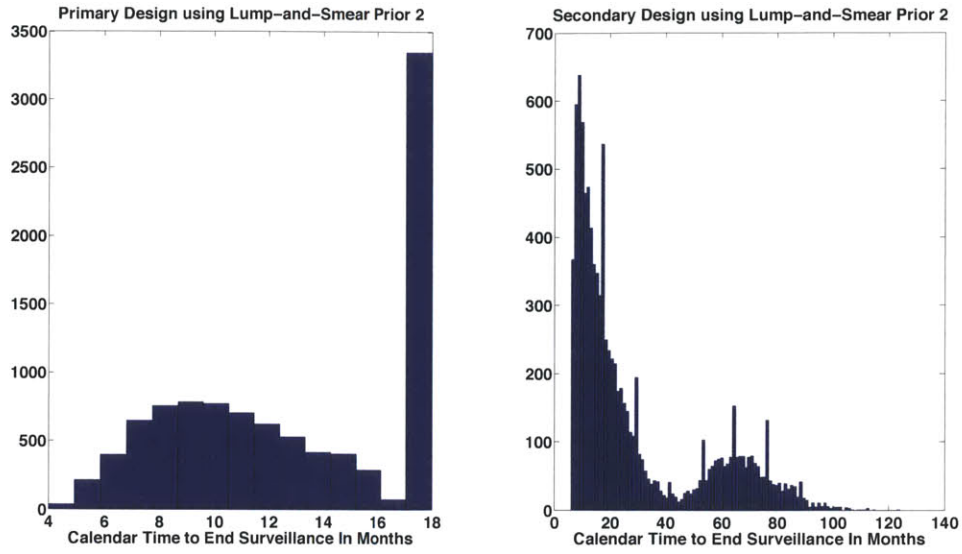

Figure 23. Distribution of the Calendar Time to End Surveillance in Months Applying Lump-and-Smear Prior Assumption 2 in the Vaccine Example
The left panel is the primary design/analysis (i.e., cohort design with Poisson MaxSPRT model with a minimum of four events) and the right panel is the secondary design/analysis (i.e., self-controlled design with binomial MaxSPRT model). Nsim=10,000. All parameters are as stated in Table 5.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test

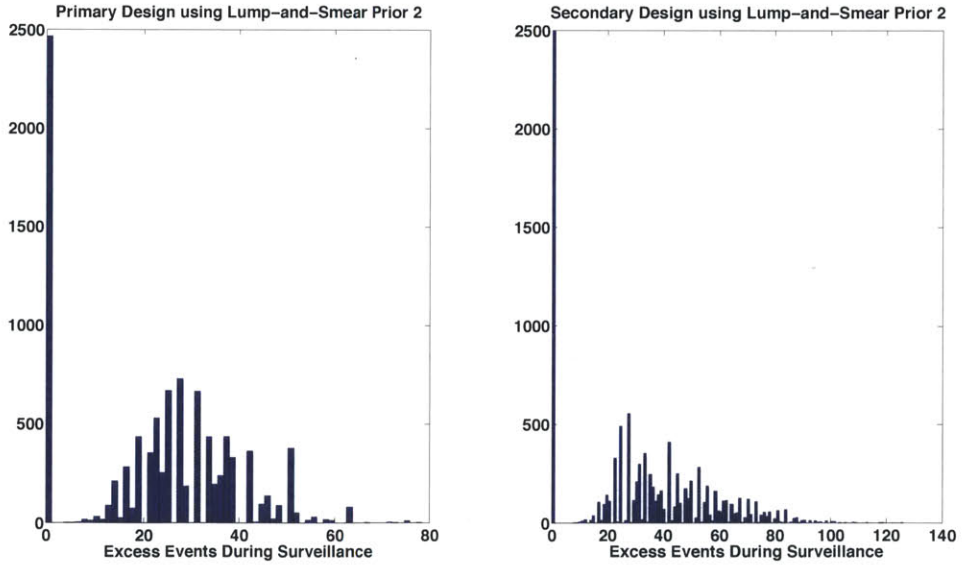

Figure 24. Distribution of Excess Events that Occur During Surveillance Applying Lump-and-Smear Prior Assumption 2 in the Vaccine Example
The left panel is the primary design/analysis (i.e., cohort design with the Poisson MaxSPRT model with a minimum of four events) and the right panel is the secondary design/analysis (i.e., self-controlled design with the binomial MaxSPRT model). Nsim=10,000. All parameters are as stated in Table 5.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test

7.5 Summary

In this example, I have not delved into a choice among surveillance configurations in the interest of showing the user how the simulator operates with a simple example. Clearly, the user could initiate surveillance configuration choices by varying the detection criteria, the choice of primary design, and the many epidemiologic choices related to how exposures and outcomes are counted and compared. Next, I turn to an example with a surveillance configuration choice to illustrate how the user might draw comparisons and learn about potential scenarios. I do this by relaxing the assumptions relating to zero misclassification, which are quite unrealistic, but have thus far allowed the reader to explore the simulator without additional levels of complication.

8 VACCINE EXAMPLE WITH MISCLASSIFICATION

8.1 Modeling Misclassification

Let us take our existing example, and examine the effects of outcome misclassification on statistical power, and the two sample sizes of interest, maximum and median sample size. For these purposes, I will always assume non-differential misclassification although in certain circumstances, this assumption may be quite strong.[387] Many statisticians and epidemiologists have studied non-differential misclassification in traditional retrospective epidemiologic studies[388] and several have published *post hoc* correction techniques.[389] However, very little has been written about misclassification with respect to sequential analysis. This absence is likely because it was a problem not previously encountered. Most sequential analysis has been done in support of clinical trials that analyze primary data. Because these data are gathered explicitly, there is a much lower likelihood of misclassification. To wit, misclassification might occur in these circumstances because of scientific differences in interpretation of data. Sequential statistical analysis of secondary data is expected to contain some degree of misclassification but these effects have not been evaluated. I extend the same mathematical proofs of prior authors who examined non-sequential misclassification to a sequential setting.

For each database $i=1,2,...,m$ under consideration, I now assign a positive predictive value (PPV) in the comparator group (PPV_i^0) and a sensitivity (ϕ_i). I assign the PPV to the *comparator group only* because Green has shown that the PPV in the treatment group (PPV_i^1) is always higher than in the comparator group when an excess risk exists while assuming non-differential misclassification.[390] The reason is that the treatment group is actually composed of two different sub-groups that are impossible for the user to differentiate. The sub-groups are 1) those that experience the outcome for reasons

---

[387] See section 4.2.3.2 for a discussion on differential and non-differential outcome misclassification.
[388] S Greenland, "Basic Methods for Sensitivity Analysis of Biases," *International Journal of Epidemiology* 25, no. 6 (December 1996): 1107–1116.
[389] Mullooly, "Misclassification Model for Person-time Analysis of Automated Medical Care Databases"; Brenner and Gefeller, "Use of the Positive Predictive Value to Correct for Disease Misclassification in Epidemiologic Studies"; Green, "Use of Predictive Value to Adjust Relative Risk Estimates Biased by Misclassification of Outcome Status."
[390] Green, "Use of Predictive Value to Adjust Relative Risk Estimates Biased by Misclassification of Outcome Status."

independent of their exposure at the background incidence rate of the outcome and 2) those that experience the outcome because of their exposure. The first sub-group resembles the comparator group and would have the same PPV. The second sub-group is responsible for the numerically higher PPV (i.e., PPV_{vx} or PPV due to vaccination) in the treatment group, which is modeled using the incidence rate ratio (IRR). Equation (5) describes this circumstance with the first half of the equation describing the first sub-group and the second half of the equation describing the second sub-group:

$$PPV_1 = (PPV_0)*(1/IRR) + (PPV_{vx})*((IRR-1)/IRR) \quad (5)$$

As the reader can see, when the IRR=1 (i.e., there is no excess risk), the second part of equation 5 is zeroed out and $PPV_1=PPV_0$.

For these purposes, PPV_1 can be derived from PPV_0, which depends on additional information about the relative size of the comparator population to the treatment population and the true effect size. Equation (6) is that algebraic derivation under the assumption of non-differential disease misclassification.

$$PPV_1=(IRR*z)/(IRR*z -1 + (1/PPV_0)) \quad (6)$$

z=ratio of person-time contributed in the comparator population to person-time contributed in the treatment population.

Given an incidence rate in the comparator group (ρ) that is assumed true, the database-specific sensitivity and PPV in the comparator group as described above, it is possible to calculate true positive cases, false positive cases, and false negative cases of the outcome of interest in both the treatment and comparator populations. A compact explanation of these rates is shown in Table 9 below. However, these rates will produce deterministic totals of cases in calendar time according to these average values, and create non-integer valued case totals. Such an approach is inconsistent with how these data actually arrive and so I treat each of these rates as an input to a Poisson process and simulate actual case arrivals.

Exposed	Coded with Outcome	Had Outcome	Case Type	Modeled Rate of Occurrence
Treatment Population				
1	0	0	TN	N/A
1	0	1	FN	$[(IRR-1)*\rho+\rho]*(1-\phi)$
1	1	0	FP	$IRR*\rho*(\phi/PPV_1)$
1	1	1	TP	$[(IRR-1)*\rho+\rho]*\phi$
Comparator Population				
0	0	0	TN	N/A
0	0	1	FN	$\rho*(1-\phi)$
0	1	0	FP	$\rho*(\phi)[(1/PPV_0)-1)]$
0	1	1	TP	$\rho*\phi$

Table 9. Rates of Occurrence of Outcomes of Interest when Modeling Misclassification
Abbreviations: TN, true negative; FN, false negative; FP, false positive; TP, true positive; IRR, incidence rate ratio.

8.1.1 Primary Design: Cohort Design with the Poisson MaxSPRT Model

Prior to assuming any misclassification, in the vaccine example in the previous section, I set surveillance to detect a fivefold incidence rate ratio with 90% power. Keeping those same detection criteria, I run my simulation again and set the true effect size to the effect size I wish to detect, i.e., a fivefold incidence rate ratio. Now, to illustrate the negative effects of misclassification on sequential database surveillance, I vary sensitivity from 0 to 1 and PPV in the comparator group from 0 to 1 and explore the changes to my surveillance scenario in Figure 25. Aside from these new changes to misclassification, all other parameters remain as they were in Table 5. That is, all exposure accruals occur as they did before in section 7 with the assumption of no exposure misclassification. However, now outcomes are simulated as explained in the previous subsection and sequential database surveillance is performed on these new data. Figure 25 addresses the primary design and analysis only (i.e., cohort design analyzed with the Poisson MaxSPRT model).

10,000 simulations were performed for each combination of sensitivity and PPV_0 listed in Figure 25. In this set of three panels, the boxed upper righthand cell reflects what happens with zero misclassification and is identical to analyses shown previously. The top, middle and bottom panels respectively show the changes in statistical power, median

sample size, and maximum sample size that occur as a result of the presence of misclassification. As one moves away from zero misclassification (i.e., the upper righthand cell), some combination of the following effects happen: statistical power deteriorates, and sample sizes to detect the true effect size increase.

POWER

Sensitivity	PPV0 = 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1.0	0.10	0.15	0.26	0.45	0.54	0.68	0.75	0.79	0.87	0.90
0.9	0.07	0.16	0.31	0.40	0.58	0.62	0.75	0.81	0.86	0.90
0.8	0.06	0.21	0.27	0.42	0.57	0.69	0.75	0.81	0.87	0.90
0.7	0.04	0.18	0.31	0.43	0.55	0.66	0.75	0.83	0.87	0.91
0.6	0.13	0.19	0.32	0.48	0.53	0.65	0.74	0.83	0.86	0.91
0.5	0.07	0.21	0.35	0.43	0.57	0.67	0.74	0.84	0.88	0.92
0.4	0.11	0.20	0.35	0.45	0.59	0.65	0.76	0.82	0.88	0.91
0.3	0.10	0.19	0.33	0.44	0.61	0.71	0.78	0.83	0.88	0.92
0.2	0.10	0.21	0.35	0.47	0.57	0.69	0.77	0.83	0.88	0.92
0.1	0.12	0.24	0.35	0.49	0.58	0.71	0.76	0.83	0.89	0.91

MEDIAN SAMPLE SIZE IN CALENDAR MONTHS

Sensitivity	PPV0 = 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.0	5	6	8	9	9	10	10	10	11	11
0.9	5	7	8	10	10	11	11	11	11	11
0.8	5	7	9	11	11	11	12	12	12	13
0.7	5	8	10	12	12	13	13	13	14	14
0.6	6	8	11	13	14	14	15	15	15	16
0.5	6	9	12	15	16	16	17	17	18	18
0.4	7	11	15	18	19	20	20	21	21	22
0.3	8	13	18	23	24	25	26	27	27	28
0.2	11	18	26	34	34	36	38	39	39	41
0.1	18	34	49	64	65	69	72	75	76	79

MAXIMUM SAMPLE SIZE IN CALENDAR MONTHS

Sensitivity	PPV0 = 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.0	5	6	8	9	11	12	14	15	17	18
0.9	5	7	8	10	12	13	15	17	18	20
0.8	5	7	9	11	13	15	17	18	20	22
0.7	5	8	10	12	14	16	18	21	23	25
0.6	6	8	11	13	16	18	21	23	26	29
0.5	6	9	12	15	18	21	24	28	31	34
0.4	7	11	15	18	22	26	30	34	37	41
0.3	8	13	18	23	29	34	39	44	49	54
0.2	11	18	26	34	41	49	56	64	72	79
0.1	18	34	49	64	79	94	109	125	140	155

Figure 25. Misclassification Matrix for the Vaccine Example using the Primary Design/Analysis
Nsim =10,000. All parameters, with the exception of misclassification estimates, are as stated in Table 5. Abbreviations: PPV, positive predictive value.

First, the reader should note that declines in sensitivity (i.e., increases in the number of missed cases) do not change the surveillance's statistical power to detect the signal since it will still require the same number of outcomes to reach one of the two stopping points. Missed cases simply cause surveillance to take longer to complete, which is evidenced by the higher median and maximum sample sizes in the middle and lower panels of Figure 25.

Second, as PPV in the comparator group declines, false positive cases (i.e., noise) begin to dilute the true signal and the surveillance is less frequently able to correctly detect the excess risk (i.e., statistical power decreases). This result is consistent with what would be expected for non-differential disease misclassification in a non-sequential analysis. Both median and maximum sample sizes decrease, which would seem like an improvement over a zero misclassification case because now both stopping points to surveillance occur earlier. However, this interpretation would be mistaken. Recall that in the Poisson MaxSPRT model, the comparison group outcomes are deterministically calculated by incrementing the background rate (i.e., the expected incidence rate of the outcome of interest in that population). With non-differential disease misclassification, noise (in the form of false positive cases) is systematically added to this rate, meaning that one arrives at the maximum sample size uniformly earlier than would otherwise have occurred. Therefore, the entire distribution is shifted leftward or earlier in terms of calendar months, including the median. So, while it is tempting to think that the earlier median sample size is an improvement, it simply reflects how the Poisson MaxSPRT model is systematically more affected by noise.

To be clear, in the less frequent scenarios when *a signal is detected*, surveillance is still correctly detecting it (i.e., it would be considered a *true positive signal* when defining a classification matrix based on the signal). However, it is detecting it using some mixture of true cases and noise. If one were to perform medical chart validation on all the electronic cases (i.e., true positive and false positive cases) that led to signal detection, then one would likely still see evidence of an excess risk but the precision of that estimate would be worse than anticipated in the presence of misclassification.

Third, the median sample size when the statistical power is less than 0.50 represents surveillance that (incorrectly) fails to reject the null hypothesis (i.e., detect a signal) a majority of the time. These numbers are shown, but do not get the same highlighting treatment as the others because they are less meaningful.

8.1.2 Secondary Design: Self-Controlled Design with the Binomial MaxSPRT Model

I produce the same figure for the secondary design/analysis (i.e., self-controlled design with the binomial MaxSPRT model) in Figure 26. Again, all parameters with the exception of the outcome misclassification estimates are the same as those listed in Table 5. Also, as before, I set the true effect size to be equivalent to an incidence rate ratio of 5. Overall, the patterns are similar in the secondary design/analysis as they are in the primary, however there are some important and notable differences that really allow the user to compare the Poisson MaxSPRT model to the binomial MaxSPRT model.

First, the binomial MaxSPRT model retains statistical power better in the face of misclassification. This result is because the binomial MaxSPRT model is not systematically adding noise in the same way the Poisson MaxSPRT model is. In the binomial MaxSPRT model, the occurrence of false positive cases is stochastic for both the treatment and comparison groups as compared to the Poisson MaxSPRT model, which is stochastic for the treatment group *alone*. Therefore, the binomial MaxSPRT model is less sensitive to the presence of noise in the form of false positive cases. In that way, the deterministic component of the Poisson MaxSPRT model works as a double-edged sword. One is able to build statistical power faster and reach a stopping point to surveillance uniformly faster in the Poisson MaxSPRT model as compared to the binomial MaxSPRT model under zero misclassification conditions. However, for the same reasons, misclassification erodes statistical power faster in the Poisson MaxSPRT model as compared to the binomial MaxSPRT model.

Second, the median sample size in the binomial MaxSPRT model slightly increases with lower PPV as opposed to decreasing as it does in the Poisson MaxSPRT model. Again, this result is because the overall distribution of the binomial MaxSPRT model is less affected by the presence of false positives. Intuitively, this result is expected and

consistent with non-sequential analysis. That is, in the presence of noise, the surveillance is less able to distinguish differences between the treatment and comparison groups and takes longer to detect a signal.

POWER

Sensitivity	PPV$_0$ 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.0	0.18	0.40	0.57	0.71	0.76	0.80	0.87	0.86	0.89	0.91
0.9	0.19	0.45	0.58	0.72	0.75	0.83	0.87	0.89	0.89	0.92
0.8	0.20	0.41	0.61	0.69	0.80	0.81	0.85	0.88	0.91	0.93
0.7	0.20	0.46	0.58	0.69	0.77	0.80	0.85	0.90	0.90	0.93
0.6	0.23	0.43	0.63	0.74	0.79	0.82	0.87	0.89	0.90	0.91
0.5	0.20	0.43	0.60	0.71	0.81	0.81	0.87	0.88	0.90	0.91
0.4	0.22	0.44	0.60	0.71	0.80	0.83	0.86	0.88	0.92	0.92
0.3	0.25	0.45	0.63	0.73	0.79	0.86	0.88	0.89	0.91	0.92
0.2	0.23	0.44	0.61	0.72	0.79	0.85	0.87	0.90	0.92	0.93
0.1	0.23	0.46	0.64	0.73	0.80	0.84	0.87	0.91	0.91	0.94

MEDIAN SAMPLE SIZE IN CALENDAR MONTHS

Sensitivity	PPV$_0$ 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.0	14	17	17	16	16	16	15	15	14	14
0.9	14	18	18	18	17	16	15	15	15	15
0.8	16	19	20	19	18	18	17	16	16	16
0.7	17	21	22	21	20	20	19	18	18	18
0.6	19	24	24	22	22	22	21	20	21	19
0.5	22	28	28	28	25	25	24	23	23	22
0.4	26	33	34	32	30	29	29	28	27	27
0.3	32	42	43	40	39	36	36	35	33	33
0.2	46	61	64	58	57	52.5	51	49	48	48
0.1	87	114	115	113	101	99	97	92	91	89

MAXIMUM SAMPLE SIZE IN CALENDAR MONTHS

Sensitivity	PPV$_0$ 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.0	17	27	35	42	48	53	58	62	66	70
0.9	18	29	38	46	53	59	64	69	73	77
0.8	20	32	43	51	59	65	71	76	81	85
0.7	22	36	48	58	66	74	81	87	92	97
0.6	25	41	55	67	77	86	94	100	107	112
0.5	29	48	65	79	91	102	111	119	127	134
0.4	35	59	79	97	113	126	137	148	157	166
0.3	44	77	104	127	148	167	182	196	208	220
0.2	64	112	154	188	219	246	269	291	310	327
0.1	121	220	301	371	433	488	530	574	614	648

Figure 26. Misclassification Matrix for the Simulated Vaccine Example using the Secondary Design/Analysis

Nsim =10,000. All parameters, with the exception of misclassification estimates, are as stated in Table 5. Abbreviations: PPV, positive predictive value.

Finally, recall that the maximum sample size in calendar months for the binomial MaxSPRT model is calculated by taking the median of the time to end surveillance when the incidence rate ratio is 1. For significant decreases in sensitivity, the maximum sample size is likely unacceptable to any user (i.e., ranges in excess of 120 calendar months).

8.2 Likely Misclassification in the Simulated Vaccine Example

Continuing with the vaccine example I have been using throughout, let us examine what kind of misclassification might be expected in this example based on previous validation studies. Outcome misclassification depends on the electronic algorithm used to detect the outcome of interest. As Chubak et al. point out, when an electronic algorithm is very inclusive (i.e., casts a wide net), there are likely to be more false positive cases that occur and fewer missed cases (i.e., PPV is low and sensitivity is high).[391] Conversely, when an electronic algorithm is fairly narrow, there are likely to be more missed cases that occur but there is a higher probability of correct identification when a case is identified (i.e., sensitivity is low and PPV is high). The user will likely choose the algorithm based on the priorities for the study, i.e. whether false positives or false negatives are more costly. Again, the user's sense of the likelihood that a true excess risk exists is likely a large determinant of their preferences. Let us examine three algorithms that have been studied that could be used to detect the outcome of interest, idiopathic thrombocytopenic purpura (ITP).

The first algorithm detects the occurrence of ICD-9[392] code 287.3. Terrell et al. performed a validation study using this algorithm in both children and adults, but I limit the application of these results to the studies in children, which is reflective of the example.[393] She finds a combined PPV of 0.54 (225/323 records in the outpatient setting and 12/118 records in the inpatient setting) for a definitive diagnosis of ITP. There were no data reported on sensitivity. The reader should note the significant differences in

[391] Chubak, Pocobelli, and Weiss, "Tradeoffs Between Accuracy Measures for Electronic Health Care Data Algorithms."
[392] An ICD-9 code is used in medical billing and coding to describe diseases, injuries, symptoms and conditions.
[393] Terrell et al., "Determining a Definite Diagnosis of Primary Immune Thrombocytopenia by Medical Record Review."

misclassification depending on the healthcare delivery setting as this report reinforces previous findings.[394]

The second algorithm (hereafter Algorithm 2 in subsequent figures) requires the occurrence of any of the following ICD-9 codes (287, 287.0, 287.1, 287.2, 287.3 287.31, 287.39, 287.4, 287.5, 287.8, 287.9) plus a laboratory value of <50,000 platelets. This study was done in five databases that are now a part of the Mini-Sentinel System.[395] The use of additional laboratory data is because these five databases originate from integrated delivery systems where laboratory data are available. I will use the published results from the years 2005-2008. The PPV for this algorithm is 0.53 and the sensitivity for this algorithm is 0.99.[396]

The third algorithm (hereafter Algorithm 3 in subsequent figures) comes from the same study and requires the occurrence of ICD-9 code 287.31 plus a lab value of <50,000 platelets. The PPV for this code is 0.79 and the sensitivity is 0.59.[397] As expected, this narrower algorithm has a higher PPV and lower sensitivity when compared with the broader Algorithm 2.

To illustrate the impact of these different algorithms on the analysis, I reperform the baseline analyses from the previous section while relaxing my assumption regarding perfect outcome classification. As before, I first create figures showing what happens when the true effect size happens to be equivalent to the effect size specified as the effect size of interest (i.e., a fivefold incidence rate ratio). Again, I do this for both the primary and secondary designs/analyses. Figure 27 shows the results when using Algorithm 2. As discussed earlier, low positive predictive values degrade statistical power in the Poisson MaxSPRT model (shown on the left) to a greater degree than the binomial MaxSPRT model (shown on the right). Compare a statistical power of 0.6045 in the former to 0.7351 in the latter. In comparison to the zero misclassification results shown in Figure 18 on page 136, statistical power is lower and the median sample sizes are lower since false positive cases are contributing to signal detection. For the Poisson MaxSPRT

---

[394] Greene et al., "Near Real-time Vaccine Safety Surveillance with Partially Accrued Data."
[395] O'Leary et al., "The Risk of Immune Thrombocytopenic Purpura After Vaccination in Children and Adolescents."
[396] I calculate the sensitivity for this algorithm by counting the number of confirmed missed cases with respect to other algorithms.
[397] See *supra* at note 396.

model, the zero misclassification baseline has a statistical power of 0.8993 and a median sample size of 11 calendar months whereas Algorithm 2 has a statistical power of 0.6045 and a median sample size of 10 calendar months. For the binomial MaxSPRT model, the zero misclassification baseline has a statistical power of 0.9126 and a median sample size of 14 calendar months whereas Algorithm 2 applied to this model has a statistical power of 0.7351 and a median sample size of 16 calendar months.

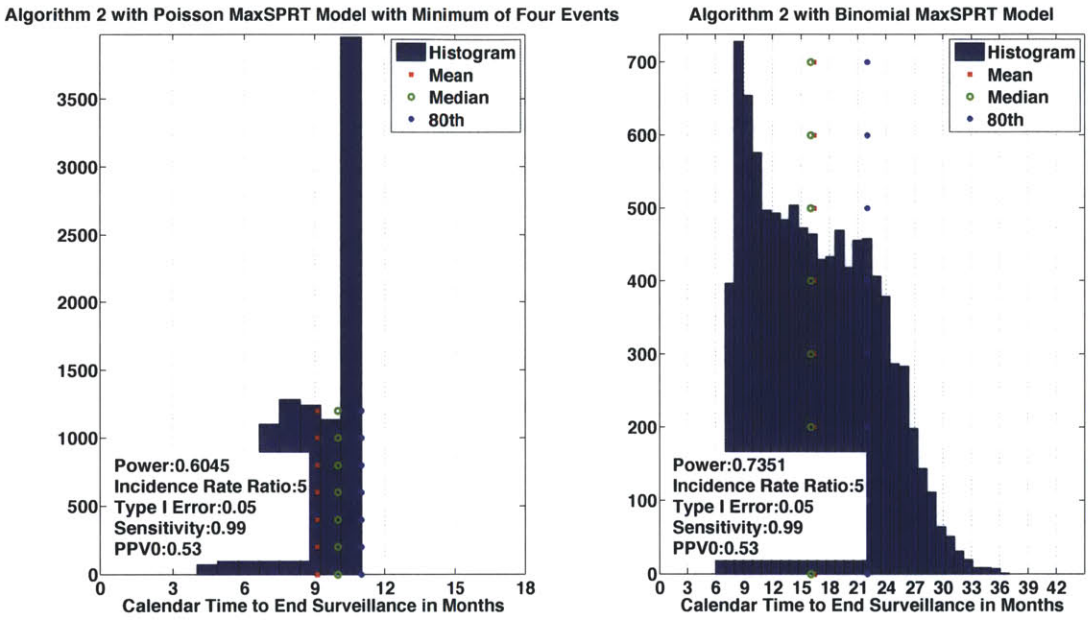

Figure 27. Distribution of Calendar Time to End Surveillance in Vaccine Example with Misclassification per Algorithm 2
The left panel is the primary design/analysis (i.e., cohort design with the Poisson MaxSPRT model with a minimum of four events) and the right panel is the secondary design/analysis (i.e., self-controlled design with the binomial MaxSPRT model). Misclassification parameters are given by Algorithm 2 and shown on the graph. Nsim=10,000. All other parameters are as stated in Table 5.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test, PPV, positive predictive value.

Figure 28 shows the results when using Algorithm 3, which creates a different misclassification pattern. Again, the left panel shows the primary design/analysis whereas the right panel shows the secondary design/analysis. Statistical power is not nearly so low as it is when using Algorithm 2. However, the reduced sensitivity has increased sample sizes relative to both the zero misclassification results shown in Figure 18 on page 136 and to the Algorithm 2 results. As explained earlier, statistical power in the secondary

design/analysis (i.e., binomial MaxSPRT Model) is less responsive to changes in PPV when compared to the primary design/analysis.

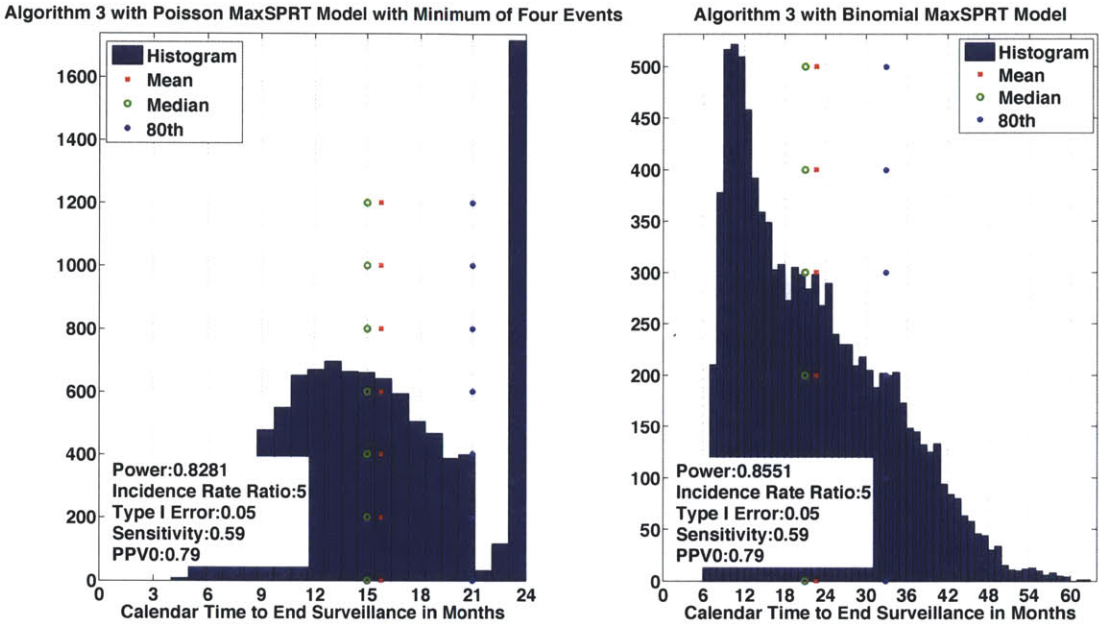

Figure 28. Distribution of Calendar Time to End Surveillance in Vaccine Example with Misclassification per Algorithm 3
The left panel is the primary design/analysis (i.e., cohort design with the Poisson MaxSPRT model with a minimum of four events) and the right panel is the secondary design/analysis (i.e., self-controlled design with the binomial MaxSPRT model). Misclassification parameters are given by Algorithm 3 and shown on the graph. Nsim=10,000. All other parameters are as stated in Table 5.
Abbreviations: MaxSPRT, Maximized Sequential Probability Ratio Test, PPV, positive predictive value.

As before, to evaluate how surveillance performs over a range of effect sizes, I extract key points in these distributions and present them in Table 10. The patterns observed in the figures above play out over the range of true effect sizes. In the middle panel of Table 10, with respect to Algorithm 2, statistical power falls below 0.5 for all effect sizes numerically less than the fivefold incidence rate ratio that the surveillance was powered to detect.

