When Drugs Kill: The Social Structure of Evidence Production

Mathijs de Vaan

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2015

ABSTRACT

When Drugs Kill: The Social Structure of Evidence Production

Mathijs de Vaan

An Adverse Drug Reaction (ADR) is defined by the World Health Organization as "a noxious response to a medication that is unintended at doses usually administered for diagnosis, prophylaxis, or treatment." Estimates suggest that such episodes – in which prescription drugs cause negative health consequences – account for more than 2 million hospitalizations and more than 100,000 deaths per year in the United States alone, making ADRs one of the leading causes of death. To put these numbers into perspective: death from treatment with prescription drugs is about 10 times as common as death from suicide. This dissertation aims to understand why these numbers are so high.

Prior work has focused mainly on the politics of drug approval to show that factors such as deadlines, status of pharmaceutical firms, and foreign approval can account for variation in regulatory decision making by the Food and Drug Administration. I take another route and focus on the production of evidence about the safety of prescription drugs. The way in which medical scientists have typically used evidence is by extracting meaning through aggregation or classification of pieces of evidence. The argument that I am making in this dissertation is that rather than aggregating or classifying evidence, one needs to account for the relationships between pieces of evidence. In particular, the dissertation shows how social theories about the structures of evidence production can be used to better understand the harm that drugs can do and, as a result, allow us to identify unsafe drugs more rapidly.

The dissertation presents analyses based on data from the two main sources of evidence that the Food and Drug Administration has at its disposal to identify unsafe drugs. The first is the Adverse Event Reporting System (AERS).

AERS is an FDA maintained system through which patients and physicians can voluntarily report ADRs to the FDA. The FDA uses this system by monitoring disproportional increases in the number of ADRs reported for a given drug. The second source of evidence is the scientific literature about prescription drugs. The FDA uses this literature to inform regulatory action.

The first set of findings in this dissertation demonstrate that ADR reports for a specific drug are more likely to be submitted if a drug has been publicly scrutinized or when a drug treats the same health condition as a drug that was publicly scrutinized. Patients and physicians differ in the ways in which their reporting behavior changes in response to increased scrutiny. Preliminary findings suggest that these episodes of changes in reporting behavior are associated with delays in regulatory action compared to drugs in which reporting behavior did not change. These findings are consistent with the hypothesis that the detection of signals in massive yet sparse data benefits from social theories of evidence production.

The second set of findings show that the social structure in which scientific evidence about the safety and efficacy of prescription drugs is not uniformly cumulative. In particular, in some cases the scientific debate about the safety and efficacy of prescription drugs is characterized by a disconnect between the claims made before a drug is approved for marketing and the claims made after approval. Moreover, the results from the study demonstrate that debates characterized by a strong disconnect are more likely to be the target of regulatory action. This suggests that a discontinuity in scientific closure is consistent with the idea that the quality of pre-approval scientific evidence predicts post-approval regulatory action.

In sum, this dissertation identifies salient structures in collective production processes and it demonstrates that the structure of collective production reveals meaning that could reduce ambiguity in interpretation.

Contents

List of Tables **v**

List of Figures **vi**

1 Introducing Evidence Production in the Context of Drug Safety **1**

1.1 The Social Organization of Drug Safety 2

1.2 The Salience of Understanding Post-approval Evidence Production 9

1.3 Post-Approval Evidence Production as a Complex Social System 13

2 Structure as the Guiding Principle in Evidence Production **19**

2.1 The Emergence of Regulation and Jurisdiction 20

2.2 A Discontinuous Shift in the Boundaries of Expertise 26

2.3 The Thalidomide Episode, the Kefauver-Harris Amendments, and the Birth of Modern Drug Regulation 35

2.4 The Science of Post-Approval Regulation 41

2.5 Concluding Remarks 45

3 Cognitive Structures and the Effect of Events on the Reporting of Adverse Drug Reactions **47**

3.1 Introduction 48

3.1.1 Post-Approval Learning About Safety and Regulating Prescription Drugs . . . 52
3.1.2 Sense Making, Relational Structures, and Heterogeneity within Audiences . . . 56
3.1.3 Roadmap . . . 67
3.2 Data and methods . . . 70
3.2.1 Study population . . . 70
3.2.1.1 AERS . . . 70
3.2.1.2 NDAs . . . 74
3.2.1.3 ATC . . . 75
3.2.1.4 Drug Withdrawals . . . 77
3.2.2 Empirical Strategy . . . 77
3.2.3 Descriptive Statistics . . . 81
3.3 Results . . . 84
3.4 Discussion and conclusion . . . 90
Appendices . . . 96

4 Scientific Evidence Production and Regulatory Action 98
4.1 Introduction . . . 99
4.1.1 Industry Science, Academic Science, and the Integration Discourses . . . 102
4.2 Roadmap . . . 107
4.3 Empirical Strategy . . . 109
4.4 Data Collection . . . 115
4.4.1 Study Population . . . 115
4.4.2 Variable Descriptions . . . 117
4.4.2.1 Logistic Regressions . . . 117
4.4.2.2 ERGMs . . . 118

4.4.2.3 Negative Binomial and Cox Proportional Hazard Regressions . . . 123
4.5 Results . . . 124
4.6 Discussion and Conclusion . . . 133
Appendices . . . 138

5 Conclusion – Structure in the Production of Evidence **140**
5.1 Synthesis and Discussion . . . 140
5.2 Conclusion . . . 144

Bibliography **146**

List of Tables

2.1 The evidence hierarchy 44

3.1 Descriptive Statistics for AERS 73
3.2 Levels in ATC 75
3.3 Event - report interval per day 83
3.4 Reports per day 83
3.5 Point estimate for treatment effect on focal drug using Poisson regression 84
3.6 RD Estimates 87
3.7 Point estimate for treatment effect on related drugs using Poisson regression 88
3.8 RD Estimates 90
A3.9 Subsets in AERS 96