In general, these results allow the user to consider how different algorithms will shift the statistical power and sample size requirements of surveillance. The problem specifics will likely influence whether the user favors maintaining a robust statistical power or minimizing median sample size.

Zero Misclassification Case

True Risks		Signal (%)		Mean (months)		Median (months)		80th (months)		95th (months)	
IRD	IRR	Poi	Bin	Poi	Bin	Poi	Bin	Poi	Bin	Poi	Bin
0	1.0	0.05	0.05	17.8	69.3	18	70	18	81.5	18	94
1	1.5	0.14	0.16	17.3	52.9	18	55	18	65	18	75
2	2.0	0.28	0.32	16.5	41.5	18	44	18	54	18	62
4	3.0	0.56	0.64	14.7	27.2	16	28	18	38	18	45
8	5.0	0.90	0.91	11.2	15.4	11	14	14	21	18	27
18	10.0	0.999	0.993	7.3	9.2	7	8	9	11	11	14

Algorithm 2: Sensitivity=0.99 and PPV=0.53

True Risks		Signal (%)		Mean (months)		Median (months)		80th (months)		95th (months)	
IRD	IRR	Poi	Bin	Poi	Bin	Poi	Bin	Poi	Bin	Poi	Bin
0	1.0	0.05	0.06	10.9	50.0	11	50	11	58	11	67
1	1.5	0.06	0.12	10.8	41.7	11	43	11	50	11	57
2	2.0	0.11	0.24	10.7	35.3	11	37	11	44	11	50
4	3.0	0.31	0.48	10.2	26.0	11	28	11	34	11	40
8	5.0	0.60	0.74	9.1	16.3	10	16	11	22	11	27
18	10.0	0.95	0.97	6.9	9.5	7	9	8	12	11	15

Algorithm 3: Sensitivity=0.59, PPV=0.79

True Risks		Signal (%)		Mean (months)		Median (months)		80th (months)		95th (months)	
IRD	IRR	Poi	Bin	Poi	Bin	Poi	Bin	Poi	Bin	Poi	Bin
0	1.0	0.05	0.05	23.7	100.4	24	101	24	119	24	137
1	1.5	0.13	0.16	23.1	78.7	24	82	24	96	24	111
2	2.0	0.21	0.30	22.3	62.5	24	67	24	81	24	94
4	3.0	0.47	0.59	20.2	41.5	24	44	24	59	24	70
8	5.0	0.83	0.86	15.8	22.7	15	21	21	33	24	43
18	10.0	0.995	0.992	9.9	11.9	9	11	12	15	16	21

Table 10. Descriptive Statistics of the Calendar Time to End Surveillance in Months for the Primary and Secondary Analyses over a Range of True Risks with Misclassification
Nsim=10,000 Simulations, Overall Type I Error: 0.05, Shading indicates when the maximum sample size is reached and the null hypothesis is not rejected. Signal is the percent of time the null hypothesis is rejected. Incidence Rate Difference is given in events per 100,000 person-years.
Abbreviations: IRD, incidence rate difference; IRR, incidence rate ratio; Poi, Poisson MaxSPRT Model; Bin, Binomial MaxSPRT Model; PPV, positive predictive value.

In these figures, I have shown the user what might occur if the user has set their detection criteria and assumed perfect misclassification and then actually performed

surveillance in the presence of the misclassification. However, if the user has obtained estimates of misclassification, then an obvious solution is to adjust the way that the surveillance detection criteria is set *a priori* based on these misclassification estimates. That adjustment entails artificially increasing the maximum sample size. By performing this adjustment, one can try to isolate the variations related to statistical power above (at least for one particular effect size), and examine performance in terms of sample size only. I do that next when I show a partitioning of the Mini-Sentinel System. Adjustments were performed by varying the adjustment factor until the appropriate statistical power was obtained. Adjustment factors differed for the two models. The primary design/analysis (i.e., cohort design analyzed using the Poisson MaxSPRT model) required a larger adjustment factor than the secondary design/analysis (i.e., self-controlled design analyzed using the binomial MaxSPRT model). This difference in factors is unsurprising when considering the statistical power of the primary design is more affected by changes in positive predictive value when compared to the secondary design.

8.3 Partitioning the Mini-Sentinel System

Up until now, I have treated the Mini-Sentinel System as one aggregate database and obscured the differences among its components in order to show the capabilities of the Sequential Database Surveillance Simulator. However, these component databases are qualitatively different, and this variation across these databases provides opportunities for a user to tailor the performance of sequential database surveillance to the specific circumstances of the public health question being evaluated. For example, of the three algorithms discussed above, only Algorithm 1 (claims only) can be executed across the Mini-Sentinel System. Algorithms 2 and 3 both require laboratory data that is only available in a subset of component databases. These qualitative differences merit further investigation.

There are three important sources of variation across the Mini-Sentinel System that affect the timeliness and accuracy with which signals of excess risk are detected. First, uptake/adoption of medical products is highly dependent on formulary status, clinical guidelines, and the practice of medicine within particular healthcare systems. Second,

misclassification error or noise in component databases varies across algorithms and the databases themselves. Third, the costs, in time and money, associated with the correction of misclassification bias via medical chart validation procedures are highly variable across the databases that comprise the Mini-Sentinel System.

In general, component databases of the Mini-Sentinel System fall into two categories: 1) databases from large, national health insurers that are aggregations of hundreds of regional healthcare plans, and 2) databases from smaller integrated delivery systems. Databases in the first category are typically claims-only data with open (i.e., less restrictive) formularies. These databases also account for the bulk of data in the Mini-Sentinel System. Databases in the second category have richer clinical data (e.g., laboratory data) to supplement claims data, but have more controlled formularies. Medical chart validation procedures are significantly less time-consuming and costly in the integrated delivery system databases.

Why is this important? A subset of the Mini-Sentinel System data is high quality with minimal noise. Does this high quality data contribute unique value that should be considered when estimating the demand for the Mini-Sentinel System? What are the tradeoffs between the quantity and the quality of the sample size? Let us re-examine the problem and disaggregate the Mini-Sentinel System. If we take the cohort of 0-1 year olds available in the Mini-Sentinel System and partition it according to the capability to get laboratory data, then we can expect 40% of the original cohort to have laboratory data available. As a note, this is not necessarily the typical breakdown of the Mini-Sentinel System today. This example reflects circumstances that occurred for this cohort of interest at the time the dataset was created. The Mini-Sentinel System is a dynamic data system and circumstances may be quite different for other tracked safety issues evaluated at other timepoints. However, for the user planning on conducting surveillance, they will have be able to execute modular programs to find out this exact information.

In this portion of the example, I imagine the user is either choosing to apply Algorithm 1 broadly across the Mini-Sentinel System or to apply Algorithm 3 to the subset of the data that also has laboratory data (i.e., 40%). I assume sensitivity is 0.99 with Algorithm 1. I will rerun the analyses with these database sizes but also by applying an adjustment factor to the maximum sample size (i.e., the upper limit on the length of

surveillance). I perform this step to try to force these analyses to have comparable statistical power. Results are shown in Table 11.

Algorithm 1 (PPV=0.52, Sensitivity=0.99), Database Size=564,000											
True Risks		**Signal (%)**		**Mean (months)**		**Median (months)**		**80th (months)**		**95th (months)**	
IRD	**IRR**	**Poi**	**Bin**	**Poi**	**Bin**	**Poi**	**Bin**	**Poi**	**Bin**	**Poi**	**Bin**
8	5.0	0.90	0.91	13.7	18.0	12	16	21	26	28	34
Algorithm 3 (PPV=0.79, Sensitivity=0.59), Database Size=.4*564,000											
True Risks		**Signal (%)**		**Mean (months)**		**Median (months)**		**80th (months)**		**95th (months)**	
IRD	**IRR**	**Poi**	**Bin**	**Poi**	**Bin**	**Poi**	**Bin**	**Poi**	**Bin**	**Poi**	**Bin**
8	5.0	0.91	0.92	38.0	50.1	33	43	55	77	77	109

Table 11. Descriptive Statistics of the Calendar Time to End Surveillance in Months for the Primary and Secondary Analyses with *A Priori* Adjustments for Anticipated Misclassification
Nsim=10,000 Simulations, Overall Type I Error: 0.05, Shading indicates when the maximum sample size is reached and the null hypothesis is not rejected. Signal is the percent of time the null hypothesis is rejected. Incidence Rate Difference is given in events per 100,000 person-years.
Abbreviations: IRD, incidence rate difference; IRR, incidence rate ratio; Poi, Poisson MaxSPRT; Bin, Binomial MaxSPRT Model, PPV, positive predictive value.

The important finding is that when one adjusts the analysis to preserve statistical power in the face of noise created from false positives, there appears to be no time advantage to using the more accurate (i.e., higher PPV) algorithm on a smaller subset of higher quality data. That is, if the true risk existed at the level that we have powered surveillance to detect (i.e., fivefold incidence rate ratio), then the size advantage outweighs the data quality because lower median sample sizes still imply quicker detection. Also, by artificially inflating statistical power, theoretically there still would be enough true positive cases to signal if the signal were based on true positive cases alone.

As a sensitivity analysis, I tested how low the PPV of Algorithm 1 would have to be before the size advantage was negated. In this case, it turned out to be a PPV of 0.06 for the binomial MaxSPRT model, and closer to a PPV of 0.13 for the Poisson MaxSPRT model. Those results are shown in Table 12. The reader should note that this might not always be the case and is entirely dependent on the numbers that pertain to this tracked safety issue. In this example, I have assumed that the user would cope with poor PPV by adjusting the analysis with a correction factor. However, it is possible that a user may

have a threshold below which they would not consider using a particular detection algorithm even if a correction factor were applied because they would be unsure about the validity of the final results.

If the user decided that chart validation were required for the more poorly performing algorithm, then the decision calculus would change considerably. First, recall that the largest data holders in the Mini-Sentinel System have the most expensive and least timely chart validation procedures. The time to end surveillance considered in Table 12 does not consider additional time for chart validation. If these time and cost considerations are accounted for, then – depending on the specific values attributed to the time and cost of chart validation – it is possible that the size advantage is negated and the user might prefer the slower uptake and more accurate electronic algorithm without chart validation.

The frequency of the outcome also plays a considerable role in this decision. When detecting very rare outcomes, such as ITP, it may be plausible to perform chart validation on every outcome. However, more frequent outcomes like acute myocardial infarctions may require some subset or sampling of outcomes.

Algorithm 1, (PPV of POI=0.13, PPV of BIN= 0.06, Sensitivity=0.99), Database Size=564,000											
True Risks		**Signal (%)**		**Mean (months)**		**Median (months)**		**80th (months)**		**95th (months)**	
IRD	**IRR**	**Poi**	**Bin**	**Poi**	**Bin**	**Poi**	**Bin**	**Poi**	**Bin**	**Poi**	**Bin**
8	5.0	0.89	0.91	37.3	53.9	33	49	61	85	78	110
Algorithm 3 (PPV=0.79, Sensitivity=0.59), Database Size=.4*564,000											
True Risks		**Signal (%)**		**Mean (months)**		**Median (months)**		**80th (months)**		**95th (months)**	
IRD	**IRR**	**Poi**	**Bin**	**Poi**	**Bin**	**Poi**	**Bin**	**Poi**	**Bin**	**Poi**	**Bin**
8	5.0	0.91	0.92	38.0	50.1	33	43	55	77	77	109

Table 12. Sensitivity Analysis on Calendar Time to End Surveillance in Months for the Primary and Secondary Analyses with *A Priori* Adjustments for Anticipated Misclassification
Nsim=10,000 Simulations, Overall Type I Error: 0.05, Shading indicates when the maximum sample size is reached and the null hypothesis is not rejected. Signal is the percent of time the null hypothesis is rejected. Incidence Rate Difference is given in events per 100,000 person-years.
Abbreviations: IRD, incidence rate difference; IRR, incidence rate ratio; Poi, Poisson MaxSPRT Model; Bin, Binomial MaxSPRT Model; PPV, positive predictive value.

8.4 Summary

When performing sequential database surveillance for a particular exposure-outcome pair, the user has many choices. A more realistic rendering of a database's capabilities will account for misclassification. Strategies to cope with misclassification are still emerging in the sequential database surveillance setting when the goal is take advantage of available electronic data in a timely fashion to provide early warning of medical product-associated risks in the environment. However, there is a real balance to be struck between generating both accurate and timely estimates of those risks. As the reader will see in the next section, many aspects of sequential database surveillance are beyond the user's control. When misclassification itself is one of those aspects, the user can experiment with different misclassification control policies in the Sequential Database Surveillance Simulator to learn more about the effects of various policies on both the timeliness and accuracy of signal detection.

THIS PAGE INTENTIONALLY LEFT BLANK

9 MODELING ADOPTION OF MEDICAL PRODUCTS

Recall that sequential database surveillance requires the simultaneous management of information time and calendar time. That is, for sequential database surveillance within the Mini-Sentinel System to be a useful evidence generation capability, information/sample size accrual (i.e., the adoption and utilization of the medical product being evaluated) has to occur within a calendar timeframe appropriate to regulatory decision-making. The ability to reach a stopping point (i.e., either rejection or acceptance of the null hypothesis of no excess risk) in sequential database surveillance requires a threshold level of information (i.e., sample size) that clearly varies with the statistical power desired to detect a true effect size and the frequency of the outcome being detected, among other things. In short, the advantage of using a simulator such as the one presented herein is the ability to sort tracked safety issues into instances when sequential database surveillance might be more or less useful than other research approaches (e.g., randomized controlled trials) to evaluate safety.

However, the Sequential Database Surveillance Simulator requires the user to begin with assumptions regarding parameters that describe medical product adoption and utilization, and these parameters are among the most uncertain to be modeled. Not only are they unknown, but, as stated earlier, they are also beyond the user's control. These parameters include the identification of the potential pool of adopters of a medical product, the anticipated percentage of non-adopters of a medical product, and the function describing adoption itself. For vaccines and other "point"/discrete exposures, adoption parameters *alone* are sufficient to describe exposure. However, for continuous exposures (e.g., most drugs and therapeutic biologics), additional information is needed to describe medical product utilization, or the behavior of patients after initial adoption. For example, it is important to estimate how long adopters of new medical products continue to use such products, i.e., adherence. For antibiotics and other short course medical products, this task is less challenging than for products administered over the course of a lifetime (e.g., beta blockers).

While sequential database surveillance can occur at any time during a product's lifecycle, it is anticipated that such surveillance will occur *early*, typically in the near-term following approval and/or commercial launch of a new medical product. If

surveillance is not immediately postapproval, then it may be possible to gather preliminary data on adoption and utilization patterns. These data may be obtained through execution of modular programs, and may form the basis for further adoption and utilization projections. However, if surveillance is planned immediately postapproval, then it is left to the user to speculate about possible adoption and utilization patterns.

In the routine childhood vaccination example that I have used throughout this dissertation, an adoption function coincident with a one-year old well visit is not unreasonable because of public policy requirements that demand immunizations for entry to school or daycare. Appendix A describes validation data to support linear adoption patterns for routine childhood vaccinations. However, adoption of new drugs is influenced by more factors (e.g., formulary policy, co-payments, the availability of substitute therapies, treatment guidelines, the practice of medicine, advertising, etc.), and is considerably more complex.

In this section, I evaluate adoption data from previously approved new molecular entities in an attempt to provide a rational basis for assuming potential adoption functions in future surveillance. To be clear, manufacturers perform extensive market research and modeling prior to the launch of a new medical product. They have a unique and detailed understanding of the pool of potential adopters and their likely adoption patterns. My purpose is not to try to recreate these models. Rather, public health planners need some general functional forms to describe medical product adoption that can be used alongside a sensitivity analysis in sequential database surveillance models employed in the simulator.

9.1 New Molecular Entity Cohort

To establish these functional forms, I looked to previously approved new molecular entities[398] in the years 2004-2006. I selected these years so I could observe enough data to be comparable to time frames associated with completed sequential database surveillance

[398] "Certain drugs are classified as new molecular entities (NMEs) for purposes of FDA review. Many of these products contain active moieties that have not been approved by FDA previously, either as a single ingredient drug or as part of a combination product; these products frequently provide important new therapies for patients." See U.S. Food and Drug Administration and Center for Drug Evaluation and Research, "Innovation in Development of Drugs and Biological Products," WebContent, n.d., http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/default.htm.

analyses.[399] The new molecular entities selected for evaluation were limited to outpatient medications, which are commonly captured in pharmacy dispensing files based on national drug codes. Recall that medical products such as medically-attended infusions and injections, particularly newly available ones, are not well captured in the Mini-Sentinel System.[400] Of the 78 new molecular entities approved between 2004-2006, 40 were included in the "cohort," and are listed in Table 13. The national drug codes associated with each product in the cohort were identified using the FDA's public database[401], and checked against a third-party commercial database. Additionally, once I began working with these data, it became apparent that particular new molecular entities experienced significant delays between when the product was approved and when it was commercially available/launched. For example, see Apidra® (insulin glulisine injection) and Omacor®/Lovaza® (omega-3 acid ethyl esters) in Table 13 below. This mismatch created a problem for trend analysis. Therefore, I had to obtain "launch dates" for the products in the cohort from the manufacturer's press releases.

Trade Name (n=40)	Generic Name (n=40)	Days Supplied	Approval Date	Launch Date
Amitiza	lubiprostone	30	01-31-2006	04-01-2006
Apidra	insulin glulisine injection	30	04-16-2004	03-08-2006
Aptivus	tipranavir	30	06-22-2005	06-22-2005
Azilect	rasagiline mesylate	30	05-16-2006	08-01-2006
Baraclude	entecavir	30	03-29-2005	04-01-2005
Byetta	exenatide	30	04-28-2005	06-01-2005
Campral	acamprosate	30	07-29-2004	01-01-2005
Chantix	varenicline	30	05-10-2006	08-02-2006
Cymbalta	duloxetine	30	08-03-2004	08-03-2004
Enablex	darifenacin	30	12-22-2004	02-09-2005
Exjade	deferasirox	30	11-02-2005	01-16-2006
Fosrenol	lanthanum	30	10-26-2004	01-01-2005
Invega	paliperidone	30	12-19-2006	01-07-2007
Januvia	sitagliptin phosphate	30	10-16-2006	10-16-2006
Levemir	insulin detemir	30	06-16-2005	03-01-2006
Lunesta	eszopiclone	30	12-16-2004	04-03-2005

[399] Yih et al., "Active Surveillance for Adverse Events: The Experience of the Vaccine Safety Datalink Project."
[400] See *supra* at note 294.
[401] Center for Drug Evaluation and Research and Food and Drug Administration, "National Drug Code Directory", n.d., http://www.accessdata.fda.gov/scripts/cder/ndc/default.cfm.

Trade Name (n=40)	Generic Name (n=40)	Days Supplied	Approval Date	Launch Date
Lyrica	pregabalin	30	12-31-2004	09-22-2005
Nevanac	nepafenac	15	08-19-2005	09-01-2005
Nexavar	sorafenib toylate	30	12-20-2005	12-20-2005
Noxafil	posaconazole	14	09-15-2006	09-15-2006
Omacor/Lovaza	omega-3 acid ethyl esters	30	11-10-2004	10-05-2005
Omnaris	ciclesonide	30	10-20-2006	05-09-2008
Prezista	darunavir	30	06-23-2006	07-01-2006
Pylera	biskalcitrate potassium, metronidazole and tetracycline hydrochloride	10	09-28-2006	05-07-2007
Ranexa	ranolazine	30	01-27-2006	03-24-2006
Revlimid	lenalidomide	28	12-27-2005	01-01-2006
Rozerem	ramelteon	30	07-22-2005	09-01-2005
Sanctura	trospium	30	05-28-2004	08-23-2004
Sensipar	cinacalcet	30	03-08-2004	04-01-2004
Spiriva	tiotropium oral inhalation	30	01-30-2004	05-25-2004
Sprycel	dasatinib	30	06-28-2006	07-01-2006
Sutent	sunitinib malate	28	01-26-2006	04-01-2006
Symlin	pramlintide acetate	30	03-16-2005	04-01-2005
Tarceva	erlotinib	30	11-18-2004	11-27-2004
Tindamax	tinidazole	5	05-17-2004	07-01-2004
Tyzeka	telbivudine	30	10-25-2006	03-16-2007
Veregen	sinecatechins	30	10-31-2006	12-14-2008
Vesicare	solifenacin	30	11-19-2004	01-21-2005
Xifaxan	rifaximin	15	05-25-2004	07-01-2004
Zolinza	vorinostat	30	10-06-2006	10-06-2006

Table 13. Cohort of New Molecular Entities Evaluated for Adoption Patterns
Days Supplied indicates average days supplied per prescription.

9.2 Medicaid Data Explorations

Prior to the availability of data from the Mini-Sentinel System, I used U.S. Medicaid dispensing data[402] to evaluate the adoption and utilization patterns (i.e., sample size accrual) of medical products in the cohort. These data are reported quarterly by state to the Centers for Medicare and Medicaid Services and are available for outpatient drugs paid for by state Medicaid agencies per the Medicaid Drug Rebate Program. I calculated exposure profiles from these quarterly Medicaid dispensing data from the time of a

[402] Centers for Medicare and Medicaid Services, "Medicaid Drug Programs Data & Resources", n.d., http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Benefits/Prescription-Drugs/Medicaid-Drug-Programs-Data-and-Resources.html.

product's commercial launch date until the end of the third calendar quarter of 2010. These publicly available dispensing summaries are based on administrative claims data and represent utilization for approximately 60 million covered lives.[403]

Several adjustments were made to account for extreme values and limited granularity in the Medicaid data. I removed extreme values[404] and replaced them with linearly interpolated data points based on the nearest neighbors. I generally assumed 30 exposed days (i.e., days supplied) per dispensing because states limit Medicaid prescription drug utilization accordingly;[405] exceptions are noted in Table 13. Because these data were not quality-controlled, I assumed exposure could range from half to double the calculated value. These exposure accrual estimates were likely high because I assumed that every dispensing contributed exposed person-time to surveillance. In actual surveillance activities, many exposed days would be excluded because of incident user criteria, disqualifying prior events, and lapses in drug or medical benefit coverage.

Using Medicaid claims data as a temporary "substitute" for data in the Mini-Sentinel System had advantages and disadvantages. The primary advantages were its free, public availability and similarity in structure to Mini-Sentinel Systems claims data. However, the lack of quality control in these data was a significant disadvantage. Also, these raw dispensing totals did not provide much insight into "incident user" utilization, which is typically the design for sequential database surveillance studies. Finally, Medicaid populations underwent substantial upheaval before and after January 1, 2006 due to changes in eligibility as a result of the 2003 Medicare Modernization Act.[406] At the time of design and subsequent analysis, I chose the cohort so that at least 15 quarters of data were available. If I were to re-perform this analysis using Medicaid data in the future, I would limit my cohort to new molecular entities approved after January 1, 2006.

---

[403] U.S. Congressional Budget Office, *The Budget and Economic Outlook: Fiscal Years 2011 to 2021* (Washington, DC: Congress of the United States, Congressional Budget Office, 2011), http://cbo.gov/doc.cfm?index=12039; U. S. Congressional Budget Office, *The Long-term Outlook for Health Care Spending* (Washington, DC: Congress of the U.S., Congressional Budget Office, 2007), http://www.cbo.gov/doc.cfm?index=8758; Jean Hearne and Congressional Research Service, *Prescription Drug Coverage Under Medicaid: CRS Report RL30726* (Washington, D.C.: Congressional Research Service, Library of Congress, February 6, 2008), http://opencrs.com/document/RL30726.

[404] I defined extreme values as greater than 3 times the median absolute deviation from the median.

[405] Hearne and Congressional Research Service, *Prescription Drug Coverage Under Medicaid: CRS Report RL30726*.

[406] For more, see section 4.3.3.

Nonetheless, my intention is using these data with few constraints was to provide rough estimates of detectable effect sizes based on actual utilization. To do this, I used the calculated exposure pattern and varied the frequency of the outcome of interest to evaluate whether simulated true effect sizes were detectable given the sample size. In other words, if the simulated effect size were the true effect size, then that effect size was detectable if sequential database surveillance using the Poisson Maximized Sequential Probability Ratio Test (MaxSPRT) model could detect the risk with a certain success rate or reach an end to surveillance without signaling (i.e., type II error was constrained to be ≤0.20). Several tables in Appendix B show these results that I summarize next. I perform this analysis with *common* event rates of 1 event/100 person-years, *infrequent* event rates of 1 event/1000 person-years, *rare* event rates of 1 event/10,000 person-years, and *very rare* event rates of 1 event/100,000 person-years.[407] Others have performed similar studies on a potential European equivalent of the Mini-Sentinel System and found similar results.[408]

9.2.1 Common Event Rates

Table 22 in Appendix B reports the simulated effect sizes that were achievable when trying to detect an event rate of 1/100 person-years under the Poisson MaxSPRT model with detection criteria that sets type I error to 0.05 and type II error to 0.20. An example of a common event rate is the event rate of interest in the Mini-Sentinel System pilot project: 9 acute myocardial infarctions/1000 person-years among users of oral hypoglycemics.[409] Of the 40 products listed in Table 13, I found that sequential database surveillance on 8 quarters of data could identify a simulated incidence rate ratio of ≥2.5 for 30 products with the most generous estimates of exposures and for 23 products with the most conservative estimates. Expanding surveillance to 15 quarters would have identified a simulated incidence rate ratio of ≥2.5 for 34 products with generous estimates and 24 products with conservative estimates.

---

[407] The use of the terms *common, infrequent, rare* and *very rare* are defined per The Council for International Organizations of Medical Sciences (CIOMS) Working Group III, *Guidelines for Preparing Core Clinical Safety Information on Drugs.*

[408] Preciosa M Coloma et al., "Electronic Healthcare Databases for Active Drug Safety Surveillance: Is There Enough Leverage?," *Pharmacoepidemiology and Drug Safety* 21, no. 6 (June 2012): 611–621.

[409] Fireman et al., "A Protocol for Active Surveillance of Acute Myocardial Infarction in Association with the Use of a New Antidiabetic Pharmaceutical Agent."

9.2.2 Infrequent Event Rates

Table 23 in Appendix B shows results with a different event rate assumption: 1 event/1000 person-years. A comparable event rate was found in a recent study of bipolar and depressed adults using anti-epileptics, which estimated 1.2-4.4 suicide-related events/1000 person-years.[410] Under this event rate assumption, I found that sequential database surveillance on 8 quarters of data could identify a simulated incidence rate ratio of ≥2.5 for 16 products with the most generous estimates of exposures and for 7 products with the most conservative estimates. Using 15 quarters of data, sequential database surveillance would have identified a simulated incidence rate ratio of ≥2.5 for 21 products with generous estimates and 15 products with conservative estimates.

9.2.3 Rare Event Rates

Table 24 in Appendix B repeats the analysis with a rare event rate: 1 event/10,000 person-years. Sequential database surveillance could identify a simulated IRR of ≥2.5 for 5 products within 8 quarters and 8 products within 15 quarters under generous exposure estimate assumptions. For reference, the rate of acute renal failures among statin users has been estimated at 4 events/10,000 person-years[411]; and the rate of upper gastrointestinal bleeding episodes among users of non-steroidal anti-inflammatory drugs has been estimated at ranges from 3.9 to 11 events/10,000 person-years.[412]

9.2.4 Very Rare Event Rates

I performed the analysis with very rare event rates - 1 event/100,000 person-years – and the results are shown in Table 25 in Appendix B. Comparable rates are 1-2 cases of Guillain-Barré Syndrome among 100,000 person-years contributed by adolescents

---

[410] A. Arana et al., "Suicide-related Events in Patients Treated with Antiepileptic Drugs," *The New England Journal of Medicine* 363, no. 6 (2010): 542–551.

[411] L. A. Garcia Rodriguez, R. Herings, and S. Johansson, "Use of Multiple International Healthcare Databases for the Detection of Rare Drug-associated Outcomes: a Pharmacoepidemiological Programme Comparing Rosuvastatin with Other Marketed Statins," *Pharmacoepidemiology and Drug Safety* 19, no. 12 (2010): 1218–1224.

[412] P. M. Coloma et al., "Combining Electronic Healthcare Databases in Europe to Allow for Large-scale Drug Safety Monitoring: The EU-ADR Project," *Pharmacoepidemiology and Drug Safety* 20, no. 1 (2011): 1–11.

eligible for vaccination against meningococcal disease.[413] A simulated incidence rate ratio of 2.5 is not achievable for any products after 15 quarters of data have accrued even under the most generous exposure assumptions.

9.3 Utilization Comparison Among Three Data Sources

To understand the greater context of these results and also how they compared to data actually in the Mini-Sentinel System, I obtained comparable data from a proprietary drug use database licensed by the FDA and summary tables[414] from the Mini-Sentinel System.

9.3.1 SDI Vector One®: National (VONA)

The VONA database measures retail dispensing of prescriptions. Prescriptions are captured from a sample of approximately 59,000 retail pharmacies throughout the U.S. These data are then presented as nationally projected monthly dispensing summaries. Dispensings from mail order pharmacies and non-retail settings are not represented in these data. Prior to the availability of the Mini-Sentinel System, these data were the primary source of utilization data for the FDA. Generally, these data are considered reliable for outpatient pharmacy dispensings that are not dispensed in specialty pharmacies or via restricted distribution programs. For medical products with limited/restricted dispensing patterns, these projections are less reliable because it is unclear how representative the sample is or how well this sample can be nationally projected in these circumstances.

9.3.2 Mini-Sentinel System Summary Table Data

Once it became available, I obtained Mini-Sentinel System Summary Table data on the cohort. These data were also quarterly dispensings for prevalent users. At that time, enrollment in the Mini-Sentinel System was estimated to cover 38.8 million persons. However, some data partners were not able to contribute data until after 2007, and because I was concerned with exposure trends immediately post-approval, I censored any

[413] Priscilla Velentgas et al., "Risk of Guillain-Barré Syndrome After Meningococcal Conjugate Vaccination," *Pharmacoepidemiology and Drug Safety* (July 16, 2012), http://www.ncbi.nlm.nih.gov/pubmed/22807266.
[414] See section 2.2.2.1 for a description of summary tables.

data partners that did not contribute to the system for the entire evaluation period (2004-2010). After censoring, my reduced version of the data was estimated to cover 10 million persons.

9.3.3 Correlation

Unfortunately, these three datasources (i.e., Medicaid, SDI Vector One®, and the Mini-Sentinel System) uniformly do not support calculation of dispensings per capita, and the estimated persons covered by the three diverge greatly. Therefore, to compare the systems and evaluate their correlation, I standardized them so that each quarter's utilization (U_t) is equal to the percent of utilization out of the total utilization over four years using equation (7) where R_t is raw dispensings per quarter.

$$U_t = \frac{R_t}{\sum_{t=1}^{16} R_t} \tag{7}$$

I chose this formula because it is easy to understand and it prioritizes utilization patterns early in a medical product's lifecycle, which would presumably be most important to understand for support of sequential database surveillance. This reduced the total cohort to 37 new molecular entities because not all of the medical products had sixteen quarters of data across the three sources. Figure 29 shows these 37 results, each medical product is a different color. Individual medical products are not identified to due to limitations on public sharing of these data.

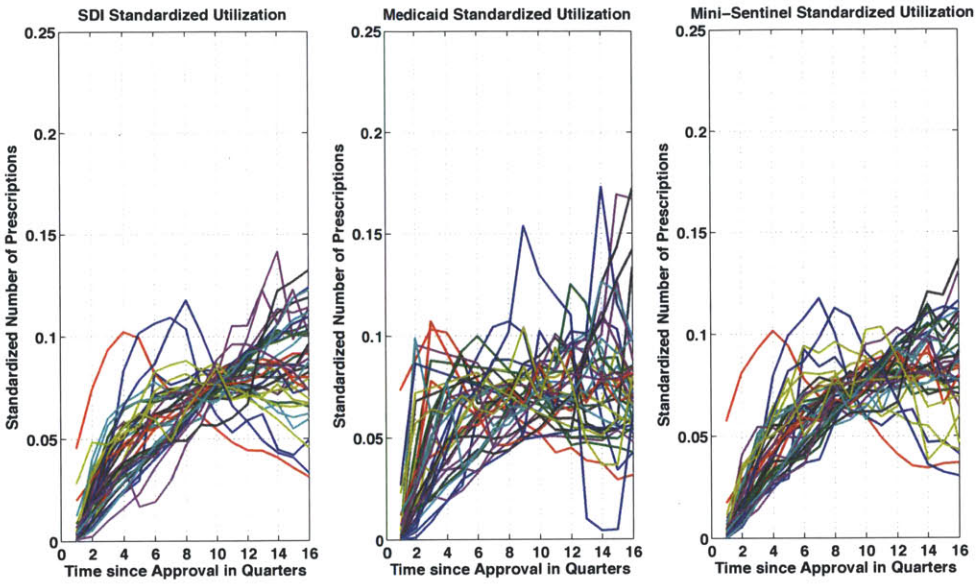

Figure 29. Standardized Utilization Patterns for 37 New Molecular Entities Over Four Years.

To calculate the correlation, I treated each utilization pattern as a sequence of values and calculated the pairwise Pearson correlation distance.[415] Table 14 shows these results, roughly ordered by products with the best correlation among the three data sources to products with the least correlation. Lower numbers indicate a better correlation and are color scaled such that green indicates a better correlation than red.

First, Medicaid dispensing data is more dissimilar from the other two data sources. This dissimilarity tends to be worse in new molecular entities with dispensings that spanned the significant changes in Medicaid that occurred after January 1, 2006. New molecular entities that do not cover this span (i.e., were approved after January 1, 2006) are noted with an asterisk in Table 14. Second, the mean is greater than the median for all cases, indicating the presence of outliers associated with lower levels of correlation. Third, the areas of dissimilarity between SDI Vector One® and Mini-Sentinel System data mostly involve medical products with restricted distribution programs (e.g., Revlimid®, Sutent®, Exjade®, Nexavar®, Zolinza®) where it is assumed that SDI Vector One® data will undercount actual utilization. Fourth, correlation tends to be worse among products with low exposure prevalence due to small total numbers of dispensings.

[415] For formula, see http://www.mathworks.com/help/stats/pdist.html "correlation distance."

Trade Name	Generic Name	SDI-MS	MS-MC	SDI-MC
Tindamax	tinidazole	0.0059	0.0185	0.0250
Ranexa*	ranolazine	0.0196	0.0255	0.0475
Byetta	exenatide	0.0063	0.0388	0.0525
Symlin	pramlintide	0.0155	0.0476	0.0492
Apidra	insulin glulisine	0.0048	0.0630	0.0745
Aptivus	tipranavir	0.0278	0.0787	0.0385
Amitiza*	lubiprostone	0.0150	0.0493	0.0943
Prezista*	darunavir	0.0104	0.0725	0.0820
Baraclude	entecavir	0.0111	0.0929	0.0702
Revlimid	lenalidomide	0.0613	0.0506	0.0724
Sprycel*	dasatinib	0.0374	0.0768	0.0787
Levemir	insulin detemir	0.0059	0.0892	0.1038
Omacor/Lovaza	omega-3 acid ethyl ester	0.0109	0.0738	0.1253
Rozerem	ramelteon	0.0395	0.0774	0.1222
Sutent*	sunitinib malate	0.1023	0.0758	0.0868
Vesicare	solifenacin	0.0015	0.1277	0.1374
Azilect*	rasagiline mesylate	0.0263	0.1324	0.1145
Chantix*	varenicline	0.0066	0.1371	0.1297
Januvia*	sitagliptin phosphate	0.0077	0.1655	0.1714
Noxafil*	posaconazole	0.0653	0.1067	0.1952
Invega*	paliperidone	0.1417	0.0575	0.1713
Campral	acamprosate	0.0178	0.1814	0.1846
Exjade	deferasirox	0.2277	0.1584	0.0271
Tyzeka*	telbivudine	0.0224	0.2499	0.1591
Nexavar	sorafenib	0.2099	0.1540	0.1126
Xifaxan	rifaximin	0.0066	0.2107	0.2678
Cymbalta	duloxetine	0.0023	0.3692	0.3834
Enablex	darifenacin	0.0205	0.3682	0.3688
Lunesta	eszopiclone	0.0347	0.3691	0.5474
Tarceva	erlotinib	0.0205	0.4551	0.5289
Lyrica	pregabalin	0.0044	0.5220	0.5340
Zolinza*	vorinostat	0.3052	0.5928	0.2616
Nevanac	nepafenac	0.0205	0.5564	0.6113
Sanctura	trospium	0.0178	0.5770	0.6087
Spiriva	tiotropium	0.0060	0.9272	0.9900
Fosrenol	lanthanum	0.0446	0.9284	1.1247
Sensipar	cinacalcet	0.0053	1.3822	1.4408
Mean		0.0429	0.2611	0.2755
Median		0.0178	0.1324	0.1297
Standard Deviation		0.0689	0.3046	0.3274

Table 14. Pearson Correlation Distances Among New Molecular Entities in the Cohort
Abbreviations: MS: Mini-Sentinel System data; MC: Medicaid data.

However, these data suggest that FDA may be able to use SDI data and/or Mini-Sentinel System data to support adoption and utilization models. However, fitting these data to a few functional forms proved quite challenging and essentially non-productive. Consequently, I moved away from use of these data for this reason and two others explained below.

First, raw dispensings are a less than ideal measure of adoption and utilization patterns because of the inability to distinguish "incident users" or new adopters from continuing users. Sequential database surveillance activities are generally planned with incident user criteria such that a person typically only contributes information to the surveillance based on their initial adoption and use of the product, i.e. during their "initial treatment episode."[416] Incident user designs are an important technique for mitigation of selection bias. Second, I utilize theoretical insights and functional forms from the diffusion of innovations literature, which is briefly described next. Generally, this literature focuses on initial adoption of new products, although there has been some work on repeat purchases.[417] Thus, incident user data are a better fit with these theoretical constructs.

9.4 Mini-Sentinel System Incident User Data

I obtained Mini-Sentinel System incident user data on the cohort described above. Again, because some data partners did not have complete data for the entire evaluation period (2004-2010), I removed these data partners from formal analysis although I still observed their trend data visually. The remaining subset of the dataset is dominated by one data partner that accounts for 80-90% of the data, depending on the medical product. Had I been looking at a different evaluation period (i.e., after January 1, 2010), this would not be the case.

[416] For more on treatment episodes, see Mini-Sentinel Operations Center, "Module 3: Drug Use - Incident Outcomes."
[417] Minhi Hahn et al., "Analysis of New Product Diffusion Using a Four-Segment Trial-Repeat Model," *Marketing Science* 13, no. 3 (July 1, 1994): 224–247.

9.4.1 Diffusion of Innovations Literature – Bass Models

There is an extensive literature in the diffusion of innovations[418], and complete coverage of that literature is beyond the scope of this research. However, modeling the adoption (and subsequent) utilization patterns of new medical products begins as a diffusion-related problem, and so it is logical to begin with this literature to find functional forms to describe new medical product adoption. The most prevalent model in this literature is the Bass model, which was introduced in 1969[419] and has been extended by numerous authors.[420] Bass formulates cumulative adoptions as an S-shaped, or sigmoid curve, characterized by adoption rates that rise and then fall over time, with the slowest rates occurring at the beginning and end of the adoption period. Other sigmoid curves that may be familiar to the reader include the symmetric logistic function and the asymmetric Gompertz function[421], which have been used widely in ecology and biology to describe biological processes such as predator-prey relations and tumor growth.

The Bass model's original premise was to describe the adoption of new durable goods (e.g., televisions, cars) among a stable, homogenous adopter population that was unlikely to exhibit repeat purchase behavior. Equation (8) is Bass's cumulative density function for adoption (i.e., the cumulative probability of adoption up to time *t)*. Equation (9) is the hazard function, and is the ratio of the probability density function to the survival function. The hazard function represents the instantaneous probability of adoption at time *t* given that one has not already adopted. Note that all potential adopters will adopt in Bass's formulation so it is important to exclude known non-adopters (e.g., perhaps those with a contraindication) at the outset. Bass hypothesizes the existence of two types of adopters that contribute to the overall adoption pattern for a particular product although he does not establish *a priori* which adopters belong to which groups. This is easiest to see in the hazard equation (9) and these two groups are the innovators and imitators.

---

[418] See generally Renana Peres, Eitan Muller, and Vijay Mahajan, "Innovation Diffusion and New Product Growth Models: A Critical Review and Research Directions," *International Journal of Research in Marketing* 27, no. 2 (June 2010): 91–106; Nigel Meade and Towhidul Islam, "Modelling and Forecasting the Diffusion of Innovation – A 25-year Review," *International Journal of Forecasting* 22, no. 3 (2006): 519–545; Everett M. Rogers, *Diffusion of Innovations*, 5th ed. (New York: Free Press, 2003).

[419] Frank M. Bass, "A New Product Growth for Model Consumer Durables," *Management Science* 15, no. 5 (January 1, 1969): 215–227.

[420] Peres, Muller, and Mahajan, "Innovation Diffusion and New Product Growth Models."

[421] Charles P. Winsor, "The Gompertz Curve as a Growth Curve," *Proceedings of the National Academy of Sciences of the United States of America* 18, no. 1 (January 15, 1932): 1–8.

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}[e^{-(p+q)t}]} \quad (8)$$

$$h(t) = \frac{f(t)}{1 - F(t)} = p_1 + \frac{q_1}{M} N(t) \quad (9)$$

The innovators (also called influentials), are accounted for in the first part of the hazard function, and their adoption is described by the coefficient of innovation known as p. Essentially, they adopt independently. Imitators, on the other hand, described by coefficient of imitation, q, adopt based on the decisions of others, as operationalized by the ratio of the number of adopters at a particular time ($N(t)$) to the total market size (M). This effect is alternatively referred to as the social contagion effect, word-of-mouth effect, etc. The Bass model also has been described as the Mixed Influence Model with much scholarship devoted to time-dependent marketing efforts aimed at either the innovators or imitators.[422]

Many extensions of the Bass Model employ a component function that accounts for "marketing mix" variables (e.g., amount of money spent on promotions) that might drive overall adoption patterns, and this is known as the Generalized Bass Model.[423] I do not make use of that model here because I do not have the data to support it, nor is such data likely to be available to the FDA at the time of launch of a new medical product. Other relevant extensions of the Bass Model include those that more formally segment the aforementioned two populations and describe their behaviors.[424] These models have been generically categorized as "two segment mixture models." Some authors describe these two segments (i.e., the early and late markets) as belonging to a single market and model one pooled market size. Others have posited two separate markets with two separate

---

[422] Vijay Mahajan and Robert A. Peterson, *Models for Innovation Diffusion*, Sage University Papers Series. Quantitative Applications in the Social Sciences no. 07-048 (Beverly Hills: Sage Publications, 1985).
[423] Frank M. Bass, Trichy V. Krishnan, and Dipak C. Jain, "Why the Bass Model Fits Without Decision Variables," *Marketing Science* 13, no. 3 (July 1, 1994): 203–223.
[424] See Demetrios Vakratsas and Ceren Kolsarici, "A Dual-market Diffusion Model for a New Prescription Pharmaceutical," *International Journal of Research in Marketing* 25, no. 4 (December 2008): 282–293; Christophe Van den Bulte and Yogesh V. Joshi, "New Product Diffusion with Influentials and Imitators," *Marketing Science* 26, no. 3 (May 1, 2007): 400–421; Jacob Goldenberg, Barak Libai, and Eitan Muller, "Riding the Saddle: How Cross-Market Communications Can Create a Major Slump in Sales," *Journal of Marketing* 66, no. 2 (April 1, 2002): 1–16.

market potentials. Complete separation of the adoption populations may be more appropriate to describe situations when manufacturers receive approval for new indications for their existing product. The approval occurs discretely in time although off-label use of the product in non-labeled indications may have preceded formal approval.

There have been multiple theories to describe why a segmentation might exist[425], and most of these theories focus on the innovators as risk takers, technology enthusiasts, or those that seek to send a signal of high social status by adopting new innovations very early in a product's lifecycle. Many of these theories are appropriate in the context of a new product that is a durable good, e.g., a new smartphone. However, there are other theories more appropriate to the context of new medical product adoption.

One theory referred to as *information transfer* describes the social contagion effect as being the result of information transfer from influentials who are described as "opinion leaders." Essentially, the opinion leaders/influentials adopt and develop experience by prescribing and monitoring the effects of the new medical product, thereby reducing uncertainties associated with its unknown risks and benefits in a non-clinical trial population. These opinion leaders/influentials then exhibit contagion behavior among themselves, and to the "imitator" physician adopters who become more willing to adopt the product once a body of evidence begins to develop with respect to the medical product's risks and benefits. Appropriate empirical testing of this theory would require modeling physician adopter networks, which I do not intend to do here, and others have done ably.[426] However, it is reasonable to assume that patient populations may be divided similarly. That is, some patients may be less risk-averse and more willing to try a new medical product based on its clinical trial profile.

---

[425] See summary in Bulte and Joshi, "New Product Diffusion with Influentials and Imitators."

[426] For example models, see Raghuram Iyengar, Christophe Van den Bulte, and Thomas W. Valente, "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science* 30, no. 2 (March 1, 2011): 195–212; Mark Paich, Corey Peck, and Jason Valant, "Pharmaceutical Market Dynamics and Strategic Planning: a System Dynamics Perspective," *System Dynamics Review (Wiley)* 27, no. 1 (March 2011): 47–63; Mary A. Burke, Gary M. Fournier, and Kislaya Prasad, "The Diffusion of a Medical Innovation: Is Success in the Stars?," *Southern Economic Journal* 73, no. 3 (January 1, 2007): 588–603; Christophe Van den Bulte and Gary L. Lilien, "Medical Innovation Revisited: Social Contagion Versus Marketing Effort," *American Journal of Sociology* 106, no. 5 (March 1, 2001): 1409–1435; James Samuel Coleman and University Columbia, *Medical Innovation; a Diffusion Study* (Indianapolis: Bobbs-Merrill Co, 1966).

Vakratsas et al. follow up this idea and postulate that this phenomenon may be particularly true for diseases that encompass varying levels of severity, e.g., diabetes, obesity, cardiovascular disease.[427] They postulate that the early market segment consists of individuals who are performing poorly on all existing therapies and so are willing to try new products as soon as they become available due to unmet medical needs. The late market segment of adopters consists of patients who may be satisfied with their current therapy, but may be more convinced to switch therapies as more information develops about the new product's risks and benefits. They completely separate these populations, i.e. their primary model is not pooled.

9.4.2 Model Fitting

In general, I adopt the functional forms first described by Bass and others cited in this section and attempt to fit the Mini-Sentinel System incident user data to these forms using the LSQCURVEFIT and non-linear model fitting functions available in MATLAB® (R2012a). The goal is to assess whether these functional forms appropriately describe the data, and how planners might use these forms to speculate as to the adoption and utilization patterns for new medical products being evaluated via sequential database surveillance. Throughout, I do not make use of the "marketing mix" components that could be present in any of these functional forms since the FDA typically does not have these data prior to the launch of the product.

As a note, I do not treat the adopter population as a "dynamic adopter population" as described by others.[428] These authors require additional modeling to describe the adopter population's growth as a natural outgrowth of population, i.e. as a continuous expansion. For example, as the population ages, there will likely be dynamic adoption of medical products that treat disease associated with older age like cardiovascular disease. Because the adoption timeframe that I am considering is relatively short, I omit these dynamics for simplicity.

Last, I narrowed my original cohort only to those medical products that had enough exposure that they might be reasonable candidates for sequential database surveillance.

---

[427] Vakratsas and Kolsarici, "A Dual-market Diffusion Model for a New Prescription Pharmaceutical."
[428] Vijay Mahajan and Robert A. Peterson, "Innovation Diffusion in a Dynamic Potential Adopter Population," *Management Science* 24, no. 15 (November 1, 1978): 1589–1597.

Therefore, I eliminated new molecular entities in the original cohort that did not have at least 10,000 incident users over the entire evaluation period. I chose 10,000 users because it mirrors a setpoint established by Congress in the 2007 Food and Drug Amendments Act, which requires the FDA to post a safety summary regarding the new product to its website.[429] This narrowed my cohort to 22 medical products, and the data that I fit was the monthly cumulative adopters who I required to be incident users. Incident users were defined as not having used the medical product in the 120 days preceding the first use, and only users with continuous drug coverage during that time were included.[430]

9.4.2.1 Single Market Model Forms

Model Form 1 (MF1) is shown as equation (10). The cumulative adopters ($N(t)$) is the market size (M) multiplied by the cumulative density function for the Bass Model, as shown in equation (8).

$$N(t) = M * F(t) \tag{10}$$

9.4.2.2 Dual Market Model Forms: Dual Bass

Model Form 2 (MF2) is a pooled dual-Bass market model. It consists of two Bass models coupled together with a simple, constant probability (π) to describe the likelihood of belonging to the first or second market, and is shown in equation (11). Again the cumulative density functions (F_1 and F_2) are shown in equation (8).

$$N(t) = [\,\pi F_1(t) + (1-\pi)F_2(t)] * M \tag{11}$$

Model Form 3 (MF3) is also a pooled dual-Bass market model as shown in equation (11). However, it uses a different formulation for the probability, a time-dynamic one that requires the early market to precede the later market and requires the highest probability

[429] § 915 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*..."[The Secretary via the FDA shall provide drug safety information to patients and prescribers by] preparing, by 18 months after approval of a drug or after use of the drug by 10,000 individuals, whichever is later, a summary analysis of the adverse drug reaction reports received for the drug, including identification of any new risks not previously identified, potential new risks, or known risks reported in unusual number;". Codified at 21 U.S.C. § 355(r)(2)(D).

[430] A 45-day enrollment gap was allowed.

of being part of the early market to occur earliest and then fade to zero. This formulation was first used in Vakratsas et al.'s paper[431]:

$$\pi(t) = \frac{e^{-\theta t}}{(1 + e^{-\theta t})} \tag{12}$$

Model Form 4 (MF4) is a non-pooled version (i.e., note the two M variables) of MF2, a dual-Bass market model, with a static probability. It is shown in equation (13).

$$N(t) = [\, \pi M_1 F_1(t) + (1 - \pi) M_2 F_2(t)] \tag{13}$$

Model Form 5 (MF5) is a non-pooled version of MF3 using the dynamic probability. It is represented as equation (13) with the probability equivalent to equation (12).

9.4.2.3 Dual Market Model Forms: Exponential-Bass Model

Vakratsas et al.'s main dual market models begin by linking together two Bass models.[432] However, they hypothesize that the early market is made up only of the innovator sub-group with no contribution from imitators. Thus, they set q in equation (8) *equal* to zero for the early market. The result is an early market described by an exponential cumulative density function and a later market described by the Bass cumulative density function. They also reason that the delay in the availability of the new medical product during the process of licensure creates a pent-up demand for the product that is well-modeled by an exponential adoption function.

They examine this dual-market combination as both a pooled and non-pooled market and with both a static and dynamic probability using the same sort of iterations discussed above. However, to ensure that the early market precedes the later market, their static probability formulation is slightly more complex. Essentially, they take the dynamic formulation listed in equation (12) and set time equal to 1. Thus, by constraining θ to be positive, the early market must be smaller than the later market.

Model Form 6 (MF6) is a pooled dual market model with static probability as described above where the first market is exponential and the second market is a Bass

[431] Vakratsas and Kolsarici, "A Dual-market Diffusion Model for a New Prescription Pharmaceutical."
[432] Ibid.

model. Model Form 7 (MF7) is a pooled version of this model with a dynamic probability as formulated in equation (12). Model Form 8 (MF8) is a non-pooled version of MF6. Model Form 9 (MF9) is a non-pooled version of MF7. Table 15 is a brief summary of these model forms.

Model Form	Description	Pooled?	Probability?	Parameters
MF1	Single Bass	N/A	N/A	M, p, q
MF2	Dual Bass	Yes	Static	M, p1, q1, π, p2, q2
MF3	Dual Bass	Yes	Dynamic	M, p1, q1, p2, q2, θ
MF4	Dual Bass	No	Static	M1, M2, p1, q1, π, p2, q2
MF5	Dual Bass	No	Dynamic	M1, M2, p1, q1, p2, q2, θ
MF6	Exponential-Bass	Yes	Static	M, p1, π, p2, q2
MF7	Exponential-Bass	Yes	Dynamic	M, p1, p2, q2, θ
MF8	Exponential-Bass	No	Static	M1, M2, p1, π, p2, q2
MF9	Exponential-Bass	No	Dynamic	M1, M2, p1, p2, q2, θ

Table 15. Summary of Model Functional Forms used with Mini-Sentinel System Incident User Data
Abbreviations: MF; Model Form.

9.4.3 Findings

9.4.3.1 Single Market Bass Model (MF1)

First, all uptake patterns could be fit to a Single Market Bass Model. With the exception of one medical product in the cohort, all parameter estimates were significant at the 0.05 level and all signs were in the correct direction. As measures of the goodness of model fit, I list mean squared error (MSE), median absolute deviation from the median (MAD), and the Bayesian Information Criterion (BIC) in Table 16. These metrics are consistent with what others have previously reported.[433] R^2 and adjusted R^2 are usually quite high and generally not informative (i.e., nearly all are greater than .995 and some are 1.)

[433] Ibid.

	MSE	MAD	BIC
Amitiza	1.55E+06	867.98	983.27
Byetta	8.40E+05	707.49	1113.68
Campral	9.53E+03	51.95	873.84
Chantix	5.71E+07	5071.88	1105.94
Cymbalta	7.05E+06	1559.70	1442.67
Enablex	3.36E+05	443.57	1114.69
Invega	3.43E+04	138.36	646.10
Januvia	3.02E+06	1203.80	914.44
Levemir	8.04E+04	235.58	828.91
Lunesta*	1.42E+07	2283.34	1341.92
Lyrica	1.95E+06	1129.92	1118.05
Nevanac	4.08E+05	435.17	1018.00
Omacor/Lovaza	7.01E+05	582.82	1036.18
Ranexa	2.80E+04	113.78	767.73
Rozerem	4.34E+05	594.25	1021.94
Sanctura	8.76E+04	249.99	1104.87
Sensipar	8.71E+03	45.36	974.94
Spiriva	5.67E+06	1425.23	1481.26
Tarceva	3.11E+03	33.51	814.97
Tindamax	7.14E+03	47.14	923.56
Vesicare	4.13E+05	380.56	1145.21
Xifaxan	6.14E+05	514.99	1271.05

*one parameter not significant at the 0.05 level

Table 16. Measures of Model Fit for New Molecular Entity Cohort with Single Market Bass Model
Abbreviations: MSE, Mean Squared Error; MAD, Median Absolute Deviation from the Median; BIC, Bayesian Information Criterion.

9.4.3.2 Dual Market Bass Models (MF2-MF5)

Generally, the MSE, MAD, and BIC were lower for the dual market Bass models than for the single market bass models and the results of the dual market Bass models are shown in Table 17. However, many of the parameters in the various model fits were not significant at the 0.05 level. Models in which all the parameters were significant are marked with an asterisk. MF3 (i.e., the pooled, dynamic market model) had a particularly hard time fitting the θ parameter (i.e., the dynamic probability of belonging to one market or the other) and MF4 (the non-pooled static market model) could fit neither of the market sizes nor the π parameter at a significant level. The signs of all the parameters were in the right direction but I had to relax the constraint requiring θ to be positive, meaning that some models fit better when the early market was larger than the latter market.

	MSE				MAD				BIC			
	MF2	MF3	MF4	MF5	MF2	MF3	MF4	MF5	MF2	MF3	MF4	MF5
Amitiza	2.00E+05	2.60E+05	2.04E+05	2.04E+05	265.03	327.22	265.01	263.34	875.71	890.65	879.88	879.88
Byetta	2.02E+05	1.82E+05	2.05E+05	1.14E+05	249.95	209.83	250.05	168.54	1027.71	1020.98	1032.02	992.61
Campral	6.64E+03	6.30E+03	6.75E+03	2.70E+03	34.57	36.19	34.58	33.10	857.70	853.82	862.08	796.05
Chantix	1.49E+06	1.46E+06	1.52E+06	1.50E+06	825.73	768.84	825.62	770.91	921.58	920.68	925.69	924.79
Cymbalta	6.94E+06	6.76E+06	7.02E+06	4.31E+06	891.64	976.29	946.65	773.48	1451.52	1449.57	1455.82	1418.30
Enablex*	1.15E+04	1.29E+04	1.17E+04	1.14E+04	56.46	68.95	56.46	53.89	885.14	893.19	889.50	887.76
Invega	5.14E+03	5.45E+03	5.27E+03	4.63E+03	49.19	39.72	49.19	36.38	563.61	566.37	567.63	561.50
Januvia	6.61E+05	7.56E+05	6.77E+05	6.78E+05	713.69	642.85	713.74	500.96	845.84	852.66	849.91	850.05
Levemir*	1.38E+04	1.38E+04	1.41E+04	1.37E+04	52.03	51.99	52.03	67.37	735.90	735.90	740.09	738.68
Lunesta*	2.09E+06	1.55E+06	2.12E+06	1.28E+06	1055.01	854.11	1055.01	812.08	1219.31	1198.66	1223.65	1188.95
Lyrica	1.10E+06	9.32E+05	1.33E+06	1.62E+06	577.86	499.76	499.42	864.39	1090.78	1080.27	1105.96	1118.76
Nevanac	1.15E+05	1.01E+05	1.17E+05	6.07E+04	168.22	143.07	165.65	143.85	946.39	937.77	950.56	908.65
Omacor/Lovaza	2.66E+05	1.21E+05	2.71E+05	1.22E+05	352.63	260.84	352.73	258.67	984.64	934.78	988.90	938.40
Ranexa*	2.10E+03	3.62E+03	2.14E+03	2.06E+03	19.53	43.16	19.53	20.87	626.69	658.30	630.87	628.56
Rozerem	8.58E+04	4.50E+05	8.73E+04	7.07E+04	109.28	586.98	109.24	129.94	927.63	1033.60	931.90	918.41
Sanctura*	1.13E+04	1.75E+04	1.14E+04	1.83E+04	59.11	80.16	59.12	75.31	957.00	990.70	961.44	997.81
Sensipar	7.27E+03	7.12E+03	3.63E+03	6.70E+03	45.45	45.27	40.58	48.25	970.48	968.73	917.51	967.19
Spiriva*	2.42E+06	1.55E+06	2.46E+06	1.36E+06	1418.04	1028.39	1418.05	910.32	1423.41	1387.74	1427.88	1380.63
Tarceva	1.37E+03	1.41E+03	1.44E+03	8.80E+02	14.84	16.60	16.93	16.88	764.39	766.23	771.43	734.82
Tindamax	6.74E+03	6.03E+03	6.84E+03	5.74E+03	49.74	33.62	48.51	44.36	929.17	920.45	933.59	919.91
Vesicare*	3.29E+04	3.49E+04	3.34E+04	3.03E+04	115.03	96.49	115.02	120.29	972.85	977.13	977.22	970.12
Xifaxan*	1.43E+04	1.33E+04	1.96E+04	1.32E+04	60.27	58.67	91.74	55.48	987.60	982.09	1015.88	985.11

*At least one of the Dual Bass Model Forms had all significant parameters

Table 17. Measures of Model Fit for New Molecular Entity Cohort with Dual Market Bass Model Forms

Abbreviations: MSE, Mean Squared Error; MAD, Median Absolute Deviation from the Median; BIC, Bayesian Information Criterion.

The implication that a dual-market model is a better fit than the single market model, particularly for models in which all parameters are significant, is an important finding epidemiologically-speaking for sequential database surveillance. Sequential database surveillance is envisioned as a technique to be used soon after licensure of a new medical product. Thus, surveillance conclusions would likely be based on an "early market" for the new medical product. If this early market of users were substantially different than the later market of users, then there may be less "transportability"[434] of sequential database surveillance findings to these later users. Additionally, the dual market model may indicate the presence of channeling bias. Regulatory action resulting from these findings likely needs to be more tailored than perhaps originally envisioned. Finally, because sequential database surveillance may be performed with less covariate control than a traditional retrospective epidemiologic study, it may be more difficult to identify characteristics that tend to place a user in the "early" v. "late" markets.

[434] Recall that transportability is a preferred term in pharmacoepidemiology but is generally equivalent to the more familiar, external validity or generalizability.

9.4.3.3 Dual Market Exponential-Bass Models (MF6-MF9)

In general, most of these data visually do not support an early market exponential model. Consequently, there were significant sign problems associated with fitting the dual market exponential-Bass model in many fits. Most commonly, *p1* was negative, which essentially creates a single market Bass model. Again, many of the parameters in the various model fits were not significant at the 0.05 level. In MF8 and MF9, there was not a single new molecular entity with parameters that were all significant. The results are shown in Table 18. I show results as "NA" if the model did not have correct signs since those fits did not really reflect an exponential-Bass model.

	MSE				MAD				BIC			
	MF6	MF7	MF8	MF9	MF6	MF7	MF8	MF9	MF6	MF7	MF8	MF9
Amitiza	1.19E+06	#N/A	1.20E+06	#N/A	630.74	#N/A	602.25	#N/A	974.58	#N/A	977.97	#N/A
Byetta	3.06E+05	#N/A	3.10E+05	3.05E+05	251.14	#N/A	251.88	233.25	1052.51	#N/A	1056.56	1055.38
Campral	6.88E+03	#N/A	6.91E+03	6.91E+03	32.99	#N/A	34.14	33.31	856.92	#N/A	860.49	860.49
Chantix	6.48E+06	#N/A	6.62E+06	5.42E+06	1839.19	#N/A	1839.17	1461.89	996.54	#N/A	1000.63	990.03
Cymbalta*	5.27E+06	6.86E+06	5.34E+06	5.31E+06	1086.32	1358.11	1086.28	1084.39	1426.99	1447.26	1431.42	1430.97
Enablex	1.74E+05	#N/A	1.73E+05	2.29E+05	309.27	#N/A	305.68	365.86	1074.62	#N/A	1077.37	1097.25
Invega	1.08E+04	#N/A	1.08E+04	1.07E+04	41.25	#N/A	42.02	43.77	596.17	#N/A	599.12	598.72
Januvia	9.69E+05	#N/A	9.82E+05	9.48E+05	478.70	#N/A	483.67	471.50	862.38	#N/A	866.00	864.21
Levemir	#N/A	#N/A	8.51E+04	5.43E+04	#N/A	#N/A	235.58	158.02	#N/A	#N/A	841.34	815.29
Lunesta*	4.68E+06	#N/A	4.76E+06	3.73E+06	1319.87	#N/A	1320.04	1135.36	1271.78	#N/A	1276.10	1259.40
Lyrica*	1.29E+06	1.66E+06	1.31E+06	7.99E+05	563.80	992.93	563.84	434.50	1097.81	1113.92	1102.06	1070.41
Nevanac	#N/A	#N/A	4.30E+05	3.15E+05	#N/A	#N/A	435.17	306.34	#N/A	#N/A	1030.70	1010.85
Omacor/Lovaza	#N/A	#N/A	#N/A	2.36E+05	#N/A	#N/A	#N/A	285.43	#N/A	#N/A	#N/A	976.99
Ranexa	#N/A	#N/A	2.96E+04	1.14E+04	#N/A	#N/A	113.78	79.50	#N/A	#N/A	780.16	724.99
Rozerem*	5.71E+04	6.39E+04	5.81E+04	5.61E+04	119.23	114.06	119.23	87.73	898.43	905.62	902.68	900.45
Sanctura	1.81E+04	2.23E+04	1.84E+04	1.83E+04	32.69	87.27	32.69	32.44	990.10	1006.27	994.52	994.51
Sensipar	3.55E+03	8.33E+03	3.60E+03	3.60E+03	40.85	39.44	40.85	40.84	909.03	978.07	913.49	913.49
Spiriva*	3.53E+06	3.15E+06	3.57E+06	3.29E+06	1564.48	1477.37	1564.69	1453.34	1450.01	1441.02	1454.47	1447.90
Tarceva	1.52E+03	#N/A	1.49E+03	1.41E+03	14.88	#N/A	16.11	16.18	768.49	#N/A	770.56	766.63
Tindamax	#N/A	6.41E+03	#N/A	6.50E+03	#N/A	39.99	#N/A	39.99	#N/A	921.90	#N/A	926.33
Vesicare	#N/A	#N/A	4.32E+05	1.28E+05	#N/A	#N/A	380.56	221.19	#N/A	#N/A	1158.24	1070.67
Xifaxan	5.09E+05	#N/A	6.42E+05	4.69E+05	514.99	#N/A	510.30	450.23	1263.04	#N/A	1284.52	1260.05

*Indicates at least one Exponential-Bass Dual Market Model had all significant parameters

Table 18. Measures of Model Fit for New Molecular Entity Cohort with Dual Market Exponential-Bass Model Forms
Abbreviations: MSE, Mean Squared Error; MAD, Median Absolute Deviation from the Median; BIC, Bayesian Information Criterion.

In general, regardless of model form, each uptake pattern can be described with one of the models described above. However, in this retrospective analysis, I have the benefit of fitting the regression on four years worth of uptake data. In modeling and simulating future sequential database surveillance on a completely new product, the user may have little or no data to rely on to speculate about uptake. Thus, it is important to use a model form that may be fit on very little data. Future work should more thoroughly examine these model fits in the presence of little data and assess how well they forecast future utilization.

Appendix C contains individual figures with the best model fits displayed. The pattern and the model fits are observable but the cumulative adoption numbers are deliberately removed because they are not authorized to be shared publicly.

9.5 Summary

In general, there is much work to be done with respect to modeling adoption/uptake functions, particularly when challenged with a new molecular entity. The main idea of modeling adoption and utilization is to support the simulation of sequential database surveillance scenarios, and sensitivity analyses on adoption and utilization functions are appropriate in these circumstances.

THIS PAGE INTENTIONALLY LEFT BLANK

10 SUMMARY OF FINDINGS AND FUTURE DIRECTIONS

10.1 Summary

This research develops qualitative and quantitative tools to aid the U.S. Food and Drug Administration (FDA) in evaluating the Mini-Sentinel System's sequential database surveillance capabilities to support regulatory decision-making. The qualitative tool – the Mini-Sentinel System Pre-Screening Checklist – is designed to help determine whether the Mini-Sentinel System is well suited, *on its face*, to evaluate a pre-specified exposure-outcome pair. This checklist does not provide definitive answers but rather is intended to prompt thoughtful analysis as a first step. Necessary inputs to the Mini-Sentinel System Pre-screening Checklist are: 1) a tracked safety issue that identifies a particular exposure-outcome pair of interest; and 2) a regulator's goal with respect to the strength of causal inference necessary to support regulatory decision-making.

The quantitative tool is a Sequential Database Surveillance Simulator that allows the user to explore virtually whether sequential database surveillance of a particular exposure-outcome pair is likely to generate evidence to identify and assess safety risks in a timely manner to support regulatory decision-making. The simulator is intended to be a learning tool that allows regulators/investigators to explore the many potential surveillance scenarios they could face. Specifically, the tool is designed to allow the regulator/public health investigator to explore the performance limitations and capabilities of sequential database surveillance *virtually* and in a *low-cost* way. That is, in this planning stage, there is no need to "learn-by-doing" while expending public health resources. In general, this tool is not intended to be strictly predictive or to forecast exactly how sequential database surveillance of a particular tracked safety issue will occur. It is intended to allow regulators to explore possible scenarios they may face in the hopes that they may learn through experimentation how to more precisely deploy the evidence generation capabilities of the Mini-Sentinel System, and further refine their assessments of its sufficiency for evidence generation.

By using a simple vaccine example to illustrate the use of the simulator, this dissertation also demonstrates the tradeoffs associated with sample size calculations in sequential statistical analysis, particularly the tradeoff between statistical power and median sample size. In some circumstances, decreasing statistical power (i.e., increasing

type II error and the chance of a missed signal) results in faster detection of a signal (i.e., smaller median sample sizes). Is this tradeoff worth it? Should a user take on the bigger risk of missing the signal if it means they can find it faster? These tradeoffs have more concrete meaning by translating the information time concepts into calendar time. With an understanding of the time to takes to detect a signal of excess risk in calendar time, it is possible for the user to estimate the potential public health harm that may occur depending on the speed and confidence with which a safety signal is detected or ruled out.

Second, the dissertation demonstrates differences in performance between various surveillance configurations that are possible when using distributed database systems. Specifically, I look at the performance of two continuous sequential testing methods in the Mini-Sentinel System. However, I also address the ability to reconfigure the Mini-Sentinel System into component configurations, particularly segregating the component databases by their data types.

Third, the dissertation demonstrates the effects of misclassification error on sequential database surveillance, and specifically how such errors may be accounted for in the design of surveillance. I find that imperfect positive predictive value has more strongly deleterious effects on statistical power whereas imperfect sensitivity can significantly increase the median and maximum sample sizes required to end surveillance.

Fourth, this dissertation considers the complexities of modeling *new* medical product adoption, and specifically, the existence of a "dual market" phenomenon for these new medical products. This finding raises a non-trivial generalizability concern regarding evidence generated via sequential database surveillance when performed immediately post-licensure.

10.2 Future Work

The current version of the Sequential Database Surveillance Simulator accommodates two specific sequential statistical models that have been frequently used in prior vaccine safety surveillance. I began this process using these models because they were well-established in this still developing field. However, this simulator could be built out to

accommodate other models, such as the group sequential models reviewed in subsection 5.2.2. An important aspect of future work will be to increase representation of group sequential models to better under their comparative performance characteristics.

The Mini-Sentinel System is still developing modular programs aimed at better understanding drug utilization patterns after a drug has been adopted. These data will be important for modeling continuous exposures. In this dissertation, I have focused on point exposures as a proof-of-concept. However, the simulator will remain a quite limited tool without expansion into continuous exposures. Once these modular programs detailing utilization patterns are complete, the simulator can be extended to support tracked safety issues with continuous exposures.

Research on modeling adoption patterns has just begun. Important future steps will be to show the predictive power of various functional forms of adoption that are based on little or no data. That is, it will be important to segregate adoption patterns into general categories and then try to understand *a priori* the contributing factors that led to that adoption pattern.

Finally, as the FDA considers the larger scope of exposure-outcome pairs that it needs to evaluate and uses the Sequential Database Surveillance Simulator repeatedly, it will be possible to establish overall demand for this capability. Understanding the level of demand for this capability and the others associated with the Mini-Sentinel System is an important determinant in public policy with respect to its ongoing public funding. That is, a more reasoned annual budget can be devoted to this piece of infrastructure when it becomes clear how often and for what purposes it is used.

THIS PAGE INTENTIONALLY LEFT BLANK

11 REFERENCES

Ahmad, S. R. "Adverse Drug Event Monitoring at the Food and Drug Administration." *Journal of General Internal Medicine* 18, no. 1 (2003): 57–60.

ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial. "Major Outcomes in High-risk Hypertensive Patients Randomized to Angiotensin-converting Enzyme Inhibitor or Calcium Channel Blocker Vs Diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT)." *Journal of the American Medical Association* 288, no. 23 (2002): 2981–2997.

Almenoff, J S, E N Pattishall, T G Gibbs, W DuMouchel, S J W Evans, and N Yuen. "Novel Statistical Tools for Monitoring the Safety of Marketed Drugs." *Clinical Pharmacology and Therapeutics* 82, no. 2 (August 2007): 157–166.

American Medical Association. "CPT® Process - How a Code Becomes a Code", n.d. http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt/cpt-process-faq/code-becomes-cpt.page.

Arah, Onyebuchi A, Yasutaka Chiba, and Sander Greenland. "Bias Formulas for External Adjustment and Sensitivity Analysis of Unmeasured Confounders." *Annals of Epidemiology* 18, no. 8 (August 2008): 637–646.

Arana, A., C. E. Wentworth, J. L. Ayuso-Mateos, and F. M. Arellano. "Suicide-related Events in Patients Treated with Antiepileptic Drugs." *The New England Journal of Medicine* 363, no. 6 (2010): 542–551.

Avorn, J. "In Defense of Pharmacoepidemiology--embracing the Yin and Yang of Drug Research." *The New England Journal of Medicine* 357, no. 22 (2007): 2219–2221.

Baggs, J., J. Gee, E. Lewis, G. Fowler, P. Benson, T. Lieu, A. Naleway, et al. "The Vaccine Safety Datalink: a Model for Monitoring Immunization Safety." *Pediatrics* 127 Suppl 1 (2011): S45–53.

Ball, R., D. Horne, H. Izurieta, A. Sutherland, M. Walderhaug, and H. Hsu. "Statistical, Epidemiological, and Risk-assessment Approaches to Evaluating Safety of Vaccines Throughout the Life Cycle at the Food and Drug Administration." *Pediatrics* 127 Suppl 1 (2011): S31–8.

Bass, Frank M. "A New Product Growth for Model Consumer Durables." *Management Science* 15, no. 5 (January 1, 1969): 215–227.

Bass, Frank M., Trichy V. Krishnan, and Dipak C. Jain. "Why the Bass Model Fits Without Decision Variables." *Marketing Science* 13, no. 3 (July 1, 1994): 203–223.

Behrman, R. E., J. S. Benner, J. S. Brown, M. McClellan, J. Woodcock, and R. Platt. "Developing the Sentinel System - A National Resource for Evidence Development." *The New England Journal of Medicine* (2011).

Belongia, E. A., S. A. Irving, I. M. Shui, M. Kulldorff, E. Lewis, R. Yin, T. A. Lieu, et al. "Real-Time Surveillance to Assess Risk of Intussusception and Other Adverse Events After Pentavalent, Bovine-Derived Rotavirus Vaccine." *The Pediatric Infectious Disease Journal* 29, no. 1 (2010): 1–5.

Berkman, Nancy D N D, Kathleen N K N Lohr, Laura C L C Morgan, Emily E Richmond, Tzy-Mey T M Kuo, Sally S Morton, Meera M Viswanathan, Douglas D Kamerow, Sue S West, and Elizabeth E Tant. *Reliability Testing of the AHRQ EPC Approach to Grading the Strength of Evidence in Comparative Effectiveness Reviews*. AHRQ Methods for Effective Health Care. Rockville (MD): Agency for Healthcare Research and Quality (US), 2012. http://www.ncbi.nlm.nih.gov/pubmed/22764383.

Berlin, Jesse A., M. Soledad Cepeda, and Carin J. Kim. "The Use of Meta-analysis in Pharmacoepidemiology." In *Pharmacoepidemiology*, edited by Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy, 723–756. Fifth. John Wiley & Sons, 2011.

Brenner, H., and O. Gefeller. "Use of the Positive Predictive Value to Correct for Disease Misclassification in Epidemiologic Studies." *American Journal of Epidemiology* 138, no. 11 (1993): 1007–1015.

Brewer, T., and G. A. Colditz. "Postmarketing Surveillance and Adverse Drug Reactions: Current Perspectives and Future Needs." *Journal of the American Medical Association* 281, no. 9 (1999): 824–829.

Brookhart, M Alan, Jeremy A Rassen, and Sebastian Schneeweiss. "Instrumental Variable Methods in Comparative Safety and Effectiveness Research." *Pharmacoepidemiology and Drug Safety* 19, no. 6 (June 2010): 537–554.

Brookhart, M Alan, Philip S Wang, Daniel H Solomon, and Sebastian Schneeweiss. "Evaluating Short-term Drug Effects Using a Physician-specific Prescribing Preference as an Instrumental Variable." *Epidemiology (Cambridge, Mass.)* 17, no. 3 (May 2006): 268–275.

Brookhart, M. A., T. Sturmer, R. J. Glynn, J. Rassen, and S. Schneeweiss. "Confounding Control in Healthcare Database Research: Challenges and Potential Approaches." *Medical Care* 48, no. 6 Suppl (2010): S114–20.

Brown, J. S., M. Kulldorff, K. A. Chan, R. L. Davis, D. Graham, P. T. Pettus, S. E. Andrade, et al. "Early Detection of Adverse Drug Events Within Population-based Health Networks: Application of Sequential Testing Methods." *Pharmacoepidemiology and Drug Safety* 16, no. 12 (2007): 1275–1284.

Brown, J. S., M. Kulldorff, K. R. Petronis, R. Reynolds, K. A. Chan, R. L. Davis, D. Graham, et al. "Early Adverse Drug Event Signal Detection Within Population-based Health Networks Using Sequential Methods: Key Methodologic Considerations." *Pharmacoepidemiology and Drug Safety* 18, no. 3 (2009): 226–234.

Brunt, Christopher S. "CPT Fee Differentials and Visit Upcoding Under Medicare Part B." *Health Economics* 20, no. 7 (July 2011): 831–841.

Budnitz, D. S., D. A. Pollock, K. N. Weidenbach, A. B. Mendelsohn, T. J. Schroeder, and J. L. Annest. "National Surveillance of Emergency Department Visits for Outpatient Adverse Drug Events." *Journal of the American Medical Association* 296, no. 15 (2006): 1858–1866.

Bulte, Christophe Van den, and Yogesh V. Joshi. "New Product Diffusion with Influentials and Imitators." *Marketing Science* 26, no. 3 (May 1, 2007): 400–421.

Bulte, Christophe Van den, and Gary L. Lilien. "Medical Innovation Revisited: Social Contagion Versus Marketing Effort." *American Journal of Sociology* 106, no. 5 (March 1, 2001): 1409–1435.

Burke, Mary A., Gary M. Fournier, and Kislaya Prasad. "The Diffusion of a Medical Innovation: Is Success in the Stars?" *Southern Economic Journal* 73, no. 3 (January 1, 2007): 588–603.

Carnahan, R. M. "Mini-Sentinel's Systematic Reviews of Validated Methods for Identifying Health Outcomes Using Administrative Data: Summary of Findings and Suggestions for Future Research." *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 90–99.

Carpenter, D. "A Proposal for Financing Postmarketing Drug Safety Studies by Augmenting FDA User Fees." *Health Affairs (Project Hope)* Suppl Web Exclusives (2005): W5–469–80.

Center for Drug Evaluation and Research. "Adverse Events Reporting System (AERS) - Potential Signals of Serious Risks/New Safety Information Identified from the Adverse Event Reporting System (AERS)." WebContent, n.d. http://www.fda.gov/drugs/guidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082196.htm.

———. "Approved Risk Evaluation and Mitigation Strategies (REMS): Avandia (rosiglitazone)." WebContent, n.d. http://www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/UCM255624.pdf.

———. "Drug Safety and Availability - FDA Drug Safety Communication: Safety Review of a Reported Death After the First Dose of Multiple Sclerosis Drug Gilenya (fingolimod)." WebContent, December 20, 2011. http://www.fda.gov/Drugs/DrugSafety/ucm284240.htm.

———. "Drug Safety and Availability - FDA Drug Safety Communication: Updated Information About the Risk of Blood Clots in Women Taking Birth Control Pills Containing Drospirenone." WebContent, April 10, 2012. http://www.fda.gov/Drugs/DrugSafety/ucm299305.htm.

———. "Drug Safety Labeling Changes." WebContent, n.d. http://www.fda.gov/Safety/MedWatch/SafetyInformation/Safety-RelatedDrugLabelingChanges/default.htm.

Center for Drug Evaluation and Research, and Food and Drug Administration. "National Drug Code Directory", n.d. http://www.accessdata.fda.gov/scripts/cder/ndc/default.cfm.

Center for Drug Evaluation and Research, Food and Drug Administration, and U.S. Department of Health and Human Services. "Advances in FDA's Safety Program for Marketed Drugs". FDA, April 2012. http://www.fda.gov/downloads/Drugs/DrugSafety/UCM300946.pdf.

Center for Drug Evaluation and Research. "Medication Guides." WebContent, n.d. http://www.fda.gov/Drugs/DrugSafety/ucm085729.htm.

———. "News & Events - Slides for the June 7, 2012 Risk Evaluation and Mitigation Strategy (REMS) Assessments Public Workshop." WebContent, n.d. http://www.fda.gov/Drugs/NewsEvents/ucm307675.htm.

———. "Postmarket Drug Safety Information for Patients and Providers - Approved Risk Evaluation and Mitigation Strategies (REMS)." WebContent, n.d. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm111350.htm.

———. "Postmarket Drug Safety Information for Patients and Providers - HHS FDA: Briefing on Avandia." WebContent, September 23, 2010. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm227934.htm.

———. "Postmarket Requirements and Commitments Database", n.d. http://www.accessdata.fda.gov/scripts/cder/pmc/index.cfm.

———. "Postmarketing Requirements and Commitments: Reports." WebContent, n.d. http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/PostmarketingPhaseIVCommitments/ucm064436.htm.

———. "Pulmonary-Allergy Drugs Advisory Committee - Briefing Information for the March 10-11, 2010 Joint Meeting of the Pulmonary-Allergy Drugs Advisory Committee and Drug Safety and Risk Management Committee." WebContent, March 10, 2010. http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/PulmonaryAllergyDrugsAdvisoryCommittee/ucm202692.htm.

———. "Reproductive Health Drugs Advisory Committee - Briefing Information for the December 9, 2011 Joint Meeting of the Advisory Committee for Reproductive Health Drugs and the Drug Safety and Risk Management Advisory Committee." WebContent, December 9, 2011. http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ReproductiveHealthDrugsAdvisoryCommittee/ucm282631.htm.

Centers for Disease Control and Prevention. "National Vital Statistics System", n.d. http://www.cdc.gov/nchs/nvss.htm.

Centers for Medicare and Medicaid Services. "Healthcare Common Procedural Coding System (HCPCS) Public Meetings", August 3, 2012. http://www.cms.gov/Medicare/Coding/MedHCPCSGenInfo/HCPCSPublicMeetings.html.

———. "Medicaid Drug Programs Data & Resources", n.d. http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Benefits/Prescription-Drugs/Medicaid-Drug-Programs-Data-and-Resources.html.

Chubak, Jessica, Gaia Pocobelli, and Noel S Weiss. "Tradeoffs Between Accuracy Measures for Electronic Health Care Data Algorithms." *Journal of Clinical Epidemiology* 65, no. 3 (March 2012): 343–349.e2.

Cole, Stephen R, and Miguel A Hernán. "Constructing Inverse Probability Weights for Marginal Structural Models." *American Journal of Epidemiology* 168, no. 6 (September 15, 2008): 656–664.

Coleman, James Samuel, and University Columbia. *Medical Innovation; a Diffusion Study*. Indianapolis: Bobbs-Merrill Co, 1966.

Coloma, P. M., M. J. Schuemie, G. Trifiro, R. Gini, R. Herings, J. Hippisley-Cox, G. Mazzaglia, et al. "Combining Electronic Healthcare Databases in Europe to Allow

for Large-scale Drug Safety Monitoring: The EU-ADR Project." *Pharmacoepidemiology and Drug Safety* 20, no. 1 (2011): 1–11.

Coloma, Preciosa M, Gianluca Trifirò, Martijn J Schuemie, Rosa Gini, Ron Herings, Julia Hippisley-Cox, Giampiero Mazzaglia, et al. "Electronic Healthcare Databases for Active Drug Safety Surveillance: Is There Enough Leverage?" *Pharmacoepidemiology and Drug Safety* 21, no. 6 (June 2012): 611–621.

Cook, A. J., R. C. Tiwari, R. D. Wellman, S. R. Heckbert, L. Li, P. Heagerty, T. Marsh, and J. C. Nelson. "Statistical Approaches to Group Sequential Monitoring of Postmarket Safety Surveillance Data: Current State of the Art for Use in the Mini-Sentinel Pilot." *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 72–81.

Cooper, W. O., L. A. Habel, C. M. Sox, K. A. Chan, P. G. Arbogast, T. C. Cheetham, K. T. Murray, et al. "ADHD Drugs and Serious Cardiovascular Events in Children and Young Adults." *The New England Journal of Medicine* 365, no. 20 (2011): 1896–1904.

Council for International Organizations of Medical Sciences. Working Group VIII. *Practical aspects of signal detection in pharmacovigilance : report of CIOMS Working Group VIII.* Geneva: CIOMS, 2010.

Curtis, L. H., M. G. Weiner, D. M. Boudreau, W. O. Cooper, G. W. Daniel, V. P. Nair, M. A. Raebel, et al. "Design Considerations, Architecture, and Use of the Mini-Sentinel Distributed Data System." *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 23–31.

Cutrona, Sarah L, Sengwee Toh, Aarthi Iyer, Sarah Foy, Gregory W Daniel, Vinit P Nair, Daniel Ng, et al. "Validation of Acute Myocardial Infarction in the Food and Drug Administration's Mini-Sentinel Program." *Pharmacoepidemiology and Drug Safety* (June 29, 2012). http://www.ncbi.nlm.nih.gov/pubmed/22745038.

Danaei, Goodarz, Mohammad Tavakkoli, and Miguel A Hernán. "Bias in Observational Studies of Prevalent Users: Lessons for Comparative Effectiveness Research from a Meta-analysis of Statins." *American Journal of Epidemiology* 175, no. 4 (February 15, 2012): 250–262.

Davidoff, Amy J., Bruce Stuart, Thomas Shaffer, J. Samantha Shoemaker, Melissa Kim, and Christopher Zacker. "Lessons Learned: Who Didn't Enroll In Medicare Drug Coverage In 2006, And Why?" *Health Affairs* 29, no. 6 (June 2010): 1255–63.

Davis, R. L., M. Kolczak, E. Lewis, J. Nordin, M. Goodman, D. K. Shay, R. Platt, S. Black, H. Shinefield, and R. T. Chen. "Active Surveillance of Vaccine Safety: a System to Detect Early Signs of Adverse Events." *Epidemiology (Cambridge, Mass.)* 16, no. 3 (2005): 336–341.

Davis, R., and Kulldorff, M. "Statistical Methods Development Details | Vaccine Safety Monitoring - Adverse Events", November 16, 2010. http://mini-sentinel.org/methods/methods_development/details.aspx?ID=1028.

Delaney, Joseph A C, Mary L Biggs, Richard A Kronmal, and Bruce M Psaty. "Demographic, Medical, and Behavioral Characteristics Associated with over the Counter Non-steroidal Anti-inflammatory Drug Use in a Population-based Cohort: Results from the Multi-Ethnic Study of Atherosclerosis." *Pharmacoepidemiology and Drug Safety* 20, no. 1 (January 2011): 83–89.

Demets, D L. "Group Sequential Procedures: Calendar Versus Information Time." *Statistics in Medicine* 8, no. 10 (October 1989): 1191–1198.

DeMets, D L, and K K Lan. "Interim Analysis: The Alpha Spending Function Approach." *Statistics in Medicine* 13, no. 13–14 (July 15, 1994): 1341–1352; discussion 1353–1356.

Department of Health and Human Services. "Code of Federal Regulations Title 45 Part 46, Protection of Human Subjects", July 14, 2009. http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html.

Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, and Center for Biologics Evaluation and Research. "Guidance for Industry: Format and Content of Proposed Risk Evaluation and Mitigation Strategies (REMS), REMS Assessments, and REMS Modifications (Draft)", October 1, 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM184128.pdf.

———. "Guidance for Industry: Postmarketing Studies and Clinical Trials - Implementation of Section 505(o)(3) of the Federal Food, Drug, and Cosmetic Act (Final)", March 31, 2011. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM172001.pdf.

———. "Guidance for Industry: Safety Labeling Changes - Implementation of Section 505(o)(4) of the Federal Food, Drug, and Cosmetic Act (Draft)", April 12, 2011. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM250783.pdf.

Department of Health and Human Services Office of the Inspector General. *FDA's Monitoring of Postmarket Study Commitments*. Vol. OEI-01–04–00390. Washington, DC: OIG, 2006.

Dore, D. D., G. L. Bloomgren, M. Wenten, C. Hoffman, C. R. Clifford, S. G. Quinn, D. K. Braun, R. A. Noel, and J. D. Seeger. "A Cohort Study of Acute Pancreatitis in Relation to Exenatide Use." *Diabetes, Obesity & Metabolism* 13, no. 6 (2011): 559–566.

Dore, D. D., J. D. Seeger, and K. Arnold Chan. "Use of a Claims-based Active Drug Safety Surveillance System to Assess the Risk of Acute Pancreatitis with Exenatide or Sitagliptin Compared to Metformin or Glyburide." *Current Medical Research and Opinion* 25, no. 4 (2009): 1019–1027.

"Drugs@FDA: FDA Approved Drug Products", n.d. http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm.

von Elm, E., D. G. Altman, M. Egger, S. J. Pocock, P. C. Gotzsche, J. P. Vandenbroucke, and STROBE Initiative. "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies." *Annals of Internal Medicine* 147, no. 8 (2007): 573–577.

von Elm, E., and M. Egger. "The Scandal of Poor Epidemiological Research." *BMJ (Clinical Research Ed.)* 329, no. 7471 (2004): 868–869.

Emerson, Scott S, John M Kittelson, and Daniel L Gillen. "Frequentist Evaluation of Group Sequential Clinical Trial Designs." *Statistics in Medicine* 26, no. 28 (December 10, 2007): 5047–5080.

Evans, B. J. "Seven Pillars of a New Evidentiary Paradigm: The Food, Drug, and Cosmetic Act Enters the Genomic Era." *Notre Dame Law Review* 85, no. 2 (2010): 419–524.

Fireman, B., S. Toh, M. G. Butler, A. S. Go, H. V. Joffe, D. J. Graham, J. C. Nelson, G. W. Daniel, and J. V. Selby. "A Protocol for Active Surveillance of Acute Myocardial Infarction in Association with the Use of a New Antidiabetic Pharmaceutical Agent." *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 282–290.

Food and Drug Administration Amendments Act of 2007, Public Law 110-85, 2007.

Food and Drug Administration, and Center for Drug Evaluation and Research. "FDA Drug Safety Communication: FDA Requires Post-market Safety Trials for Long-Acting Beta-Agonists (LABAs)." WebContent, April 15, 2011. http://www.fda.gov/Drugs/DrugSafety/ucm251512.htm.

———. "Prescription Drug User Fee Act (PDUFA) V: Reauthorization Performance Goals and Procedures; Fiscal Years 2013 Through 2017." WebContent, July 19, 2012. http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM270412.pdf.

———. "Transcript of the Joint Meeting of Pulmonary-Allergy Drugs Advisory Committee and the Drug Safety and Risk Management Advisory Committee", March 11, 2010. http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/Pulmonary-AllergyDrugsAdvisoryCommittee/UCM209124.pdf.

Food and Drug Administration Safety and Innovation Act, Public Law 112-144, 2012.

Gagne, J. J., B. Fireman, P. B. Ryan, M. Maclure, T. Gerhard, S. Toh, J. A. Rassen, J. C. Nelson, and S. Schneeweiss. "Design Considerations in an Active Medical Product Safety Monitoring System." *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 32–40.

Gamble, John-Michael, Finlay A McAlister, Jeffrey A Johnson, and Dean T Eurich. "Restrictive Drug Coverage Policies Can Induce Substantial Drug Exposure Misclassification in Pharmacoepidemiologic Studies." *Clinical Therapeutics* 34, no. 6 (June 2012): 1379–1386.e3.

Garcia Rodriguez, L. A., R. Herings, and S. Johansson. "Use of Multiple International Healthcare Databases for the Detection of Rare Drug-associated Outcomes: a Pharmacoepidemiological Programme Comparing Rosuvastatin with Other Marketed Statins." *Pharmacoepidemiology and Drug Safety* 19, no. 12 (2010): 1218–1224.

Gee, J., A. Naleway, I. Shui, J. Baggs, R. Yin, R. Li, M. Kulldorff, et al. "Monitoring the Safety of Quadrivalent Human Papillomavirus Vaccine: Findings from the Vaccine Safety Datalink." *Vaccine* 29, no. 46 (2011): 8279–8284.

Glynn, R. J., J. J. Gagne, and S. Schneeweiss. "Role of Disease Risk Scores in Comparative Effectiveness Research with Emerging Therapies." *Pharmacoepidemiology and Drug Safety* 21 Suppl 2 (2012): 138–147.

Goldenberg, Jacob, Barak Libai, and Eitan Muller. "Riding the Saddle: How Cross-Market Communications Can Create a Major Slump in Sales." *Journal of Marketing* 66, no. 2 (April 1, 2002): 1–16.

Golder, Su, Yoon K Loke, and Martin Bland. "Meta-analyses of Adverse Effects Data Derived from Randomised Controlled Trials as Compared to Observational Studies: Methodological Overview." *PLoS Medicine* 8, no. 5 (May 2011): e1001026.

Green, M. S. "Use of Predictive Value to Adjust Relative Risk Estimates Biased by Misclassification of Outcome Status." *American Journal of Epidemiology* 117, no. 1 (1983): 98–105.

Greene, S. K., M. Kulldorff, E. M. Lewis, R. Li, R. Yin, E. S. Weintraub, B. H. Fireman, et al. "Near Real-Time Surveillance for Influenza Vaccine Safety: Proof-of-Concept in the Vaccine Safety Datalink Project." *American Journal of Epidemiology* (2009).

Greene, S. K., M. Kulldorff, R. Yin, W. K. Yih, T. A. Lieu, E. S. Weintraub, and G. M. Lee. "Near Real-time Vaccine Safety Surveillance with Partially Accrued Data." *Pharmacoepidemiology and Drug Safety* 20, no. 6 (2011): 583–590.

Greenland, S. "An Introduction to Instrumental Variables for Epidemiologists." *International Journal of Epidemiology* 29, no. 4 (August 2000): 722–729.

———. "Basic Methods for Sensitivity Analysis of Biases." *International Journal of Epidemiology* 25, no. 6 (December 1996): 1107–1116.

Greenland, Sander. "Multiple-bias Modelling for Analysis of Observational Data." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168, no. 2 (2005): 267–306.

Greenland, Sander, and Kenneth J Rothman. "Introduction to Stratified Analysis." In *Modern Epidemiology*, edited by Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash, 258–302. Third. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.

Habel, L. A., W. O. Cooper, C. M. Sox, K. A. Chan, B. H. Fireman, P. G. Arbogast, T. C. Cheetham, et al. "ADHD Medications and Risk of Serious Cardiovascular Events in Young and Middle-aged Adults." *JAMA : the Journal of the American Medical Association* 306, no. 24 (2011): 2673–2683.

Hahn, Minhi, Sehoon Park, Lakshman Krishnamurthi, and Andris A. Zoltners. "Analysis of New Product Diffusion Using a Four-Segment Trial-Repeat Model." *Marketing Science* 13, no. 3 (July 1, 1994): 224–247.

Hall, G. C., B. Sauer, A. Bourke, J. S. Brown, M. W. Reynolds, and R. L. Casale. "Guidelines for Good Database Selection and Use in Pharmacoepidemiology Research." *Pharmacoepidemiology and Drug Safety* 21, no. 1 (2012): 1–10.

Halsey, Neal A, Kathryn M Edwards, Cornelia L Dekker, Nicola P Klein, Roger Baxter, Philip Larussa, Colin Marchant, Barbara Slade, and Claudia Vellozzi. "Algorithm to Assess Causality After Individual Adverse Events Following Immunizations." *Vaccine* 30, no. 39 (August 24, 2012): 5791–5798.

Hammad, Tarek A, Simone P Pinheiro, and George A Neyarapally. "Secondary Use of Randomized Controlled Trials to Evaluate Drug Safety: a Review of Methodological Considerations." *Clinical Trials (London, England)* 8, no. 5 (October 2011): 559–570.

Hauben, Manfred, and Jeffrey K Aronson. "Defining 'Signal' and Its Subtypes in Pharmacovigilance Based on a Systematic Review of Previous Definitions." *Drug Safety* 32, no. 2 (2009): 99–110.

Hearne, Jean, and Congressional Research Service. *Prescription Drug Coverage Under Medicaid: CRS Report RL30726*. Washington, D.C.: Congressional Research Service, Library of Congress, February 6, 2008. http://opencrs.com/document/RL30726.

Hennekens, C. H., and D. DeMets. "Statistical Association and Causation: Contributions of Different Types of Evidence." *JAMA : the Journal of the American Medical Association* 305, no. 11 (2011): 1134–1135.

Hernán, Miguel A, Alvaro Alonso, Roger Logan, Francine Grodstein, Karin B Michels, Walter C Willett, Joann E Manson, and James M Robins. "Observational Studies Analyzed Like Randomized Experiments: An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease." *Epidemiology (Cambridge, Mass.)* 19, no. 6 (November 2008): 766–779.

Hernán, Miguel A, and Sonia Hernández-Díaz. "Beyond the Intention-to-treat in Comparative Effectiveness Research." *Clinical Trials (London, England)* 9, no. 1 (February 2012): 48–55.

Hernán, Miguel A, and James M Robins. "Estimating Causal Effects from Epidemiological Data." *Journal of Epidemiology and Community Health* 60, no. 7 (July 2006): 578–586.

Hernán, Miguel, and Jamie Robins. *Causal Inference*. v1.10.17 ed. Chapman & Hall/CRC, 2012. http://www.hsph.harvard.edu/faculty/miguel-hernan/causal-inference-book/.

Hsiao, Chun-Ju, Esther Hing, Thomas C Socey, and Bill Cai. "Electronic Health Record Systems and Intent to Apply for Meaningful Use Incentives Among Office-based Physician Practices: United States, 2001-2011." *NCHS Data Brief*, no. 79 (November 2011): 1–8.

Institute of Medicine (IOM). *Adverse Effects of Vaccines: Evidence and Causality*. Washington, D.C.: The National Academies Press, 2012.

———. *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*. Washington, D.C.: National Academies Press, 2012. http://books.nap.edu/catalog/13219.html.

———. *Improving the Presumptive Disability Decision-Making Process for Veterans*. Edited by Catherine C. Bodurow and Jonathan M. Samet. The National Academies Press, 2008. http://www.nap.edu/openbook.php?record_id=11908.

———. *The Future of Drug Safety: Promoting and Protecting the Health of the Public*. Washington, DC: National Academies Press, 2007. http://www.nap.edu/catalog/11750.html.

Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente. "Opinion Leadership and Social Contagion in New Product Diffusion." *Marketing Science* 30, no. 2 (March 1, 2011): 195–212.

Jane-wit, D., R. I. Horwitz, and J. Concato. "Variation in Results from Randomized, Controlled Trials: Stochastic or Systematic?" *Journal of Clinical Epidemiology* 63, no. 1 (2010): 56–63.

Jennison, Christopher, and Bruce W. Turnbull. *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton: Chapman & Hall/CRC, 2000.

Jha, A. K., C. M. DesRoches, E. G. Campbell, K. Donelan, S. R. Rao, T. G. Ferris, A. Shields, S. Rosenbaum, and D. Blumenthal. "Use of Electronic Health Records in

U.S. Hospitals." *The New England Journal of Medicine* 360, no. 16 (2009): 1628–1638.

Karr, Alan F., Xiaodong Lin, Ashish P. Sanil, and Jerome P. Reiter. "Privacy-Preserving Analysis of Vertically Partitioned Data Using Secure Matrix Products." *Journal of Official Statistics* 25, no. 1 (2009): 125–138.

Keeney, Ralph L. "Utility Functions for Multiattributed Consequences." *Management Science* 18, no. 5, Theory Series, Part 1 (1972): 276–287.

Kessler, David A., and David C. Vladeck. "A Critical Examination of the FDA's Efforts To Preempt Failure-To-Warn Claims." *Georgetown Law Journal* 96 (2008): 461–495.

Kim, Seo Young, and Daniel H Solomon. "Use of Administrative Claims Data for Comparative Effectiveness Research of Rheumatoid Arthritis Treatments." *Arthritis Research & Therapy* 13, no. 5 (2011): 129.

Kittelson, J M, and S S Emerson. "A Unifying Family of Group Sequential Test Designs." *Biometrics* 55, no. 3 (September 1999): 874–882.

Klein, N. P., B. Fireman, W. K. Yih, E. Lewis, M. Kulldorff, P. Ray, R. Baxter, et al. "Measles-mumps-rubella-varicella Combination Vaccine and the Risk of Febrile Seizures." *Pediatrics* 126, no. 1 (2010): e1–8.

Kulldorff, M., R. L. Davis, M. Kolczak, E. Lewis, T. A. Lieu, and R. Platt. "A Maximized Sequential Probability Ratio Test for Drug and Vaccine Safety Surveillance." *Seq Anal* 30, no. 1 (2011): 58–78.

Kulldorff, Martin. "Sequential Statistical Methods for Prospective Postmarketing Safety Surveillance." In *Pharmacoepidemiology*, edited by Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy, 852–867. Fifth. John Wiley & Sons, 2011.

Kulldorff, Martin, and Ivair Silva. "Continuous Sequential Analysis with Delayed Start", 2012.

Lan, K. K. Gordon, and David L. Demets. "Discrete Sequential Boundaries for Clinical Trials." *Biometrika* 70, no. 3 (December 1, 1983): 659–663.

Lash, Timothy L, and Aliza K Fink. "Semi-automated Sensitivity Analysis to Assess Systematic Errors in Observational Data." *Epidemiology (Cambridge, Mass.)* 14, no. 4 (July 2003): 451–458.

Lash, Timothy L, Morten Schmidt, Annette Østergaard Jensen, and Malene Cramer Engebjerg. "Methods to Apply Probabilistic Bias Analysis to Summary Estimates of Association." *Pharmacoepidemiology and Drug Safety* 19, no. 6 (June 2010): 638–644.

Lawlor, D. A., G. Davey Smith, D. Kundu, K. R. Bruckdorfer, and S. Ebrahim. "Those Confounded Vitamins: What Can We Learn from the Differences Between Observational Versus Randomised Trial Evidence?" *Lancet* 363, no. 9422 (2004): 1724–1727.

Lee, Grace M., Sharon K. Greene, Eric S. Weintraub, James Baggs, Martin Kulldorff, Bruce H. Fireman, Roger Baxter, et al. "H1N1 and Seasonal Influenza Vaccine Safety in the Vaccine Safety Datalink Project." *American Journal of Preventive Medicine* 41, no. 2 (August 2011): 121–128.

Lester, Jean, George A Neyarapally, Earlene Lipowski, Cheryl Graham, Marni Hall, and Gerald Dal Pan. "Evaluation of FDA Safety-Related Drug Label Changes in 2010", May 5, 2012.

Li, L. "A Conditional Sequential Sampling Procedure for Drug Safety Surveillance." *Statistics in Medicine* 28, no. 25 (2009): 3124–3138.

Li, L., and M. Kulldorff. "A Conditional Maximized Sequential Probability Ratio Test for Pharmacovigilance." *Statistics in Medicine* 29, no. 2 (2010): 284–295.

Li, Lingling, Martin Kulldorff, Jennifer Nelson, and Andrea Cook. "A Propensity Score-Enhanced Sequential Analytic Method for Comparative Drug Safety Surveillance." *Statistics in Biosciences* 3, no. 1 (2011): 45–62.

Lieu, T. A., M. Kulldorff, R. L. Davis, E. M. Lewis, E. Weintraub, K. Yih, R. Yin, J. S. Brown, R. Platt, and for the Vaccine Safety Datalink Rapid Cycle Analysis Team. "Real-time Vaccine Safety Surveillance for the Early Detection of Adverse Events." *Medical Care* 45, no. 10 Supl 2 (2007): S89–95.

Mahajan, Vijay, and Robert A. Peterson. "Innovation Diffusion in a Dynamic Potential Adopter Population." *Management Science* 24, no. 15 (November 1, 1978): 1589–1597.

———. *Models for Innovation Diffusion*. Sage University Papers Series. Quantitative Applications in the Social Sciences no. 07-048. Beverly Hills: Sage Publications, 1985.

Maro, J. C., and J. S. Brown. "Impact of Exposure Accrual on Sequential Postmarket Evaluations: a Simulation Study." *Pharmacoepidemiology and Drug Safety* 20, no. 11 (2011): 1184–1191.

Maro, J. C., R. Platt, J. H. Holmes, B. L. Strom, S. Hennessy, R. Lazarus, and J. S. Brown. "Design of a National Distributed Health Data Network." *Annals of Internal Medicine* 151, no. 5 (2009): 341–344.

Maro, Judith C. "Development of a public health information infrastructure for postmarket evidence". Thesis, Massachusetts Institute of Technology, 2009. http://dspace.mit.edu/handle/1721.1/53058.

McCandless, Lawrence C, Paul Gustafson, and Adrian R Levy. "A Sensitivity Analysis Using Information About Measured Confounders Yielded Improved Uncertainty Assessments for Unmeasured Confounding." *Journal of Clinical Epidemiology* 61, no. 3 (March 2008): 247–255.

McClure, D. L., J. M. Glanz, S. Xu, S. J. Hambidge, J. P. Mullooly, and J. Baggs. "Comparison of Epidemiologic Methods for Active Surveillance of Vaccine Safety." *Vaccine* 26, no. 26 (2008): 3341–3345.

McGraw, D., K. Rosati, and B. Evans. "A Policy Framework for Public Health Uses of Electronic Health Data." *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 18–22.

McMahon, Alex D. "Approaches to Combat with Confounding by Indication in Observational Studies of Intended Drug Effects." *Pharmacoepidemiology and Drug Safety* 12, no. 7 (November 2003): 551–558.

Meade, Nigel, and Towhidul Islam. "Modelling and Forecasting the Diffusion of Innovation – A 25-year Review." *International Journal of Forecasting* 22, no. 3 (2006): 519–545.

Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Public Law 108-173, 2003.

Meyboom, R H, M Lindquist, and A C Egberts. "An ABC of Drug-related Problems." *Drug Safety: An International Journal of Medical Toxicology and Drug Experience* 22, no. 6 (June 2000): 415–423.

Mini-Sentinel Operations Center. "Angiotensin II Receptor Blockers & Celiac Disease", January 17, 2012. http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Angiotensin-II-Receptor-Blockers-and-Celiac-Disease.pdf.

———. "Assessments of Diagnoses and Medical Procedures", n.d. http://mini-sentinel.org/assessments/diagnoses_and_medical_procedures/default.aspx.

———. "Assessments of Diagnoses and Medical Procedures | Acute Myocardial Infarctions", April 27, 2012. http://mini-sentinel.org/assessments/diagnoses_and_medical_procedures/details.aspx?ID=132.

———. "Distributed Query Tool - Summary Tables", August 10, 2011. http://mini-sentinel.org/data_activities/details.aspx?ID=117.

———. "Module 1: Drug Use - General Characterization", August 17, 2011. http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Modular-Program-1_v1.0.pdf.

———. "Module 2: Drug Use - By Medical Condition", August 17, 2011. http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Modular-Program-2_v1.0.pdf.

———. "Module 3: Drug Use - Incident Outcomes", August 17, 2011. http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Modular-Program-3_v1.0.pdf.

———. "Module 4: Drug Use - Concomitant Use", August 17, 2011. http://www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Modular-Program-4_v1.0.pdf.

———. "Smoking Cessation Drugs & Cardiovascular Outcomes", January 17, 2012. http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Smoking-Cessation-Drugs-and-Selected-Cardiovascular-Outcomes.pdf.

Modi, Avani C, Joseph R Rausch, and Tracy A Glauser. "Patterns of Nonadherence to Antiepileptic Drug Therapy in Children with Newly Diagnosed Epilepsy." *JAMA: The Journal of the American Medical Association* 305, no. 16 (April 27, 2011): 1669–1676.

Moore, T. J., B. M. Psaty, and C. D. Furberg. "Time to Act on Drug Safety." *Journal of the American Medical Association* 279, no. 19 (1998): 1571–1573.

Moore, Thomas J, Sonal Singh, and Curt D Furberg. "The FDA and New Safety Warnings." *Archives of Internal Medicine* 172, no. 1 (January 9, 2012): 78–80.

Mullooly, J. P. "Misclassification Model for Person-time Analysis of Automated Medical Care Databases." *American Journal of Epidemiology* 144, no. 8 (1996): 782–792.

Nathan, D. M. "Rosiglitazone and Cardiotoxicity--weighing the Evidence." *The New England Journal of Medicine* 357, no. 1 (2007): 64–66.

Nelson, J. C., A. J. Cook, O. Yu, C. Dominguez, S. Zhao, S. K. Greene, B. H. Fireman, S. J. Jacobsen, E. S. Weintraub, and L. A. Jackson. "Challenges in the Design and Analysis of Sequentially Monitored Postmarket Safety Surveillance Evaluations

Using Electronic Observational Health Care Data." *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 62–71.

Nelson, Jennifer C, Andrea J Cook, Onchee Yu, Shanshan Zhao, Lisa A Jackson, and Bruce M Psaty. "Methods for Observational Post-licensure Medical Product Safety Surveillance." *Statistical Methods in Medical Research* (December 2, 2011). http://www.ncbi.nlm.nih.gov/pubmed/22138688.

Nguyen, M., R. Ball, K. Midthun, and T. A. Lieu. "The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring Program: Strengthening the Federal Vaccine Safety Enterprise." *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 291–297.

Nguyen, Michael D., Sharon K. Greene, W. Katherine Yih, Tracy A. Lieu, Nandini Selvam, Vinit Nair, Cheryl N. McMahill-Walraven, and David B. Martin. "Monitoring for Venous Thromboembolism After Gardasil Vaccination." *Mini-Sentinel*, March 6, 2012. http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Gardasil-and-Venous-Thromboembolism-Protocol.pdf.

Nissen, S. E., and K. Wolski. "Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes." *The New England Journal of Medicine* 356, no. 24 (2007): 2457–2471.

"Notice to Readers: Update on Supply of Vaccines Containing Varicella-Zoster Virus." *JAMA: The Journal of the American Medical Association* 298, no. 7 (August 15, 2007): 736.

O'Leary, Sean T, Jason M Glanz, David L McClure, Aysha Akhtar, Matthew F Daley, Cynthia Nakasato, Roger Baxter, et al. "The Risk of Immune Thrombocytopenic Purpura After Vaccination in Children and Adolescents." *Pediatrics* 129, no. 2 (February 2012): 248–255.

Owens, D. K., Kathleen N. Lohr, D Atkins, J. R. Treadwell, J. T. Reston, E. B. Bass, S. Chang, and M. Helfand. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. AHRQ Methods for Effective Health Care. Rockville (MD): Agency for Healthcare Research and Quality (US), 2008. http://www.ncbi.nlm.nih.gov/pubmed/21433399.

Paich, Mark, Corey Peck, and Jason Valant. "Pharmaceutical Market Dynamics and Strategic Planning: a System Dynamics Perspective." *System Dynamics Review (Wiley)* 27, no. 1 (March 2011): 47–63.

Dal Pan, Gerald J., Marie Lindquist, and Kate Gelperin. "Postmarketing Spontaneous Pharmacovigilance Reporting Systems." In *Pharmacoepidemiology*, edited by Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy, 137–157. Fifth. John Wiley & Sons, 2011.

Papanikolaou, P. N., G. D. Christidi, and J. P. Ioannidis. "Comparison of Evidence on Harms of Medical Interventions in Randomized and Nonrandomized Studies." *CMAJ : Canadian Medical Association Journal* 174, no. 5 (2006): 635–641.

Patient Protection and Affordable Care Act, P.L. 111-148, 2010.

Peres, Renana, Eitan Muller, and Vijay Mahajan. "Innovation Diffusion and New Product Growth Models: A Critical Review and Research Directions." *International Journal of Research in Marketing* 27, no. 2 (June 2010): 91–106.

Pfeiffer, Paul N, Dara Ganoczy, Nicholas W Bowersox, John F McCarthy, Frederic C Blow, and Marcia Valenstein. "Depression Care Following Psychiatric Hospitalization in the Veterans Health Administration." *The American Journal of Managed Care* 17, no. 9 (September 2011): e358–364.

Pfeiffer, Paul N, Benjamin R Szymanski, Marcia Valenstein, John F McCarthy, and Kara Zivin. "Trends in Antidepressant Prescribing for New Episodes of Depression and Implications for Health System Quality Measures." *Medical Care* 50, no. 1 (January 2012): 86–90.

Platt, R., R. M. Carnahan, J. S. Brown, E. Chrischilles, L. H. Curtis, S. Hennessy, J. C. Nelson, et al. "The U.S. Food and Drug Administration's Mini-Sentinel Program: Status and Direction." *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 1–8.

Platt, R., R. Davis, J. Finkelstein, A. S. Go, J. H. Gurwitz, D. Roblin, S. Soumerai, et al. "Multicenter Epidemiologic and Health Services Research on Therapeutics in the HMO Research Network Center for Education and Research on Therapeutics." *Pharmacoepidemiology and Drug Safety* 10, no. 5 (2001): 373–377.

Platt, R., M. Wilson, K. A. Chan, J. S. Benner, J. Marchibroda, and M. McClellan. "The New Sentinel Network--improving the Evidence of Medical-product Safety." *The New England Journal of Medicine* 361, no. 7 (2009): 645–647.

Pocock, S. J., and D. R. Elbourne. "Randomized Trials or Observational Tribulations?" *The New England Journal of Medicine* 342, no. 25 (2000): 1907–1909.

Powers, Abbey, and G Elliott Cook. "Potential Safety Signals and Their Significance." *Archives of Internal Medicine* 172, no. 1 (January 9, 2012): 72–73.

Psaty, B. M., and C. D. Furberg. "COX-2 Inhibitors--lessons in Drug Safety." *The New England Journal of Medicine* 352, no. 11 (2005): 1133–1135.

———. "The Record on Rosiglitazone and the Risk of Myocardial Infarction." *The New England Journal of Medicine* 357, no. 1 (2007): 67–69.

Psaty, Bruce M, and David S Siscovick. "Minimizing Bias Due to Confounding by Indication in Comparative Effectiveness Research: The Importance of Restriction." *JAMA: The Journal of the American Medical Association* 304, no. 8 (August 25, 2010): 897–898.

Qureshi, Zaina P, Enrique Seoane-Vazquez, Rosa Rodriguez-Monguio, Kurt B Stevenson, and Sheryl L Szeinbach. "Market Withdrawal of New Molecular Entities Approved in the United States from 1980 to 2009." *Pharmacoepidemiology and Drug Safety* 20, no. 7 (July 2011): 772–777.

Rassen, J. A., J. Avorn, and S. Schneeweiss. "Multivariate-adjusted Pharmacoepidemiologic Analyses of Confidential Information Pooled from Multiple Health Care Utilization Databases." *Pharmacoepidemiology and Drug Safety* 19, no. 8 (2010): 848–857.

Rassen, J. A., and S. Schneeweiss. "Using High-dimensional Propensity Scores to Automate Confounding Control in a Distributed Medical Product Safety Surveillance System." *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 41–49.

Rassen, J. A., D. H. Solomon, J. R. Curtis, L. Herrinton, and S. Schneeweiss. "Privacy-maintaining Propensity Score-based Pooling of Multiple Databases Applied to a Study of Biologics." *Medical Care* 48, no. 6 Suppl (2010): S83–9.

Rassen, Jeremy A, M Alan Brookhart, Robert J Glynn, Murray A Mittleman, and Sebastian Schneeweiss. "Instrumental Variables I: Instrumental Variables Exploit Natural Variation in Nonexperimental Data to Estimate Causal Relationships." *Journal of Clinical Epidemiology* 62, no. 12 (December 2009): 1226–1232.

Ray, W. A. "Improving Automated Database Studies." *Epidemiology (Cambridge, Mass.)* 22, no. 3 (2011): 302–304.

Ray, Wayne A. "Evaluating Medication Effects Outside of Clinical Trials: New-user Designs." *American Journal of Epidemiology* 158, no. 9 (November 1, 2003): 915–920.

Reynolds, R. F., J. A. Lem, N. M. Gatto, and S. M. Eng. "Is the Large Simple Trial Design Used for Comparative, Post-approval Safety Research? A Review of a Clinical Trials Registry and the Published Literature." *Drug Safety : an International Journal of Medical Toxicology and Drug Experience* 34, no. 10 (2011): 799–820.

Riley, Gerald F., Jesse M. Levy, and Melissa A. Montgomery. "Adverse Selection In The Medicare Prescription Drug Program." *Health Affairs* 28, no. 6 (December 2009): 1826–37.

Robb, M. A., J. A. Racoosin, R. E. Sherman, T. P. Gross, R. Ball, M. E. Reichman, K. Midthun, and J. Woodcock. "The US Food and Drug Administration's Sentinel Initiative: Expanding the Horizons of Medical Product Safety." *Pharmacoepidemiology and Drug Safety* 21 Suppl 1 (2012): 9–11.

Robb, Melissa. "FDA's Mini-Sentinel Exceeds 100 Million Lives (and Counting)… A Major Milestone in Developing a Nationwide Rapid-response Electronic Medical Product Safety Surveillance Program." *FDA Voice*, June 29, 2012. http://blogs.fda.gov/fdavoice/index.php/2012/06/fdas-mini-sentinel-exceeds-100-million-lives-and-counting-a-major-milestone-in-developing-a-nationwide-rapid-response-electronic-medical-product-safety-surveillance-program/.

Rogers, Everett M. *Diffusion of Innovations*. 5th ed. New York: Free Press, 2003.

Rossouw, J. E., G. L. Anderson, R. L. Prentice, A. Z. LaCroix, C. Kooperberg, M. L. Stefanick, R. D. Jackson, et al. "Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results From the Women's Health Initiative Randomized Controlled Trial." *Journal of the American Medical Association* 288, no. 3 (2002): 321–333.

Rothman, Kenneth J, Sander Greenland, and Timothy L. Lash. "Design Strategies to Improve Study Accuracy." In *Modern Epidemiology*, edited by Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash, 168–182. Third. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.

———. "Validity in Epidemiologic Studies." In *Modern Epidemiology*, edited by Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash, 128–147. Third. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.

Rothman, Kenneth J., Sander Greenland, and Timothy L. Lash. *Modern Epidemiology*. Third. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.

Rubinstein, Reuven Y., and Dirk P. Kroese. *Simulation and the Monte Carlo Method*. Vol. 2. Hoboken, N.J.: John Wiley & Sons, 2008.

Salmon, D. A., A. Akhtar, M. J. Mergler, K. S. Vannice, H. Izurieta, R. Ball, G. M. Lee, et al. "Immunization-safety Monitoring Systems for the 2009 H1N1 Monovalent Influenza Vaccination Program." *Pediatrics* 127 Suppl 1 (2011): S78–86.

Schneeweiss, S., and J. Avorn. "A Review of Uses of Health Care Utilization Databases for Epidemiologic Research on Therapeutics." *Journal of Clinical Epidemiology* 58, no. 4 (2005): 323–337.

Schneeweiss, S., A. R. Patrick, T. Sturmer, M. A. Brookhart, J. Avorn, M. Maclure, K. J. Rothman, and R. J. Glynn. "Increasing Levels of Restriction in Pharmacoepidemiologic Database Studies of Elderly and Comparison with Randomized Trial Results." *Medical Care* 45, no. 10 Supl 2 (2007): S131–42.

Schneeweiss, S., J. A. Rassen, R. J. Glynn, J. Avorn, H. Mogun, and M. A. Brookhart. "High-dimensional Propensity Score Adjustment in Studies of Treatment Effects Using Health Care Claims Data." *Epidemiology (Cambridge, Mass.)* 20, no. 4 (2009): 512–522.

Schneeweiss, Sebastian, Robert J Glynn, Elizabeth H Tsai, Jerry Avorn, and Daniel H Solomon. "Adjusting for Unmeasured Confounders in Pharmacoepidemiologic Claims Data Using External Information: The Example of COX2 Inhibitors and Myocardial Infarction." *Epidemiology (Cambridge, Mass.)* 16, no. 1 (January 2005): 17–24.

Schneeweiss, Sebastian, and Jennifer Nelson. "Mini-Sentinel Methods Core: Accomplishments and Lessons Learned" presented at the International Society of Pharmacoepidemiology (ISPE), Chicago, IL, August 17, 2011.

Schneeweiss, Sebastian. "Sensitivity Analysis and External Adjustment for Unmeasured Confounders in Epidemiologic Database Studies of Therapeutics." *Pharmacoepidemiology and Drug Safety* 15, no. 5 (May 2006): 291–303.

Seeger, John, and Gregory W. Daniel. "Commercial Insurance Databases." In *Pharmacoepidemiology*, edited by Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy, 189–208. Fifth. John Wiley & Sons, 2011.

Shrank, William H., Amanda R. Patrick, and M. Alan Brookhart. "Healthy User and Related Biases in Observational Studies of Preventive Interventions: A Primer for Physicians." *Journal of General Internal Medicine* 26, no. 5 (May 2011): 546–550.

Spiegelhalter, D J, J P Myles, D R Jones, and K R Abrams. "Bayesian Methods in Health Technology Assessment: a Review." *Health Technology Assessment (Winchester, England)* 4, no. 38 (2000): 1–130.

Spiegelhalter, David J. "Incorporating Bayesian Ideas into Health-care Evaluation." *Statistical Science* 19, no. 1 (2004): 156–174.

Steiner, S H, R J Cook, and V T Farewell. "Risk-adjusted Monitoring of Binary Surgical Outcomes." *Medical Decision Making: An International Journal of the Society for Medical Decision Making* 21, no. 3 (June 2001): 163–169.

Sterman, John. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston: Irwin/McGraw-Hill, 2000.

Strom, B. L. "Methodologic Challenges to Studying Patient Safety and Comparative Effectiveness." *Medical Care* 45, no. 10 Supl 2 (2007): S13–5.

———. "Potential for Conflict of Interest in the Evaluation of Suspected Adverse Drug Reactions: a Counterpoint." *Journal of the American Medical Association* 292, no. 21 (2004): 2643–2646.

Strom, Brian L. "Basic Principles of Clinical Epidemiology Relevant to Pharmacoepidemiologic Studies." In *Pharmacoepidemiology*, edited by Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy, 38–51. Fifth. John Wiley & Sons, 2011.

———. "Overview of Automated Databases in Pharmacoepidemiology." In *Pharmacoepidemiology*, edited by Brian L. Strom, Stephen E. Kimmel, and Sean Hennessy, 158–162. Fifth. John Wiley & Sons, 2011.

Stürmer, Til, Robert J Glynn, Kenneth J Rothman, Jerry Avorn, and Sebastian Schneeweiss. "Adjustments for Unmeasured Confounders in Pharmacoepidemiologic Database Studies Using External Information." *Medical Care* 45, no. 10 Supl 2 (October 2007): S158–165.

Subramanian, Sujha. "Impact of Medicaid Copayments on Patients with Cancer: Lessons for Medicaid Expansion Under Health Reform." *Medical Care* 49, no. 9 (September 2011): 842–847.

Taubman, Sarah L, James M Robins, Murray A Mittleman, and Miguel A Hernán. "Intervening on Risk Factors for Coronary Heart Disease: An Application of the Parametric G-formula." *International Journal of Epidemiology* 38, no. 6 (December 2009): 1599–1611.

Temple, R. "Meta-analysis and Epidemiologic Studies in Drug Development and Postmarketing Surveillance." *JAMA : the Journal of the American Medical Association* 281, no. 9 (1999): 841–844.

Terrell, Deirdra R, Laura A Beebe, Sara K Vesely, Barbara R Neas, Jodi B Segal, and James N George. "Determining a Definite Diagnosis of Primary Immune Thrombocytopenia by Medical Record Review." *American Journal of Hematology* 87, no. 9 (September 2012): 843–847.

The Council for International Organizations of Medical Sciences (CIOMS) Working Group III. *Guidelines for Preparing Core Clinical Safety Information on Drugs.* Geneva: World Health Organization (WHO), 1995.

Toh, Darren, Marsha E. Reichman, Monika Houston, Hernandez, Adrian, Lingling Li, Carolyn McCloskey, Ahoaibi Azadeh, et al. "Protocol for Signal Refinement of Angioedema Events in Association with Use of Drugs That Act on the Renin-Angiotensin-Aldosterone System." *Mini-Sentinel*, July 18, 2011. http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Angioedema-and-RAAS_Protocol.pdf.

Toh, S., R. Platt, J. F. Steiner, and J. S. Brown. "Comparative-effectiveness Research in Distributed Health Data Networks." *Clinical Pharmacology and Therapeutics* 90, no. 6 (2011): 883–887.

Toh, Sengwee, Luis A García Rodríguez, and Miguel A Hernán. "Analyzing Partially Missing Confounder Information in Comparative Effectiveness and Safety Research of Therapeutics." *Pharmacoepidemiology and Drug Safety* 21 Suppl 2 (May 2012): 13–20.

Trivedi, Amal N, Regina C Grebla, Lan Jiang, Jean Yoon, Vincent Mor, and Kenneth W Kizer. "Duplicate Federal Payments for Dual Enrollees in Medicare Advantage

Plans and the Veterans Affairs Health Care System." *JAMA: The Journal of the American Medical Association* 308, no. 1 (July 4, 2012): 67–72.

Tse, Alison, Hung Fu Tseng, Sharon K Greene, Claudia Vellozzi, and Grace M Lee. "Signal Identification and Evaluation for Risk of Febrile Seizures in Children Following Trivalent Inactivated Influenza Vaccine in the Vaccine Safety Datalink Project, 2010-2011." *Vaccine* 30, no. 11 (March 2, 2012): 2024–2031.

Tunis, S. R., D. B. Stryer, and C. M. Clancy. "Practical Clinical Trials: Increasing the Value of Clinical Research for Decision Making in Clinical and Health Policy." *Journal of the American Medical Association* 290, no. 12 (2003): 1624–1632.

U. S. Congressional Budget Office. *The Long-term Outlook for Health Care Spending.* Washington, DC: Congress of the U.S., Congressional Budget Office, 2007. http://www.cbo.gov/doc.cfm?index=8758.

U.S. Congressional Budget Office. *The Budget and Economic Outlook: Fiscal Years 2011 to 2021.* Washington, DC: Congress of the United States, Congressional Budget Office, 2011. http://cbo.gov/doc.cfm?index=12039.

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, and Center for Biologics Evaluation and Research. "Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets (Draft)", February 16, 2011. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM243537.pdf.

———. "Guidance for Industry: Adverse Reactions Section of Labeling for Human Prescription Drug and Biological Products-Content and Format (Final)", January 18, 2006. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM075057.pdf.

———. "Guidance for Industry: Warnings and Precautions, Contraindications, and Boxed Warning Sections of Labeling for Human Prescription Drug and Biological Products - Content and Format (Final)", October 11, 2011. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM075096.pdf.

———. "Guidance: Medication Guides - Distribution Requirements and Inclusion in Risk Evaluation and Mitigation Strategies (REMS) (Final)", November 17, 2011. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM244570.pdf.

U.S. Department of Health and Human Services, Food and Drug Administration, and Center for Drug Evaluation and Research. "Guidance for Industry: Diabetes Mellitus: Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes (Final)", December 17, 2008. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071627.pdf.

———. "Guidance: Classifying Significant Postmarketing Drug Safety Issues (Draft)", March 8, 2012. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM295211.pdf.

U.S. Department of Health and Human Services, Food and Drug Administration. "The Sentinel Initiative". FDA, 2010. http://www.fda.gov/downloads/Safety/FDAsSentinelInitiative/UCM233360.pdf.

U.S. EPA. *Guidelines for Carcinogen Risk Assessment*. Washington, D. C.: U.S. Environmental Protection Agency, 2005. http://www.epa.gov/cancerguidelines/.

U.S. Food and Drug Administration, and Center for Drug Evaluation and Research. "Innovation in Development of Drugs and Biological Products." WebContent, n.d. http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/default.htm.

U.S. General Accounting Office. *Adverse Events: The Magnitude of Health Risk Is Uncertain Because of Limited Incidence Data*. Vol. GAO/HEHS-00–21. Washington, DC: GPO, 2000.

U.S. Government Accountability Office. *FDA Has Begun Efforts to Enhance Postmarket Safety, but Additional Actions Are Needed*. Vol. GAO-10–68. Washington, DC: GPO, 2009. http://www.gao.gov/products/GAO-10-68.

———. *FDA Needs to Enhance Its Oversight of Drugs Approved the Basis of Surrogate Endpoints*. GAO-09-866. Washington, DC: GPO, 2009. http://www.gao.gov/products/GAO-09-866.

U.S. House Committee on Energy and Commerce, and Subcommittee on Oversight and Investigations. *FDA's Role in Protecting the Public Health: Examining FDA's Review of Safety and Efficacy Concerns in Anti-depressant Use by Children*. U.S. G.P.O., 2005.

U.S. Senate Committee on Finance, and U.S. Senate Committee on Finance. *FDA, Merck, and Vioxx: Putting Patient Safety First?* U.S. G.P.O., 2005.

Uhl, K, and P Honig. "Risk Management of Marketed Drugs: FDA and the Interface with the Practice of Medicine." *Pharmacoepidemiology and Drug Safety* 10, no. 3 (May 2001): 205–208.

"Update: Recommendations from the Advisory Committee on Immunization Practices (ACIP) Regarding Administration of Combination MMRV Vaccine." *MMWR. Morbidity and Mortality Weekly Report* 57, no. 10 (March 14, 2008): 258–260.

Vakratsas, Demetrios, and Ceren Kolsarici. "A Dual-market Diffusion Model for a New Prescription Pharmaceutical." *International Journal of Research in Marketing* 25, no. 4 (December 2008): 282–293.

Vandenbroucke, J. P. "What Is the Best Evidence for Determining Harms of Medical Treatment?" *CMAJ : Canadian Medical Association Journal* 174, no. 5 (2006): 645–646.

———. "Why Do the Results of Randomised and Observational Studies Differ?" *BMJ (Clinical Research Ed.)* 343 (2011): d7020.

Vandenbroucke, Jan P, and Bruce M Psaty. "Benefits and Risks of Drug Treatments: How to Combine the Best Evidence on Benefits with the Best Data About Adverse Effects." *JAMA: The Journal of the American Medical Association* 300, no. 20 (November 26, 2008): 2417–2419.

Velentgas, P., R. L. Bohn, J. S. Brown, K. A. Chan, P. Gladowski, C. N. Holick, J. M. Kramer, et al. "A Distributed Research Network Model for Post-marketing Safety

Studies: The Meningococcal Vaccine Study." *Pharmacoepidemiology and Drug Safety* 17, no. 12 (2008): 1226–1234.

Velentgas, Priscilla, Anthony A Amato, Rhonda L Bohn, K Arnold Chan, Thomas Cochrane, Donnie P Funch, Inna Dashevsky, et al. "Risk of Guillain-Barré Syndrome After Meningococcal Conjugate Vaccination." *Pharmacoepidemiology and Drug Safety* (July 16, 2012). http://www.ncbi.nlm.nih.gov/pubmed/22807266.

Wald, A. "Sequential Tests of Statistical Hypotheses." *The Annals of Mathematical Statistics* 16, no. 2 (1945): pp. 117–186.

Whitehead, John. *The Design and Analysis of Sequential Clinical Trials*. Wiley, 1997.

Winsor, Charles P. "The Gompertz Curve as a Growth Curve." *Proceedings of the National Academy of Sciences of the United States of America* 18, no. 1 (January 15, 1932): 1–8.

Woodcock, J. "Evidence Vs. Access: Can Twenty-first-century Drug Regulation Refine the Tradeoffs?" *Clinical Pharmacology and Therapeutics* 91, no. 3 (March 2012): 378–380.

World Health Organization. *The Importance of Pharmacovigilance - Safety Monitoring of Medicinal Products*. Geneva, Switzerland: World Health Organization, 2002. http://apps.who.int/medicinedocs/en/d/Js4893e/.

Wysowski, D. K., and L. Swartz. "Adverse Drug Event Surveillance and Drug Withdrawals in the United States, 1969-2002: The Importance of Reporting Suspected Reactions." *Archives of Internal Medicine* 165, no. 12 (2005): 1363–1369.

Xing, Jian, Howard Burkom, and Jerome Tokars. "Method Selection and Adaptation for Distributed Monitoring of Infectious Diseases for Syndromic Surveillance." *Journal of Biomedical Informatics* 44, no. 6 (December 2011): 1093–1101.

Yih, W Katherine, Swati Deshpande, Candace Fuller, Dawn Heisey-Grove, John Hsu, Benjamin A Kruskal, Martin Kulldorff, et al. "Evaluating Real-time Syndromic Surveillance Signals from Ambulatory Care Data in Four States." *Public Health Reports (Washington, D.C.: 1974)* 125, no. 1 (February 2010): 111–120.

Yih, W. K., M. Kulldorff, B. H. Fireman, I. M. Shui, E. M. Lewis, N. P. Klein, J. Baggs, et al. "Active Surveillance for Adverse Events: The Experience of the Vaccine Safety Datalink Project." *Pediatrics* 127 Suppl 1 (2011): S54–64.

Yih, W. K., J. D. Nordin, M. Kulldorff, E. Lewis, T. A. Lieu, P. Shi, and E. S. Weintraub. "An Assessment of the Safety of Adolescent and Adult Tetanus-diphtheria-acellular Pertussis (Tdap) Vaccine, Using Active Surveillance for Adverse Events in the Vaccine Safety Datalink." *Vaccine* 27, no. 32 (2009): 4257–4262.

12 APPENDIX A - Vaccine Validation Data

12.1 Mini-Sentinel System Vaccine Incident User Data

In sections 7 and 8, I used a simulated vaccine example to illustrate use of the Sequential Database Surveillance Simulator. To model adoption of a new routine childhood vaccination, I proposed a linear adoption function coincident with one-year well-visits for a cohort of 0-1 year olds. Later, data became available with respect to adoption of new vaccinations in the Mini-Sentinel System and I obtained these data via execution of a modular program to validate the example that I presented. I requested data for new childhood vaccinations approved since 2005 as shown in Table 19 below. Data partners that did not have complete data for the entire evaluation period (i.e., date of approval-2010) were censored from the analysis. I fit monthly cumulative adopters who I required to be incident users. Incident users were defined as not having used the medical product in the 30 days preceding the first use, and only users with continuous drug and medical coverage during that time were included.[435]

Trade Name (n=4)	Generic Name (n=4)	Approval Date
Pentacel	Diphtheria and Tetanus Toxoids and Acellular Pertussis Adsorbed, Inactivated Poliovirus and Haemophilus b Conjugate (Tetanus Toxoid Conjugate) Vaccine	06-20-2008
Prevnar-13	Pneumococcal 13-valent Conjugate Vaccine [Diphtheria CRM197 Protein]	02-24-2010
Rotateq	Rotavirus Vaccine, Live, Oral, Pentavalent	02-03-2006
ProQuad	Measles, Mumps, Rubella and Varicella Virus Vaccine Live Lyophilized preparation for subcutaneous injection	09-06-2005

Table 19. Vaccine Validation Cohort

In early 2007, Merck, the manufacturer of ProQuad®, reported shortages of the varicella-zoster virus and its subsequent prioritization of other varicella-containing vaccinations over ProQuad®.[436] Therefore, I censored ProQuad® data at the end of 2006. Later, based on safety risk data developed during sequential database surveillance in the

[435] A 45-day enrollment gap was allowed.
[436] "Notice to Readers: Update on Supply of Vaccines Containing Varicella-Zoster Virus," *JAMA: The Journal of the American Medical Association* 298, no. 7 (August 15, 2007): 736.

Vaccine Safety Datalink, the Advisory Committee on Immunization Practices changed its recommendations on the use of ProQuad®, specifically eliminating language that indicated a preference for the combined vaccine over its separately available components (i.e., measles, mumps, and rubella vaccination AND varicella vaccination).[437] In subsequent years, ProQuad® has only been intermittently available, reflecting its low prioritization.

12.2 Model Fits

I used the generalized linear model fitting functions of MATLAB® (R2012a). I report the model fits below. Table 20 shows the linear fits when using data from the date of approval until the end of 2010. Both the intercept and coefficient terms were significant at the 0.005 level for all four models. Table 21 shows the linear fits when the first three months of adoption data are excluded, which accounts for uneven uptake among the different data partners that comprise the Mini-Sentinel System. With this adjustment, the intercept and coefficient terms were significant at the 0.0005 level for all four models. The R^2 and adjusted R^2 values are uniformly higher with these initial datapoints excluded because the adoption pattern begins to reflect the steady state of new vaccinations. In conclusion, a linear adoption function is appropriate to model adoption of new childhood vaccinations.

Trade Name (n=4)	Approval Date	RMSE	R^2	Adjusted R^2
Pentacel	06-20-2008	1.50E4	0.991	0.991
Prevnar-13	02-24-2010	2.88E4	0.983	0.982
Rotateq	02-03-2006	3.43E4	0.983	0.982
ProQuad*	09-06-2005	8.67E3	0.827	0.815

Table 20. Linear Regression Model for the Vaccine Cohort
*ProQuad® data censored from 2007-2010.
Abbreviations: RMSE, root mean square error.

[437] "Update: Recommendations from the Advisory Committee on Immunization Practices (ACIP) Regarding Administration of Combination MMRV Vaccine," *MMWR. Morbidity and Mortality Weekly Report* 57, no. 10 (March 14, 2008): 258–260.

Trade Name (n=4)	Approval Date	RMSE	R^2	Adjusted R^2
Pentacel	06-20-2008	6.48E3	0.998	0.998
Prevnar-13	02-24-2010	1.46E4	0.994	0.993
Rotateq	02-03-2006	2.65E4	0.989	0.989
ProQuad*	09-06-2005	6.93E3	0.896	0.886

Table 21. Linear Regression Model for the Vaccine Cohort with the First Three Months Excluded
*ProQuad® data censored from 2007-2010.
Abbreviations: RMSE, root mean square error.

THIS PAGE INTENTIONALLY LEFT BLANK

13 APPENDIX B – Medicaid Dispensing Data Simulations

13.1 Common Event Rates

Generic Name (n=40)	Total Exposed Days after 8 quarters	Range of detectable IRRs after 8 quarters	Total Exposed Days after 15 quarters	Range of detectable IRRs after 15 quarters
New Molecular Entities Approved in 2004				
tinidazole[a]	36,110	5-10	101,133	2.75-10
rifaximin[b]	206,795	2.25-5	485,250	1.75-2.75
erlotinib	434,685	1.75-2.25	762,915	1.75-2.75
insulin glulisine injection	446,310	1.75-2.75	1,179,420	1.5-2
lanthanum	1,024,095	1.5-2	1,634,145	1.5-2
trospium	993,940	1.5-2.25	1,758,405	1.5-2.75
acamprosate	1,727,108	1.5-1.75	3,529,605	1.33-1.75
cinacalcet	3,052,635	1.33-1.75	5,006,580	1.33-1.5
omega-3 acid ethyl esters	2,223,360	1.33-1.75	7,896,930	1.2-1.5
darifenacin	4,007,340	1.33-1.5	8,366,490	1.2-1.33
solifenacin	3,566,535	1.33-1.75	8,982,105	1.2-1.33
eszopiclone	19,705,410	1.2-1.33	33,866,880	1.2
tiotropium	33,532,470	1.2	58,154,685	1.2
pregabalin	39,717,375	1.2	79,711,665	1.2
duloxetine	45,217,020	1.2	92,261,400	1.2
New Molecular Entities Approved in 2005				
lenalidomide[c]	73,164	3-10	233,996	2.25-5
sorafenib toylate	78,345	3-10	237,585	2.25-5
tipranavir	300,900	2-3	437,745	1.75-2.75
nepafenac[b]	393,570	2-2.75	762,075	1.75-2.25
pramlintide acetate	371,190	2-3	903,300	1.75-2.25
entecavir	272,370	2-5	1,129,140	1.5-2
deferasirox	725,865	1.75-2.25	1,592,055	1.5-2
exenatide	3,341,895	1.33-1.75	8,517,255	1.2-1.33
ramelteon	4,816,590	1.33-1.5	10,779,270	1.2-1.33
insulin detemir	2,437,695	1.33-1.75	12,483,555	1.2-1.33

Table 22. Detectable Incidence Rate Ratios for Common Events using the Poisson MaxSPRT Model
Estimates based on power=.80, alpha=.05. One prescription is assumed to be a 30-day supply unless otherwise specified. Range is based on sensitivity analyses of doubling and halving the days exposed.
[a]indicates one prescription was equivalent to a 5-day supply.
[b]indicates one prescription was equivalent to a 15-day supply.
[c]indicates one prescription was equivalent to a 28-day supply.
[d]indicates that 10 quarters of data were available.
[e]indicates one prescription was equivalent to a 14-day supply.
[f]indicates one prescription was equivalent to a 10-day supply.
Abbreviations: IRR, incidence rate ratio; MaxSPRT, Maximized Sequential Probability Ratio Test.

Generic Name (n=40)	Total Exposed Days after 8 quarters	Range of detectable IRRs after 8 quarters	Total Exposed Days after 15 quarters	Range of detectable IRRs after 15 quarters
New Molecular Entities Approved in 2006				
vorinostat	6,180	100	11,910	10-100
kunecatechins[d]	6,630	100	**	**
biskalcitrate potassium, metronidazole and tetracycline hydrochloride[e]	30,100	5-10	59,480	5-10
posaconazole[f]	32,508	5-10	63,672	5-10
dasatinib	59,820	5-10	143,880	2.5-5
rasagiline mesylate	69,060	5-10	170,520	2.25-5
sunitinib malate[c]	80,668	3-10	175,672	2.25-5
telbivudine	164,580	2.5-5	381,330	2-2.75
ranolazine	553,215	1.75-2.5	1,784,115	1.5-1.75
ciclesonide[d]	821,835	1.75-2.25	**	**
darunavir	966,000	1.5-2.25	3,346,500	1.33-1.75
lubiprostone	2,066,100	1.33-1.75	5,995,890	1.2-1.5
sitagliptin phosphate	7,960,500	1.2-1.5	20,810,620	1.2-1.33
paliperidone	9,789,630	1.2-1.33	23,213,280	1.2
varenicline	17,177,985	1.2-1.33	30,531,405	1.2

Table 22 (Continued). Detectable Incidence Rate Ratios for Common Events using the Poisson MaxSPRT Model

Estimates based on power=.80, alpha=.05. One prescription is assumed to be a 30-day supply unless otherwise specified. Range is based on sensitivity analyses of doubling and halving the days exposed.

[a]indicates one prescription was equivalent to a 5-day supply.

[b]indicates one prescription was equivalent to a 15-day supply.

[c]indicates one prescription was equivalent to a 28-day supply.

[d]indicates that 10 quarters of data were available.

[e]indicates one prescription was equivalent to a 14-day supply.

[f]indicates one prescription was equivalent to a 10-day supply.

Abbreviations: IRR, incidence rate ratio; MaxSPRT, Maximized Sequential Probability Ratio Test.

13.2 Infrequent Event Rates

Generic Name (n=40)	Total Exposed Days after 8 quarters	Range of detectable IRRs after 8 quarters	Total Exposed Days after 15 quarters	Range of detectable IRRs after 15 quarters
New Molecular Entities Approved in 2004				
tinidazole[a]	36,110	100	101,133	10-100
rifaximin[b]	206,795	10-100	485,250	5-10
erlotinib	434,685	5-10	762,915	3-10
insulin glulisine injection	446,310	5-10	1,179,420	2.75-5
lanthanum	1,024,095	2.75-5	1,634,145	2.5-5
trospium	993,940	2.75-10	1,758,405	2.25-5
acamprosate	1,727,108	2.25-5	3,529,605	2-3
cinacalcet	3,052,635	2-3	5,006,580	1.75-2.5
omega-3 acid ethyl esters	2,223,360	2.25-5	7,896,930	1.75-2.25
darifenacin	4,007,340	2-2.75	8,366,490	1.75-2.25
solifenacin	3,566,535	2-3	8,982,105	1.75-2.25
eszopiclone	19,705,410	1.5-1.75	33,866,880	1.33-1.75
tiotropium	33,532,470	1.33-1.75	58,154,685	1.2-1.5
pregabalin	39,717,375	1.33-1.5	79,711,665	1.2-1.5
duloxetine	45,217,020	1.33-1.5	92,261,400	1.2-1.33
New Molecular Entities Approved in 2005				
lenalidomide[c]	73,164	10-100	233,996	10-100
sorafenib toylate	78,345	10-100	237,585	10-100
tipranavir	300,900	5-10	437,745	5-10
nepafenac[b]	393,570	5-10	762,075	3-10
pramlintide acetate	371,190	5-10	903,300	3-10
entecavir	272,370	5-100	1,129,140	2.75-5
deferasirox	725,865	5-10	1,592,055	2.5-5
exenatide	3,341,895	2-3	8,517,255	1.75-2.25
ramelteon	4,816,590	1.75-2.75	10,779,270	1.5-2
insulin detemir	2,437,695	2.25-5	12,483,555	1.5-2

Table 23. Detectable Incidence Rate Ratios for Infrequent Events using the Poisson MaxSPRT Model
Estimates based on power=.80, alpha=.05. One prescription is assumed to be a 30-day supply unless otherwise specified. Range is based on sensitivity analyses of doubling and halving the days exposed.
[a]indicates one prescription was equivalent to a 5-day supply.
[b]indicates one prescription was equivalent to a 15-day supply.
[c]indicates one prescription was equivalent to a 28-day supply.
[d]indicates that 10 quarters of data were available.
[e]indicates one prescription was equivalent to a 14-day supply.
[f]indicates one prescription was equivalent to a 10-day supply.
Abbreviations: IRR, incidence rate ratio; MaxSPRT, Maximized Sequential Probability Ratio Test.

Generic Name (n=40)	Total Exposed Days after 8 quarters	Range of detectable IRRs after 8 quarters	Total Exposed Days after 15 quarters	Range of detectable IRRs after 15 quarters
New Molecular Entities Approved in 2006				
vorinostat	6,180	100-1000	11,910	100-1000
kunecatechins[d]	6,630	100-1000	**	**
biskalcitrate potassium, metronidazole and tetracycline hydrochloride[e]	30,100	100	59,480	100
posaconazole[f]	32,508	100	63,672	100
dasatinib	59,820	100	143,880	10-100
rasagiline mesylate	69,060	100	170,520	10-100
sunitinib malate[c]	80,668	10-100	175,672	10-100
telbivudine	164,580	10-100	381,330	5-10
ranolazine	553,215	5-10	1,784,115	2.25-5
ciclesonide[d]	821,835	3-10	**	**
darunavir	966,000	2.75-10	3,346,500	2-3
lubiprostone	2,066,100	2.25-5	5,995,890	1.75-2.5
sitagliptin phosphate	7,960,500	1.75-2.25	20,810,620	1.33-1.75
paliperidone	9,789,630	1.5-2.25	23,213,280	1.33-1.75
varenicline	17,177,985	1.5-1.75	30,531,405	1.33-1.75

Table 23 (Continued). Detectable Incidence Rate Ratios for Infrequent Events using the Poisson MaxSPRT Model

Estimates based on power=.80, alpha=.05. One prescription is assumed to be a 30-day supply unless otherwise specified. Range is based on sensitivity analyses of doubling and halving the days exposed.

[a]indicates one prescription was equivalent to a 5-day supply.

[b]indicates one prescription was equivalent to a 15-day supply.

[c]indicates one prescription was equivalent to a 28-day supply.

[d]indicates that 10 quarters of data were available.

[e]indicates one prescription was equivalent to a 14-day supply.

[f]indicates one prescription was equivalent to a 10-day supply.

Abbreviations: IRR, incidence rate ratio; MaxSPRT, Maximized Sequential Probability Ratio Test.

13.3 Rare Event Rates

Generic Name (n=40)	Total Exposed Days after 8 quarters	Range of detectable IRRs after 8 quarters	Total Exposed Days after 15 quarters	Range of detectable IRRs after 15 quarters
New Molecular Entities Approved in 2004				
tinidazole[a]	36,110	1000	101,133	100-1000
rifaximin[b]	206,795	100	485,250	100
erlotinib	434,685	100	762,915	10-100
insulin glulisine injection	446,310	100	1,179,420	10-100
lanthanum	1,024,095	10-100	1,634,145	10-100
trospium	993,940	10-100	1,758,405	10-100
acamprosate	1,727,108	10-100	3,529,605	5-10
cinacalcet	3,052,635	5-10	5,006,580	5-10
omega-3 acid ethyl esters	2,223,360	10-100	7,896,930	3-10
darifenacin	4,007,340	5-10	8,366,490	3-10
solifenacin	3,566,535	5-10	8,982,105	3-10
eszopiclone	19,705,410	2.25-5	33,866,880	2-3
tiotropium	33,532,470	2-3	58,154,685	1.75-2.5
pregabalin	39,717,375	2-2.75	79,711,665	1.75-2.25
duloxetine	45,217,020	1.75-2.75	92,261,400	1.5-2.25
New Molecular Entities Approved in 2005				
lenalidomide[c]	73,164	100-1000	233,996	100
sorafenib toylate	78,345	100-1000	237,585	100
tipranavir	300,900	100	437,745	100
nepafenac[b]	393,570	100	762,075	10-100
pramlintide acetate	371,190	100	903,300	10-100
entecavir	272,370	100	1,129,140	10-100
deferasirox	725,865	100	1,592,055	10-100
exenatide	3,341,895	5-10	8,517,255	3-10
ramelteon	4,816,590	5-10	10,779,270	2.75-5
insulin detemir	2,437,695	10-100	12,483,555	2.5-5

Table 24. Detectable Incidence Rate Ratios for Rare Events using the Poisson MaxSPRT Model
Estimates based on power=.80, alpha=.05. One prescription is assumed to be a 30-day supply unless otherwise specified. Range is based on sensitivity analyses of doubling and halving the days exposed.
[a]indicates one prescription was equivalent to a 5-day supply.
[b]indicates one prescription was equivalent to a 15-day supply.
[c]indicates one prescription was equivalent to a 28-day supply.
[d]indicates that 10 quarters of data were available.
[e]indicates one prescription was equivalent to a 14-day supply.
[f]indicates one prescription was equivalent to a 10-day supply.
Abbreviations: IRR, incidence rate ratio; MaxSPRT, Maximized Sequential Probability Ratio Test.

Generic Name (n=40)	Total Exposed Days after 8 quarters	Range of detectable IRRs after 8 quarters	Total Exposed Days after 15 quarters	Range of detectable IRRs after 15 quarters
New Molecular Entities Approved in 2006				
vorinostat	6,180	>=1000	11,910	>=1000
kunecatechins[d]	6,630	>=1000	**	**
biskalcitrate potassium, metronidazole and tetracycline hydrochloride[e]	30,100	1000	59,480	100-1000
posaconazole[f]	32,508	1000	63,672	100-1000
dasatinib	59,820	100-1000	143,880	100-1000
rasagiline mesylate	69,060	100-1000	170,520	100
sunitinib malate[c]	80,668	100-1000	175,672	100
telbivudine	164,580	100	381,330	100
ranolazine	553,215	100	1,784,115	10-100
ciclesonide[d]	821,835	10-100	**	**
darunavir	966,000	10-100	3,346,500	5-10
lubiprostone	2,066,100	10-100	5,995,890	5-10
sitagliptin phosphate	7,960,500	3-10	20,810,620	2.25-5
paliperidone	9,789,630	2.75-10	23,213,280	2.25-5
varenicline	17,177,985	2.25-5	30,531,405	2-3

Table 24 (Continued). Detectable Incidence Rate Ratios for Rare Events using the Poisson MaxSPRT Model

Estimates based on power=.80, alpha=.05. One prescription is assumed to be a 30-day supply unless otherwise specified. Range is based on sensitivity analyses of doubling and halving the days exposed.

[a]indicates one prescription was equivalent to a 5-day supply.

[b]indicates one prescription was equivalent to a 15-day supply.

[c]indicates one prescription was equivalent to a 28-day supply.

[d]indicates that 10 quarters of data were available.

[e]indicates one prescription was equivalent to a 14-day supply.

[f]indicates one prescription was equivalent to a 10-day supply.

Abbreviations: IRR, incidence rate ratio; MaxSPRT, Maximized Sequential Probability Ratio Test.

13.4 Very Rare Event Rates

Generic Name (n=40)	Total Exposed Days after 8 quarters	Range of detectable IRRs after 8 quarters	Total Exposed Days after 15 quarters	Range of detectable IRRs after 15 quarters
New Molecular Entities Approved in 2004				
tinidazole[a]	36,110	>1000	101,133	>=1000
rifaximin[b]	206,795	1000	485,250	100-1000
erlotinib	434,685	100-1000	762,915	100-1000
insulin glulisine injection	446,310	100-1000	1,179,420	100-1000
lanthanum	1,024,095	100-1000	1,634,145	100
trospium	993,940	100-1000	1,758,405	100
acamprosate	1,727,108	100	3,529,605	100
cinacalcet	3,052,635	100	5,006,580	100
omega-3 acid ethyl esters	2,223,360	100	7,896,930	10-100
darifenacin	4,007,340	100	8,366,490	10-100
solifenacin	3,566,535	100	8,982,105	10-100
eszopiclone	19,705,410	10-100	33,866,880	5-10
tiotropium	33,532,470	5-10	58,154,685	5-10
pregabalin	39,717,375	5-10	79,711,665	3-10
duloxetine	45,217,020	5-10	92,261,400	3-10
New Molecular Entities Approved in 2005				
lenalidomide[c]	73,164	>=1000	233,996	1000
sorafenib toylate	78,345	>=1000	237,585	1000
tipranavir	300,900	1000	437,745	100-1000
nepafenac[b]	393,570	100-1000	762,075	100-1000
pramlintide acetate	371,190	100-1000	903,300	100-1000
entecavir	272,370	1000	1,129,140	100-1000
deferasirox	725,865	100-1000	1,592,055	100
exenatide	3,341,895	100	8,517,255	10-100
ramelteon	4,816,590	100	10,779,270	10-100
insulin detemir	2,437,695	100	12,483,555	10-100

Table 25. Detectable Incidence Rate Ratios for Very Rare Events using the Poisson MaxSPRT Model
Estimates based on power=.80, alpha=.05. One prescription is assumed to be a 30-day supply unless otherwise specified. Range is based on sensitivity analyses of doubling and halving the days exposed.
[a]indicates one prescription was equivalent to a 5-day supply.
[b]indicates one prescription was equivalent to a 15-day supply.
[c]indicates one prescription was equivalent to a 28-day supply.
[d]indicates that 10 quarters of data were available.
[e]indicates one prescription was equivalent to a 14-day supply.
[f]indicates one prescription was equivalent to a 10-day supply.
Abbreviations: IRR, incidence rate ratio; MaxSPRT, Maximized Sequential Probability Ratio Test.

New Molecular Entities Approved in 2006				
vorinostat	6,180	>1000	11,910	>=1000
kunecatechins[d]	6,630	>1000	**	**
biskalcitrate potassium, metronidazole and tetracycline hydrochloride[e]	30,100	>1000	59,480	>=1000
posaconazole[f]	32,508	>1000	63,672	>=1000
dasatinib	59,820	>=1000	143,880	>=1000
rasagiline mesylate	69,060	>=1000	170,520	1000
sunitinib malate[c]	80,668	>=1000	175,672	1000
telbivudine	164,580	1000	381,330	100-1000
ranolazine	553,215	100-1000	1,784,115	100
ciclesonide[d]	821,835	100-1000	**	**
darunavir	966,000	100-1000	3,346,500	100
lubiprostone	2,066,100	100	5,995,890	100
sitagliptin phosphate	7,960,500	10-100	20,810,620	10-100
paliperidone	9,789,630	10-100	23,213,280	10-100
varenicline	17,177,985	10-100	30,531,405	5-10

Table 25 (Continued). Detectable Incidence Rate Ratios for Very Rare Events using the Poisson MaxSPRT Model

Estimates based on power=.80, alpha=.05. One prescription is assumed to be a 30-day supply unless otherwise specified. Range is based on sensitivity analyses of doubling and halving the days exposed.

[a]indicates one prescription was equivalent to a 5-day supply.

[b]indicates one prescription was equivalent to a 15-day supply.

[c]indicates one prescription was equivalent to a 28-day supply.

[d]indicates that 10 quarters of data were available.

[e]indicates one prescription was equivalent to a 14-day supply.

[f]indicates one prescription was equivalent to a 10-day supply.

Abbreviations: IRR, incidence rate ratio; MaxSPRT, Maximized Sequential Probability Ratio Test.

14 APPENDIX C – New Molecular Entity Cohort Model Fits

This appendix contains individual figures with respect to the subset of new molecular entities that were examined for uptake patterns in the Mini-Sentinel System. A select group of the model functional forms are displayed. The pattern and the model fits are observable, but the cumulative adoptions are deliberately removed because they are not authorized to be shared publicly. Also, I have a included a brief summary of each drug's indications, approval dates, and other pieces of information that may have impacted adoption. These summaries are based on data contained in the Drugs@FDA database and are not meant to be an exhaustive discussion of the therapeutic in question. Launch dates were verified with press releases from manufacturers.

14.1 Amitiza® (lubiprostone)

Amitiza® (lubiprostone) was approved January 31, 2006 and became commercially available in the United States in April 2006. It was a first-in-class chloride channel activator approved for chronic idiopathic constipation. A competitor, Zelnorm® (tegaserod) was voluntarily withdrawn from the market on March 30, 2007 due to cardiovascular risks, and a large increase in utilization of Amitiza® is observable at that time. Zelnorm® was returned to limited use as a treatment investigational new drug, but was then withdrawn completely on April 2, 2008. Additionally, in April 2008, Amitiza® received a supplemental approval for the treatment of irritable bowel syndrome with constipation in women over 18 years old.

Visually, the dual market models are superior to the single market model. In this case, the existence of two markets may be well-explained by the indication expansion although there was presumably some level of off-label use for this indication prior to the new indication approval date. Both the indications that Amitiza® is approved for are diseases that exist on a spectrum in severity, which is one theoretical construct hypothesized to explain the dual market phenomena. All parameters of the single market Bass model were significant at the 0.05 level.

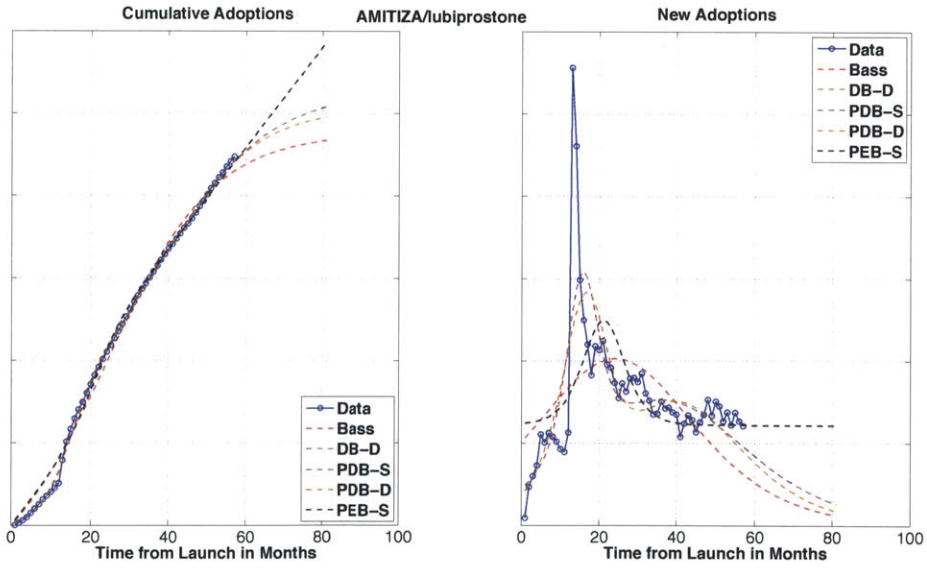

Figure 30. Amitiza® (lubiprostone) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.2 Byetta® (exenatide)

Byetta® (exenatide) was approved April 28, 2005 and became commercially available in the United States in June 2005. It was approved as a first-in-class glucagon-like peptide-1 agonist to be used an adjunct therapy for type 2 diabetics in conjunction with other diabetes medications (i.e., metformin and sulfonylurea). On October 16, 2006, a competitor product - Januvia® (sitagliptin) – was approved as a monotherapy for type 2 diabetics and as an adjunct therapy. Januvia® was also a first-in-class therapeutic in a competing class (i.e., a dipeptidyl peptidase-4 inhibitor). Warnings regarding an increased risk of acute pancreatitis were added to Byetta®'s label in October 2007. In November of 2009, Byetta® was approved as a first-line monotherapy. On January 26, 2010, the second therapeutic glucagon-like peptide-1 agonist was approved, Victoza® (liraglutide), although as an add-on therapy and with a black box warning for pancreatitis and thyroid cancer.

Visually, there are not notable differences between the single market model and most of the dual market models. The only dual market model that fits the first large spike in adoption is the non-pooled dual Bass model with a dynamic probability. All parameters of the single market Bass model were significant at the 0.05 level.

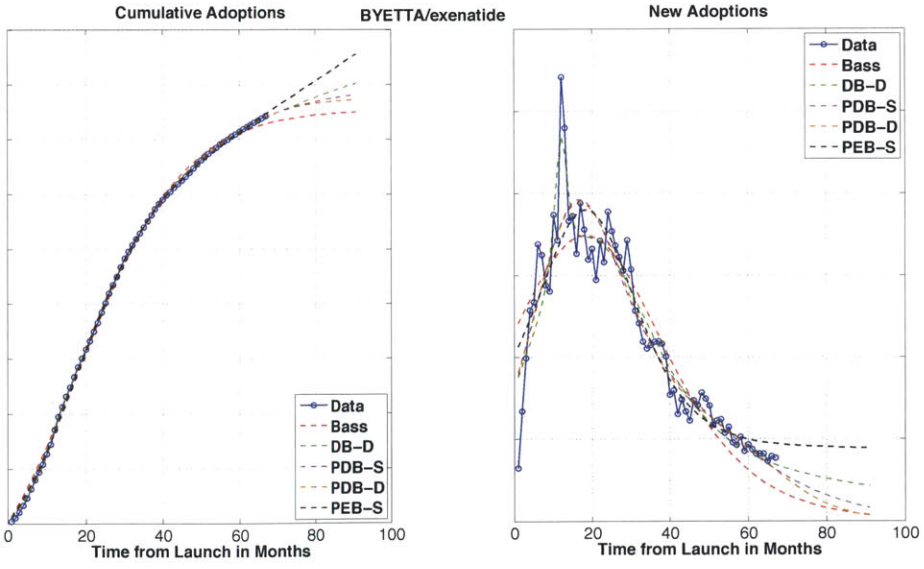

Figure 31. Byetta® (exenatide) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.3 Campral® (acamprosate)

Campral® (acamprosate) was approved July 29, 2004 and became commercially available in the United States in 2005. It was approved as a first-in-class gamma-aminobutyric acid- type A receptors for the supportive treatment for alcoholism recovery. A competitor in another pharmacologic class, Vivitrol® (naltrexone injection), was approved in April 2006.

Visually, all the models are fairly similar although the pooled dual Bass model with a dynamic probability is predicting a rather large secondary market for this therapeutic. In general, there seems to be less evidence of a distinction or clear superiority of a dual market model over a single market model. All parameters of the single market Bass model were significant at the 0.05 level.

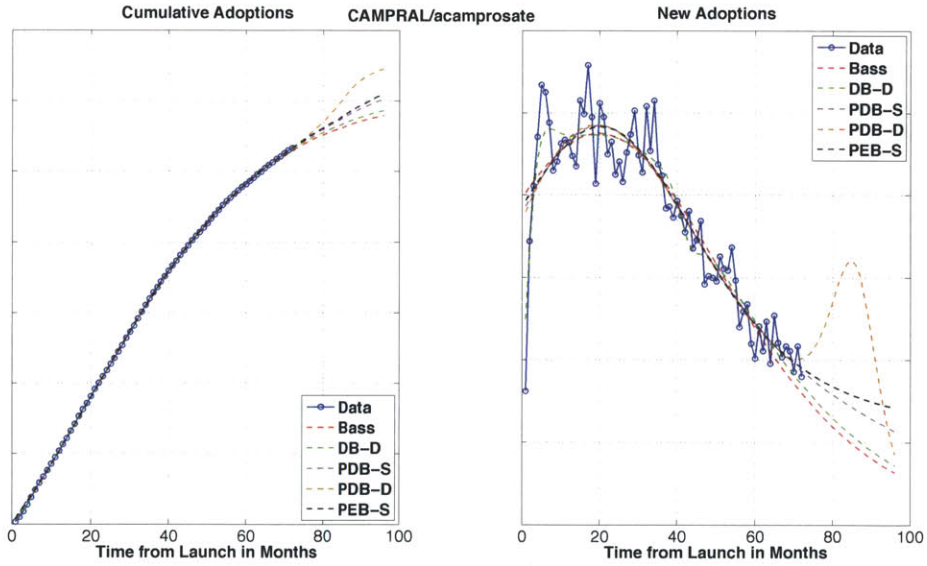

Figure 32. Campral® (acamprosate) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.4 Chantix® (varenicline)

Chantix® (varenicline) was approved May 10, 2006 and became commercially available in the United States in August 2, 2006. It was approved as a treatment for smoking cessation and was a first-in-class therapeutic. In November 2007, the FDA began reviewing a potential risk of suicidal thoughts and aggressive or erratic behavior associated with Chantix®.

Visually, the dual market Bass models are better fits than the single market Bass model. The dynamic non-pooled dual Bass model reduces to the single market Bass model. All parameters of the single market Bass model were significant at the 0.05 level.

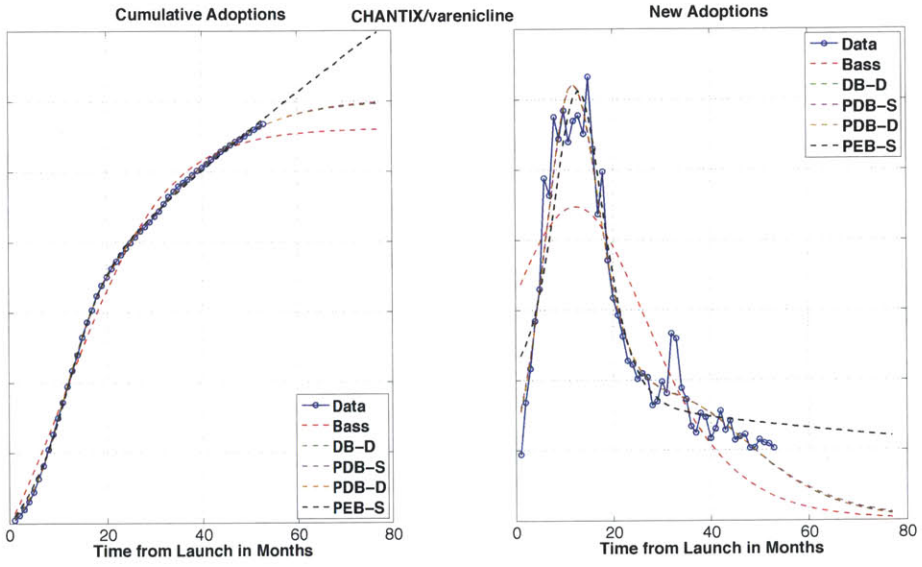

Figure 33. Chantix® (varenicline) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit.

14.5 Cymbalta® (duloxetine)

Cymbalta® (duloxetine) was approved August 3, 2004 and became commercially available in the United States immediately post-approval. It is a selective serotonin and norepinephrine reuptake inhibitor approved as a treatment for depression among adults. It was not a first-in-class therapeutic. On November 8, 2004, it received a subsequent approval for use in the management of diabetic peripheral neuropathic pain. In February 2007, it received another approval for generalized anxiety disorder. In December 2007, it received another approved as a maintenance therapy for major depressive disorder in adult patients. On March 3, 2008, a competitor drug in the same pharmacologic class, Pristiq® (desvenlafaxine), was approved. On June 13, 2008, Cymbalta® received another supplemental approval for the treatment of fibromyalgia. This approval happened to coincide with the generic entry of the first-in-class drug for this class. On November 4, 2010, Cymbalta® received another supplemental indication approval for the treatment of chronic musculoskeletal pain.

Visually, only the pooled dual Bass market model with a static probability correctly identifies the early market. The dynamic dual Bass market model reduces to the single Bass market model. All parameters of the single market Bass model and the static pooled exponential-Bass model were significant at the 0.05 level.

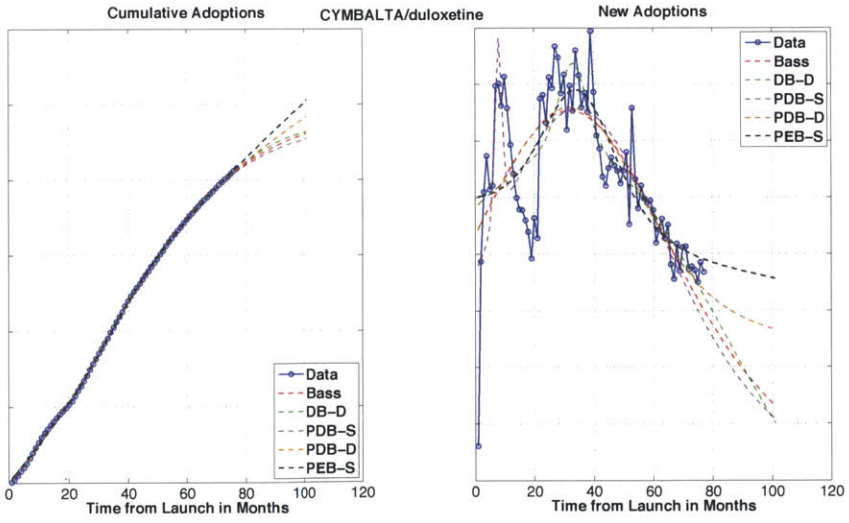

Figure 34. Cymbalta® (duloxetine) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.6 Enablex® (darifenacin)

Enablex® (darifenacin) was approved December 22, 2004 and became commercially available in the United States after February 9, 2005. It was an antispasmodic/anticholinergic approved as a treatment for overactive bladder and was the sixth-in-class. Two other drugs in this class - Sanctura® (trospium) and Vesicare® (solifenacin) - were also approved in the same year and are a part of this cohort.

Visually, the dual Bass models perform similarly. The dual market models are a better fit than the single market Bass model. All parameters of the single market Bass model and the static pooled dual Bass market model were significant at the 0.05 level.

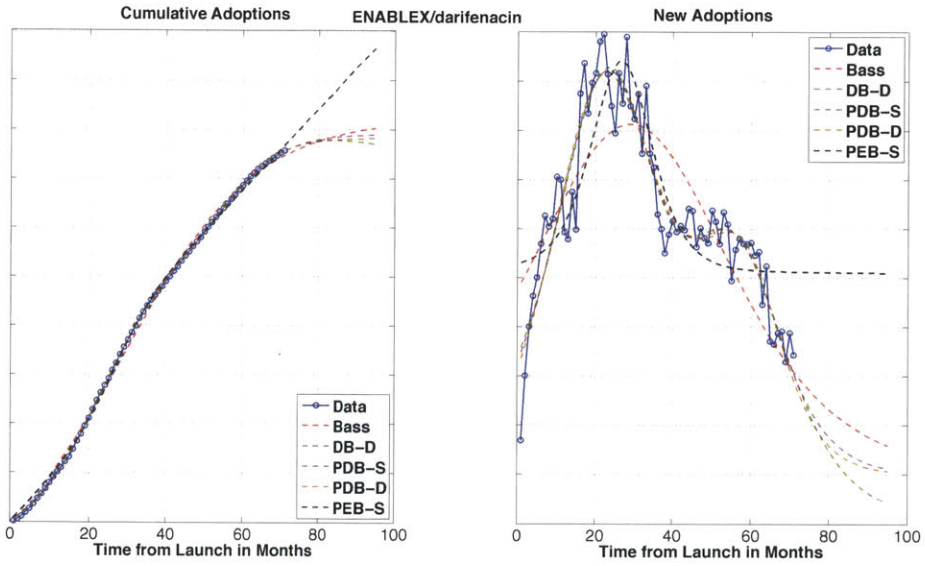

Figure 35. Enablex® (darifenacin) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.7 Invega® (paliperidone)

Invega® (paliperidone) was approved December 19, 2006 and became commercially available in the United States after January 7, 2007. It was approved for acute treatment of schizophrenia and was not a first-in-class therapeutic. As of April 30, 2007, it was approved for the maintenance (long-term) treatment of schizophrenia. On August 3, 2009, Invega® was approved for schizoaffective disorder treatment as either a monotherapy or adjunctive therapy. At the same time, an extended-release (i.e., once monthly) injectable suspension version of the drug was approved. For the purposes of incident users, I considered the injectable version to be a different drug.

Visually, the dual Bass market models are similar and clearly better than the single Bass market model. All parameters of the single market Bass model were significant at the 0.05 level.

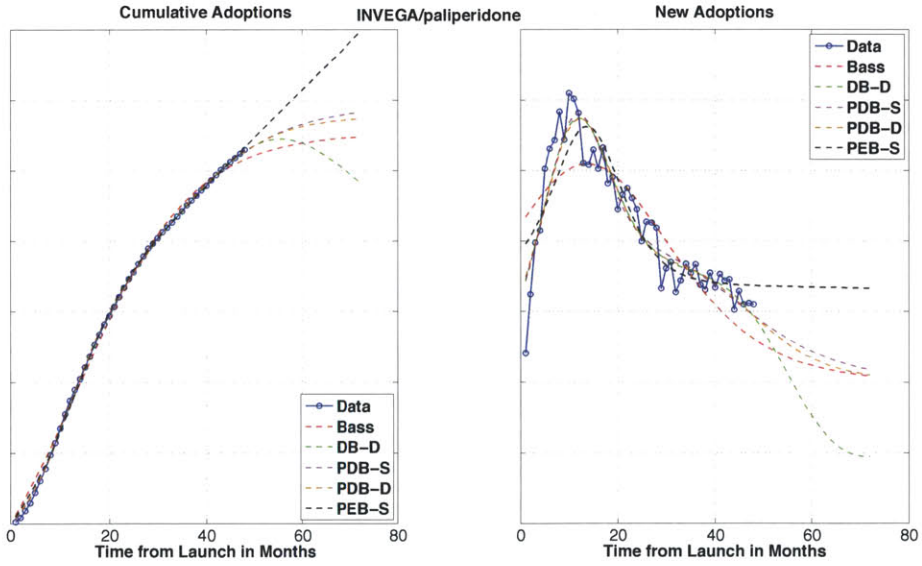

Figure 36. Invega® (paliperidone) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.8 Januvia® (sitagliptin phosphate)

Januvia® (sitagliptin phosphate) was approved October 16, 2006 and became commercially available in the United States immediately postapproval. It was the first-in-class dipeptidyl peptidase-4 inhibitor approved as a monotherapy or adjunct therapy for type 2 diabetics. On October 18, 2007, Januvia received additional approvals as an adjunct therapy. In August 2009, a competitor in the same class, Onglyza® (saxagliptin) was approved. Shortly thereafter, Januvia®'s label was amended to include reports of acute pancreatitis.

Visually, all the dual market models outperform the single market model. All parameters of the single market Bass model were significant at the 0.05 level.

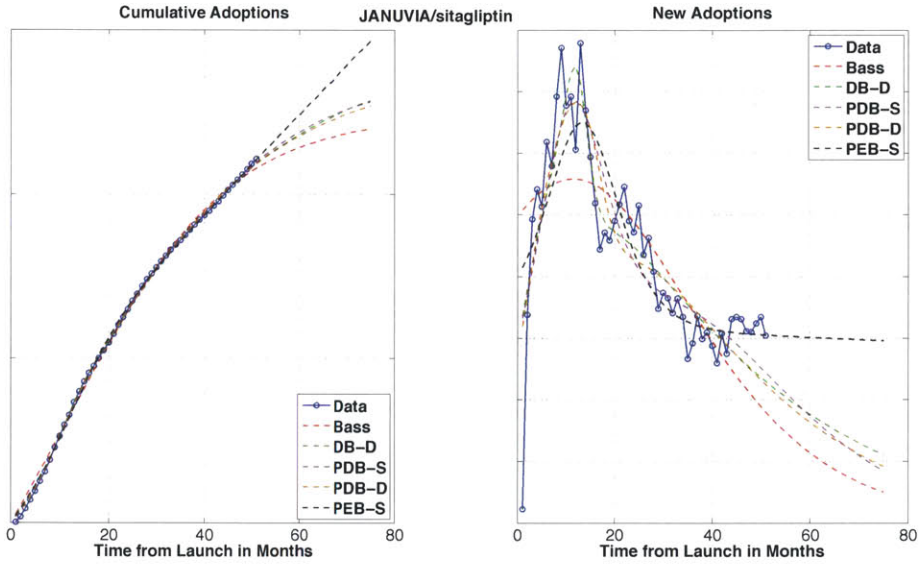

Figure 37. Januvia® (sitagliptin) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.9 Levemir® (insulin detemir)

Levemir® (insulin detemir) was approved June 16, 2005 and became commercially available in the United States in March 2006. It was approved for the treatment of adult patients with Type 1 or Type 2 diabetes mellitus who require basal (long-acting) insulin for the control of hyperglycemia. It was not a first-in-class therapeutic.

Visually, the dual market models outperform the single market models with the dynamic non-pooled dual Bass market model getting the slight edge. All parameters of the single market Bass model and the static pooled dual Bass market model were significant at the 0.05 level.

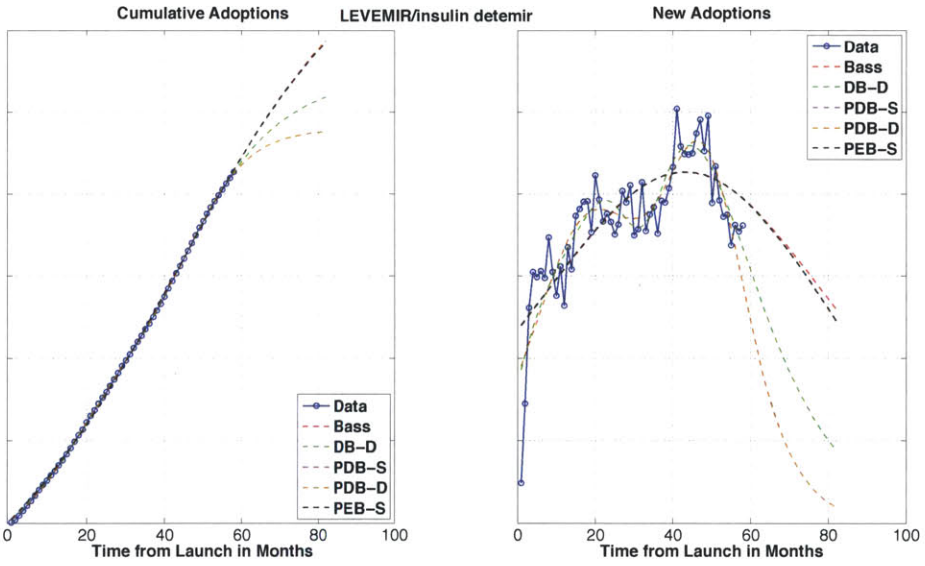

Figure 38. Levemir® (insulin detemir) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.10 Lunesta® (eszopiclone)

Lunesta® (eszopiclone) was approved December 16, 2004 and became commercially available in the United States after April 3, 2005. It was a non-benzodiazepine hypnotic agent approved for the treatment of insomnia and is a scheduled drug. A non-scheduled competitor drug, Rozerem® (ramelteon), was approved later that year.

Visually, this is one of the instances when the dual market models all outperform the single market model, but also one when the exponential-Bass dual market model does well. All parameters in the dynamic non-pooled dual Bass model and in the static pooled exponential-Bass were significant.

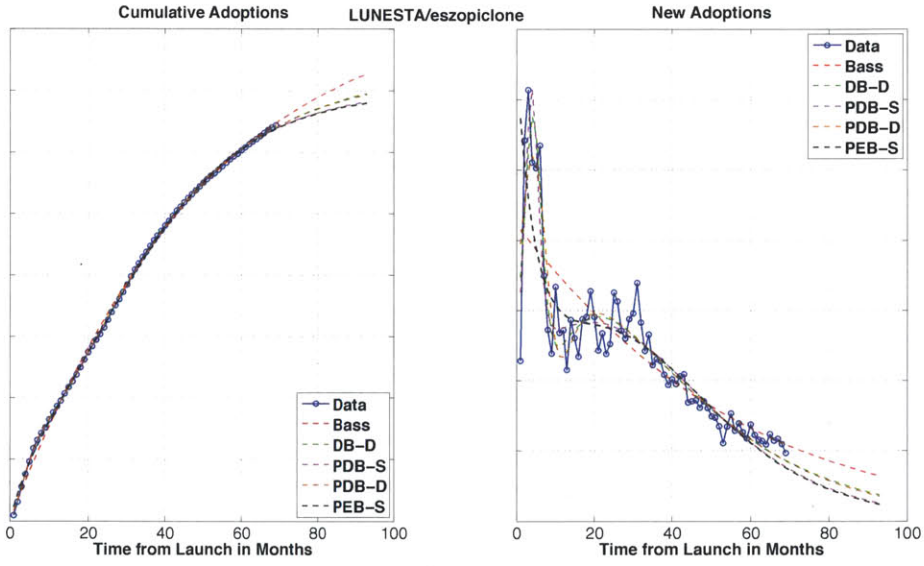

Figure 39. Lunesta® (eszopiclone) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.11 Lyrica® (pregabalin)

Lyrica® (pregabalin) was approved December 31, 2004 and became commercially available in the United States after September 22, 2005. It was approved for the management of neuropathic pain associated with diabetic peripheral neuropathy and post-herpetic neuralgia. It is not a first-in-class therapeutic. On June 14, 2005, it received an additional approval as an adjunct treatment for partial seizures. It was also designated as a controlled substance. As of June 22, 2007, Lyrica® was additionally approved for the management of fibromyalgia.

Visually, only the dynamic non-pooled dual Bass market model identifies the early market closely whereas the static pooled dual Bass market model performs better on the latter market peak. All parameters of the single market Bass model and the pooled static exponential-Bass dual market model are significant at the 0.05 level.

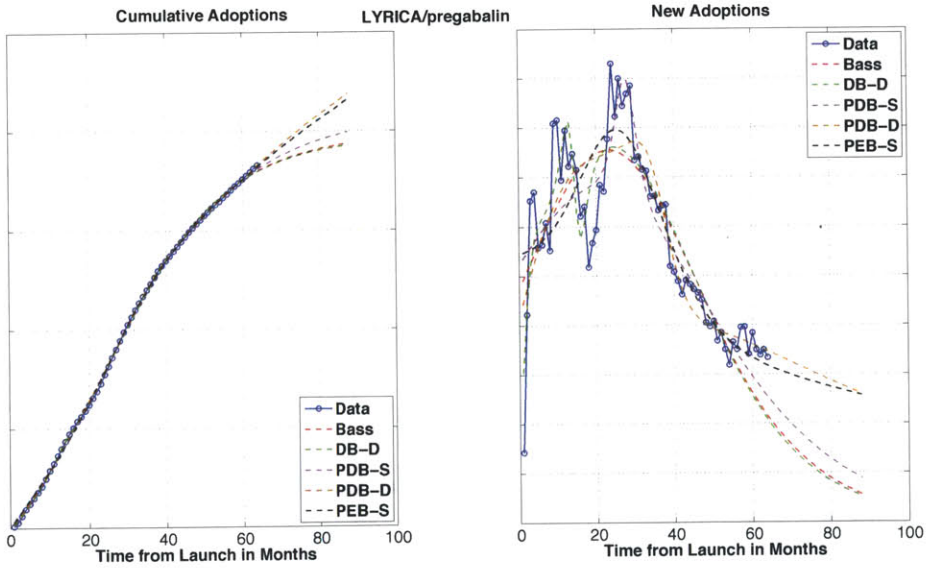

Figure 40. Lyrica® (pregabalin) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.12 Nevanac® (nepafanac)

Nevanac® (nepafanac) was approved August 19, 2005 and became commercially available in the United States in September 2005. It was approved for the treatment of pain and inflammation associated with cataract surgery and is the first-in-class ophthalmic non-steroidal anti-inflammatory drug.

All the models I fit had trouble with this therapeutic because the market appears to be continually strongly growing whereas all these models are expecting saturation effects to begin to takeover. This therapeutic is perhaps better modeled with a dynamic adopter population that accounts for age dynamics and the trends in cataract surgery.

All parameters of the single market Bass model are significant at the 0.05 level.

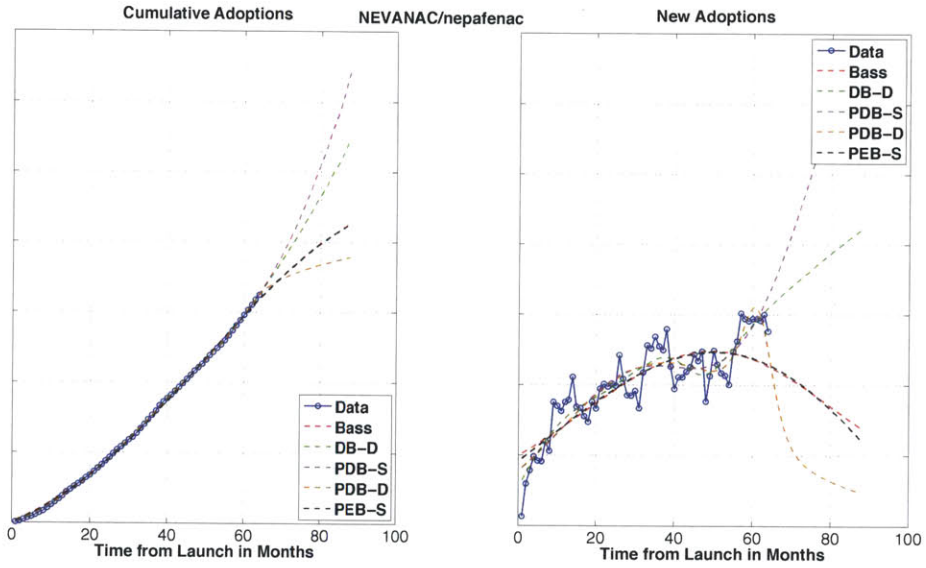

Figure 41. Nevanac® (nepafanac) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit.

14.13 Omacor®/Lovaza® (omega-3 acid ethyl ester)

Omacor® (omega-3 acid ethyl ester) was approved November 10, 2004 and became commercially available in the United States after October 5, 2005. It was approved as an adjunct to diet to reduce very high triglyceride levels (greater than or equal to 500 mg/dL) in adult patients. As of October 22, 2007, at the FDA's request, the therapeutic's name was changed from Omacor® to Lovaza® as a result of reports of prescribing errors between Omacor® and Amicar®.

Visually, it is difficult to distinguish whether a true dual market exists or a single market Bass model adequately explains the adoption pattern. All parameters of the single market Bass model are significant at the 0.05 level.

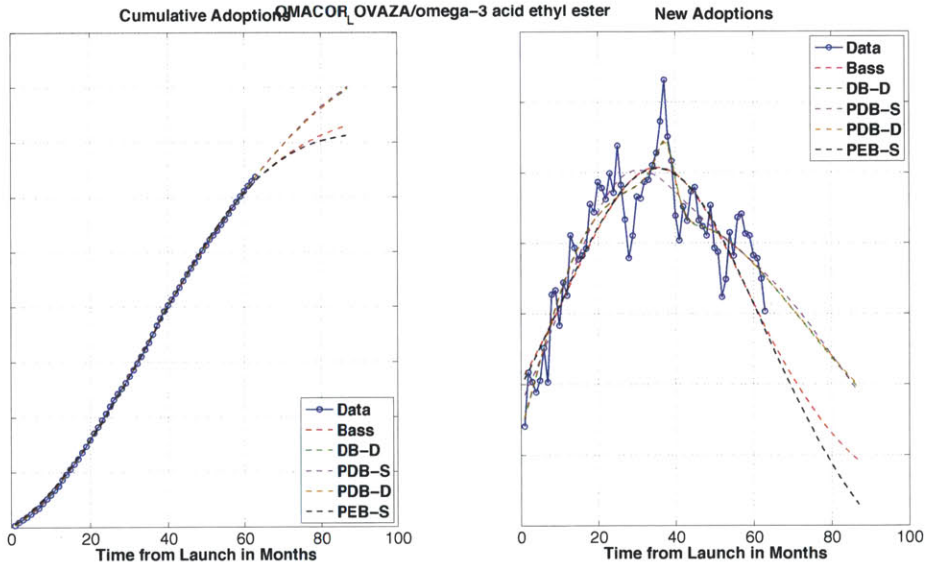

Figure 42. Omacor® /Lovaza® (omega-3 acid ethyl ester) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.14 Ranexa® (ranolazine)

Ranexa® (ranolazine) was approved January 27, 2006 and became commercially available in the United States after March 24, 2006. It was approved for the reduction of chest pain (i.e., angina) in patients who have failed to respond adequately to older anti-angina drugs and was a first-in-class therapeutic. As of November 5, 2008, an extended release version of Ranexa® was approved and it also received a new indication as a first-line treatment for chronic angina.

Visually, it appears a dual market model is superior to a single market model that may line up well with the additional indication Ranexa® received. Additionally, chronic angina is a disease with a spectrum in severity, which is one theoretical construct hypothesized to explain the dual market phenomena. All parameters of the single market Bass model and pooled static dual Bass market model were significant at the 0.05 level.

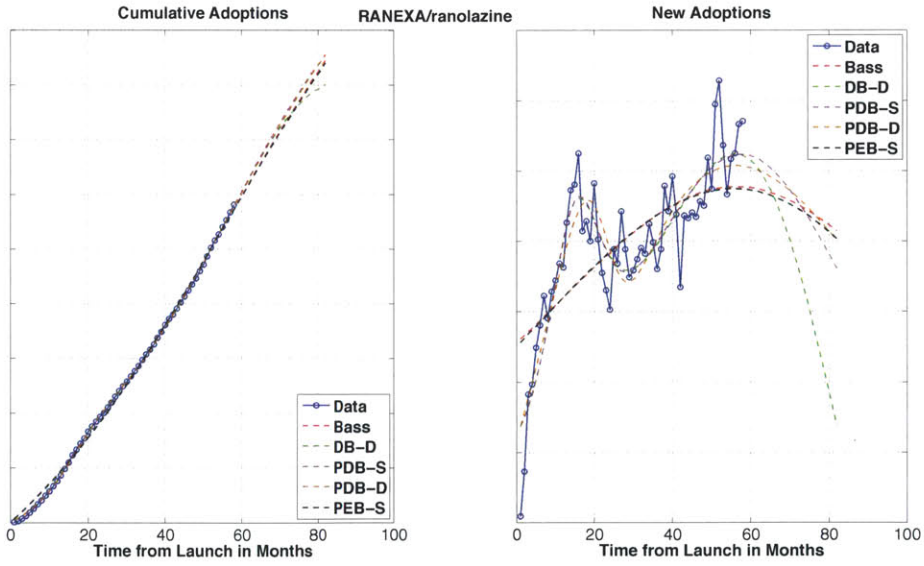

Figure 43. Ranexa® (ranolazine) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.15 Rozerem® (ramelteon)

Rozerem® (ramelteon) was approved July 22, 2005 and became commercially available in the United States in September 2005. It was approved as a non-scheduled treatment for insomnia and was a first-in-class therapeutic. The market leader for insomnia treatments became available generically around the same time Rozerem® was approved.

Visually, there does not seem to be clear evidence of a dual market, and there are not supplemental indications for this product. All parameters of the single market Bass model and pooled static exponential-Bass dual market model were significant at the 0.05 level.

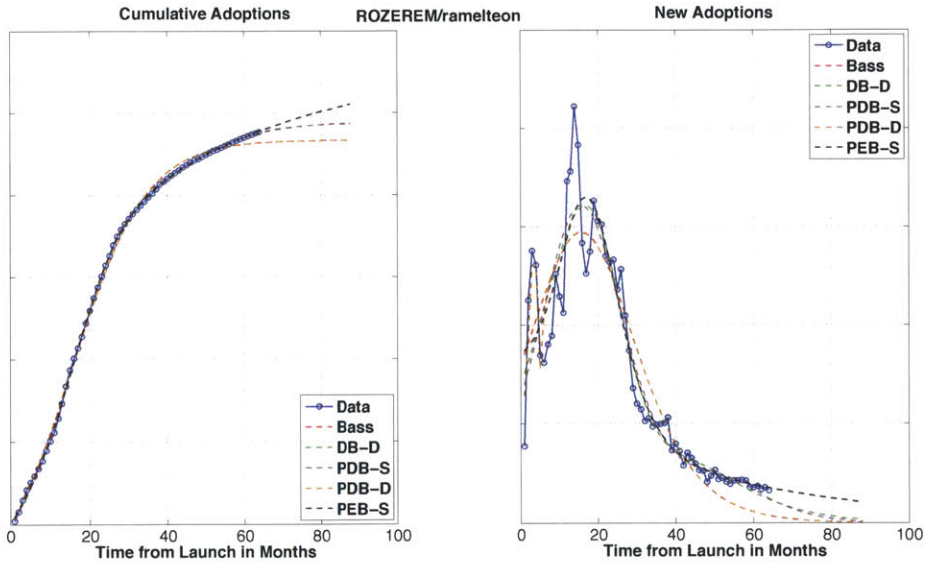

Figure 44. Rozerem® (ramelteon) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.16 Sanctura® (trospium)

Sanctura® (trospium) was approved May 28, 2004 and became commercially available in the United States after August 23, 2004. Sanctura® was approved for the treatment of overactive bladder and was the fourth-in-class. Two competitor drugs in class were approved later in 2004: Vesicare® (solifenacin) and Enablex® (darifenacin). As of August 7, 2007, an extended release version of Sanctura® was approved.

Despite a singular indication, there are clearly multiple markets. All parameters of the pooled static dual market model, the pooled dynamic dual market model, and the non-pooled dynamic dual market model were significant at the 0.05 level.

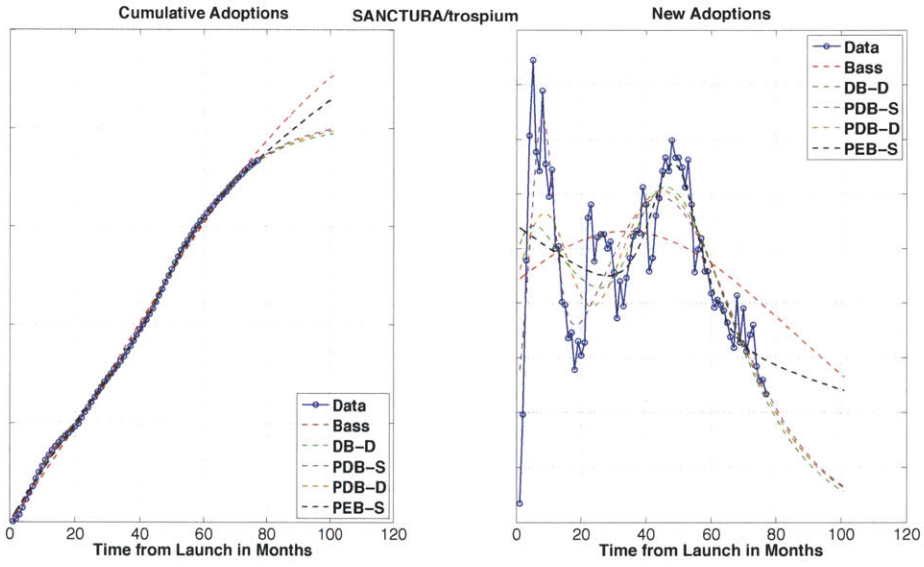

Figure 45. Sanctura® (trospium) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.17 Sensipar® (cinacalcet)

Sensipar® (cinacalcet) was approved March 8, 2004 and became commercially available in the United States in April 2004. It was approved as a first-in-class oral calcimimetic, and was indicated for the treatment of secondary hyperparathyroidism in chronic kidney disease patients on dialysis, and the treatment of elevated calcium levels in patients with parathyroid carcinoma. It was approved with an orphan drug designation.

Sensipar® appears to have at least two markets which were best identified by the pooled static dual Bass market model and the non-pooled dynamic dual Bass market model. All parameters of the single market Bass model were significant at the 0.05 level.

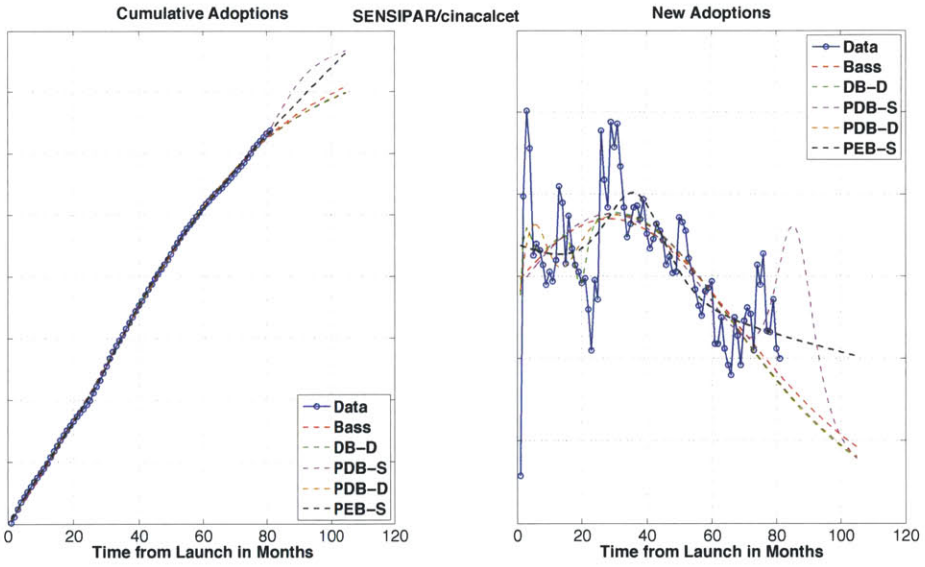

Figure 46. Sensipar® (cinacalcet) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.18 Spiriva® (tiotropium oral inhalation)

Spiriva® (tiotropium oral inhalation) was approved June 30, 2004 and became commercially available in the United States after May 25, 2004. It was approved for the treatment of chronic obstructive pulmonary disease. In October 2006, a competitor was approved for the maintenance therapy of chronic obstructive pulmonary disease, Brovana® (arformoterol inhalation). Another competitor was approved in March 2009.

Despite a single indication, there are clearly two markets for this product. All parameters of the pooled dynamic dual Bass market model, the pooled static exponential-Bass model, and the single market Bass model were significant at the 0.05 level.

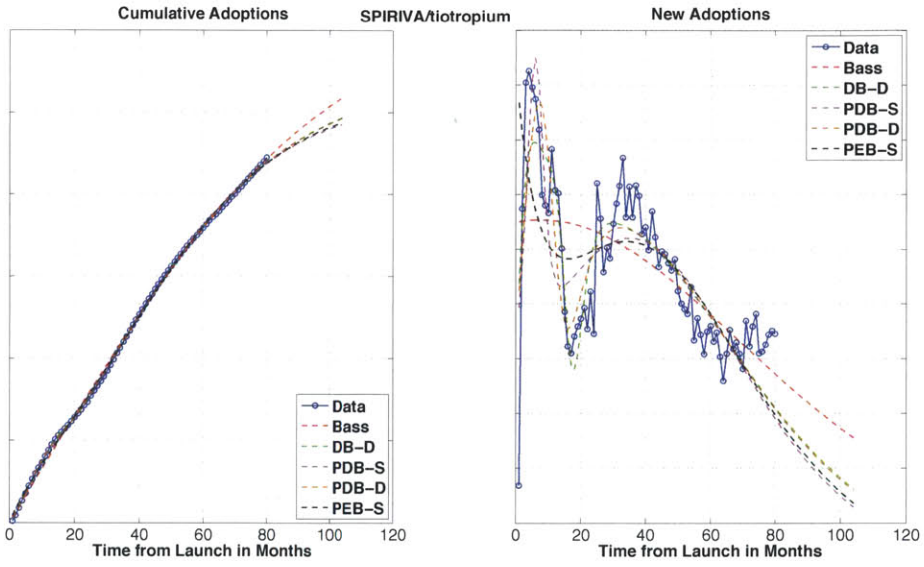

Figure 47. Spiriva® (tiotropium oral inhalation) Adoption Patterns and Nonlinear Regression Model Fits

Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.19 Tarceva® (erlotinib)

Tarceva® (erlotinib) was approved November 18, 2004 and became commercially available in the United States later that month. It was approved for the treatment of non-small cell lung cancer for patients that have already failed one treatment. In May 2005, a competitor drug, Iressa® (gefitinib) was voluntarily withdrawn from the market. On November 3, 2005, Tarceva® received an a supplemental indication for first-line treatment of pancreatic cancer.

Visually, only the non-pooled dynamic dual Bass model seemed to identify a small second market. It is unclear why a dual market model would be superior to the single market model in this case. All parameters of the single market Bass model were significant at the 0.05 level.

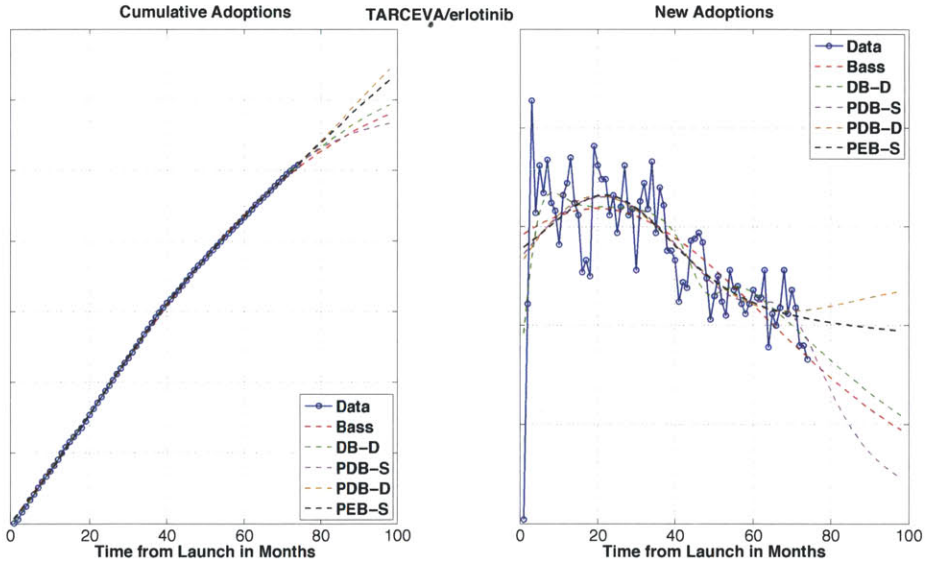

Figure 48. Tarceva® (erlotinib) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.20 Tindamax® (tinidazole)

Tindamax® (tinidazole) was approved May 17, 2004 and became commercially available in the United States in July 2004. It was approved for the treatment of trichomoniasis, a sexually-transmitted disease. On May 27, 2007, Tindamax® received a supplemental approval for the treatment of bacterial vaginosis.

In terms of model fit, all the models are very similar. There is not strong evidence to suggest the superiority of a dual market model over a single market model despite the two indications. All parameters of the single market Bass model were significant at the 0.05 level.

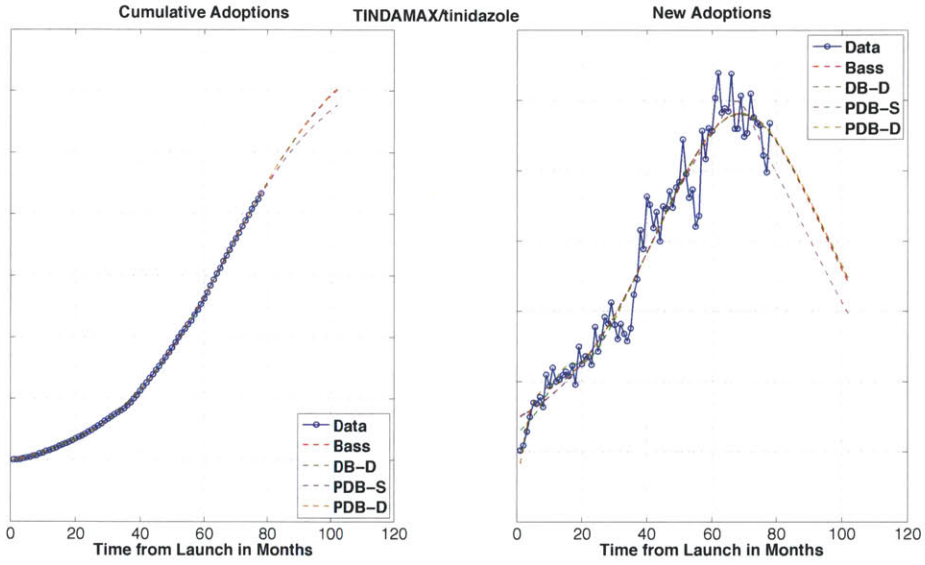

Figure 49. Tindamax® (tinidazole) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit.

14.21 Vesicare® (solifenacin)

Vesicare® (solifenacin) was approved November 19, 2004 and became commercially available in the United States after January 21, 2005. It was approved for the treatment of overactive bladder and was the fifth-in-class. It was preceded by the approval of the fourth-in-class, Sanctura® (trospium), and followed shortly by approval of the sixth-in-class, Enablex® (darifenacin).

With regard to model fits, both the single market model and the dual market models are similar. Compared to the patterns of the two other overactive bladder drugs that both suggest a dual market phenomena, this therapeutic is unusual. Notably, it also emerged as the market leader among the three. All parameters of the single market Bass model and the pooled static dual Bass market model were significant at the 0.05 level.

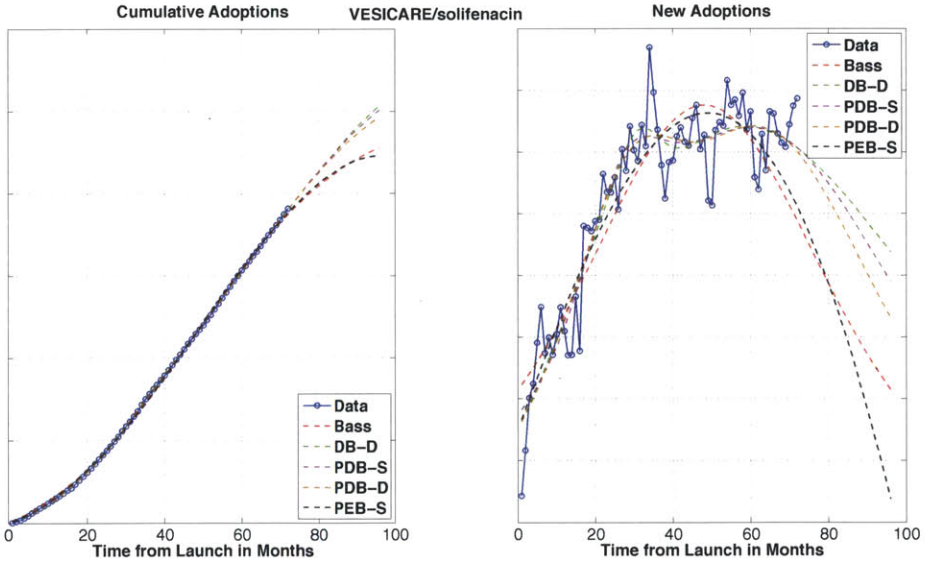

Figure 50. Vesicare® (solifenacin) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

14.22 Xifaxan® (rifaximin)

Xifaxan® (rifaximin) was approved May 25, 2004 and became commercially available in the United States in July 2005. It was approved as a treatment for traveller's diarrhea for patients 12 and older. On March 3, 2010, the FDA approved a supplemental indication for Xifaxan® for the reduction in risk of overt hepatic encephalopathy recurrence in patients aged 18 years or older. The dosage for the two indications is quite different.

The presence of two markets is pronounced and explainable due to the indication expansion. All the dual Bass models tend to perform similarly. The pooled exponential-Bass model with a static probability reduced to the single market Bass model. All parameters of the single market Bass model, the pooled dynamic dual market Bass model, the static pooled dual market Bass model, and the dynamic non-pooled dual market Bass model were significant at the 0.05 level.

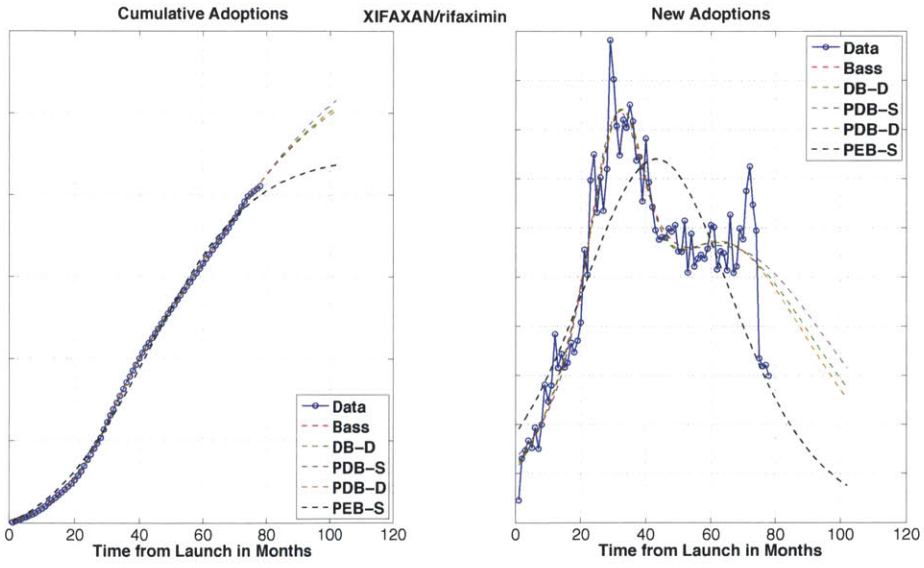

Figure 51. Xifaxan® (rifaximin) Adoption Patterns and Nonlinear Regression Model Fits
Abbreviations: Bass, Single Bass Model Fit; DB-D, Non-pooled Dual Bass–Dynamic Probability Fit; PDB-S, Pooled Dual Bass-Static Probability Fit; PDB-D, Pooled Dual Bass-Dynamic Probability Fit; PEB-S; Pooled Exponential-Bass-Static Probability Fit

THIS PAGE INTENTIONALLY LEFT BLANK

15 APPENDIX D – Glossary of Terms

This is a glossary of terms to aid the reader. All these terms are explained in the footnotes when they are referenced in the text and are repeated here. Quotations indicate that the definition is from another source, which is cited in the text. Italics indicate that the definition is a legal one, derived from the 2007 Food and Drug Administration Amendments Act.

Adjudication - Procedures that are performed to validate the data, i.e. to ensure that the electronic record actually reflects patient experiences. It often involves medical chart abstraction and confirmation of the exposures, outcomes, and covariates of interest.

Adverse drug experience – "Any adverse event associated with the use of a drug in humans, whether or not considered drug related, including—(A) an adverse event occurring in the course of the use of the drug in professional practice; (B) an adverse event occurring from an overdose of the drug, whether accidental or intentional;(C) an adverse event occurring from abuse of the drug; (D) an adverse event occurring from withdrawal of the drug; and (E) any failure of expected pharmacological action of the drug."

Association – "A statistically significant inference regarding a population."

Comparison window – In self-controlled designs, a time period when a person contributes "unexposed" time to surveillance.

Effect size - The quantitative strength of an association, which is usually a point estimate of the effect.

Epidemiologic design – The way the population of interest and the comparison population are sampled for statistical inference.

Exposure - "In epidemiology, it is customary to refer to potential causal characteristics as exposures. Thus, exposure can refer to a behavior (e.g., needle sharing), a treatment or other intervention (e.g., an educational program about hazards of needle sharing), a trait (e.g., a genotype), an exposure in the ordinary sense (e.g., an injection of contaminated blood), or even a disease (e.g., diabetes as the cause of death)."

Exposure-outcome pair - A hypothesized relationship between the exposure and outcome of interest, e.g., oral anti-diabetic medications and acute myocardial infarctions.

Induction period /latency period– The time period after a person has been exposed to a medical product but before the person is "at risk" for experiencing the outcome of interest.

Meta-analysis - "The statistical analysis of a collection of analytic results for the purpose of integrating the findings."

Near real-time data - Data on clinical experiences that arrive with a variable delay from when the experience occurred. There are two sources of delay. First, there is a processing delay, which is the time that elapses between when the experience occurs, and when it is recorded and available for analysis. Second, there is a refresh delay,

which is associated with the frequency with which an originating data source renews their dataset and makes it available for analysis.

New safety information - "Information derived from a clinical trial, an adverse event report, a postapproval study (including a study under section 505(o)(3)), or peer-reviewed biomedical literature; data derived from the postmarket risk identification and analysis system under section 505(k); or other scientific data deemed appropriate by the Secretary about— (A) a serious risk or an unexpected serious risk associated with use of the drug that the Secretary has become aware of (that may be based on a new analysis of existing information) since the drug was approved, since the risk evaluation and mitigation strategy was required, or since the last assessment of the approved risk evaluation and mitigation strategy for the drug; or (B) the effectiveness of the approved risk evaluation and mitigation strategy for the drug obtained since the last assessment of such strategy."

Outcomes - Health outcomes of interest.

Positive predictive value - The number of true positive cases/(true positive cases + false positive cases).

Postmarket - The period after licensure of a product by the U.S. Food and Drug Administration.

Postmarketing requirement - A mandate from the U.S. Food and Drug Administration to the manufacturer/sponsor of a particular product to perform a study.

Precision - The inverse of the variance of the measurements or estimates that a statistical process produces.

Processing delay time – Also known as the claims lag time, is the time that elapses between when an exposure or outcome occurs, and when it is recorded and available for analysis.

Rare event rate – Events that occur with a frequency greater than 1 event per 10,000 person-years , but less than 1 event per 1,000 person years.

Refresh delay time - The frequency with which a participating data partner renews their dataset and makes it available for analysis.

Risk window – The time period when a person is "at risk" of experiencing an outcome of interest following some exposure of interest.

Sensitivity - The number of true positive cases/(true positive cases + false negative cases).

Serious adverse drug experience - "An adverse drug experience that (A) results in—(i) death; (ii) an adverse drug experience that places the patient at immediate risk of death[...];(iii) inpatient hospitalization or prolongation of existing hospitalization; (iv) a persistent or significant incapacity or substantial disruption of the ability to conduct normal life functions; or (v) a congenital anomaly or birth defect; or (B) based on appropriate medical judgment, may jeopardize the patient and may require a medical or surgical intervention to prevent an outcome described under subparagraph (A)."

Serious risk – "The risk of a serious adverse drug experience."

Signal - 1) "Information that arises from one or multiple sources (including observations and experiments) which suggests a new potentially causal association, or a new aspect of a known association, between an intervention and an event or set of related events, either adverse or beneficial, that is judged to be of sufficient likelihood to justify verificatory action." 2) "Reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously. Usually more than a single report is required to generate a signal, depending upon the seriousness of the event and the quality of the information."

Signal evaluation - "Consists of the implementation of a full epidemiological analysis to more thoroughly evaluate the causal relationship between exposure to the medical product and the adverse outcome of interest."

Signal generation/detection – "An approach that uses statistical methods to identify medical product–adverse outcome associations that may be safety signals; no particular medical product exposure or adverse outcome is pre-specified."

Signal of a serious risk – "Information related to a serious adverse drug experience associated with use of a drug."

Signal refinement - "A process by which an identified potential safety signal is further investigated to determine whether evidence exists to support a relationship between the medical product exposure and the outcome."

Specificity - The number of true negative cases/(false positive cases + true negative cases).

Tracked safety issue – An operational term that formalizes the evaluation of a medical product safety signal because it has the potential to lead to regulatory action.

Transportability – An alternative term for what is commonly referred to as external validity or generalizability.

Type I error – The false positive rate or the incorrect rejection of the null hypothesis when it should have failed to be rejected.

Type II error – The false negative rate or the failure to reject the null hypothesis when it should have been rejected.

Unexpected Serious Risk – "A serious adverse drug experience that is not listed in the labeling of a drug, or that may be symptomatically and pathophysiologically related to an adverse drug experience identified in the labeling, but differs from such adverse drug experience because of greater severity, specificity, or prevalence."

Very Rare event rate - Events that occur with a frequency greater than 1 event per 100,000 person-years, but less than 1 event per 10,000 person years.

Washout period – In self-controlled designs, a time period when a person contributes no information to surveillance as they are considered neither exposed nor unexposed.