4.1 Predicting the probability that a paper discusses ADRs 126
4.2 Predicting volume of and duration to label changes 134

List of Figures

2.1 Advertisement for therapeutic substance in Druggist's Circular . 21
2.2 The salience of "drug safety", "drug efficacy", and "placebo" in American literature 40

3.1 Temporal Changes in Reporting Ratios of ADRs 62
3.2 The Network of Drugs that Treat the Same Disease 78
3.3 Timeline of drug withdrawals 79
3.4 Number of reports filed to AERS per day 82
3.5 Regression Discontinuity graphs 85
3.6 RD controls graphs 86
3.7 Regression Discontinuity graphs 89
3.8 RD controls graphs 89
3.9 The effect of noise on the speed of regulatory action 91

4.1 Scatterplot of pre-approval ADR papers and change in ADR paper proportion 128
4.2 Boxplot of coefficient estimates 131
4.3 Coefficient estimates of Pre-Pre and Post-Pre nodemix terms . . . 132
A4.1 Kaplan-Meier plots 139

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, Peter Bearman. Peter encouraged me to pursue novel research directions, asked questions that improved my research, and provided feedback when it was most needed. Our conversations on the 7th floor of Knox Hall were inspiring and I always came out of them with novel insights. I look forward to continuing to work together in the coming years.

I have also benefited tremendously from the invaluable comments, advice, and ideas of David Stark, Tom DiPrete, Bruce Kogut, and Damon Phillips. David and Tom have both been co-authors and working on our papers together have helped me become a better researcher.

Graduate school is very much a social endeavor and it would not have been as much fun without my fellow PhD students Philipp Brandt, Joscha Legewie, Claudius Hildebrand, Ivana Katic, Christina Ciocca, Moran Levy, Luciana De Souza Leao, Kinga Makovi, Joan Robinson, and Pilar Opazo. Thank you for making graduate school such a good experience.

I am very grateful to my sister, Eva, and my parents Math and Nelly. In addition to nurturing my intellectual curiosity, they are always there to support me. Even though we are separated by thousands of miles, the feeling of being able to count on you is amazing.

Finally, I'd like to thank Stéphanie and Alix. Steef, thank you for your love and support, for always reminding me of the important things in life, and for being the most amazing friend. I am very lucky to have you in my life and I am so proud that we can call this adventure ours. And Alix, you're to most amazing addition to my life. I will forever cherish the time I spent with you in between all those hours of dissertation work. Changing your diapers on our cheap changing table, running with you in Central Park, feeding you in the middle of the night, and discovering the streets of New York City are my most precious memories. Steef and Alix, I love you.

DEDICATION

To my wife Stephanie, for her endless support of my seemingly endless academic pursuits.

Chapter 1

Introducing Evidence Production in the Context of Drug Safety

Advancements in medical science have resulted in a wide range of prescription drugs that improve health, reduce pain, and save lives. However, recent years have also been characterized by a large volume of adverse drug reactions (ADRs) including life-threatening injuries, hospitalizations, and deaths. A meta-analysis published in 1998 based on studies conducted in the U.S. concluded that among hospitalized patients the overall incidence of serious ADRs and fatal ADRs was 6.7% (95% CI, 5.2%-8.2%) and 0.32% (95% CI, 0.23%-0.41%) respectively. These percentages imply that ADRs resulted in 2,216,000 seriously injured patients and 106,000 deaths *per year* (Lazarou and Pomeranz 1998). To put these numbers into perspective, in the U.S. the number of hospitalized patients that are killed by unsafe drugs is more than twice as large as the total number of suicides per year. The number of deaths and injured patients is even

more stunning if one considers that the study by Lazarou and Pomeranz (1998) only includes hospitalized patients. More recent studies confirm the results presented in the 1998 meta-analysis (Pirmohamed et al. 2004; Krahenbuhl-Melcher et al. 2012).[1]

1.1 The Social Organization of Drug Safety

To understand the mechanisms responsible for generating the dazzling numbers of deaths and injured patients, researchers have identified a battery of characteristics of (the sponsors of) drugs that have been disproportionally responsible for generating safety problems. These studies have predominantly focused on variation in the organizational, social, or regulatory context in which a drug is embedded prior to approval by the Food and Drug Administration (FDA). The FDA regulates the approval of prescription drugs by providing sponsors of a new therapeutic substance with a license to test the drug in clinical trials prior to approval, by interpreting the results from these clinical trials, and by approving those substances that have been shown to be safe and effective for use in larger populations of patients.

One area of study in this body of literature is concerned with changes in the regulatory environment. In a series of articles, Daniel Carpenter and colleagues (Carpenter et al. 2012, 2008) study how timing issues in the approval process of prescription drugs affect the quality of regulatory decision making.

[1] Although these numbers put ADRs in the top of the leading causes of death, one may argue that the benefits that prescription drugs provide – i.e. the number of lives that they save – might outweigh the damage that they do. This is certainly the case for many drugs, because they treat patients with life threatening health conditions and/or they account for only a small percentage of all ADRs. However, some drugs have several alternatives or do not treat life-threatening health conditions and account for the a large share of all ADRs. For example, Howard et al. (2007) show that the majority of preventable drug-related admissions to hospitals involved either antiplatelets (prevents blood cloths), diuretics (treats several conditions including high blood pressure, glaucoma, and edema), nonsteroidal anti-inflammatory drugs (painkillers), or anticoagulants (prevents blood cloths). Drugs in these classes either have several alternatives or they do not treat life threatening conditions.

These studies exploit the variation introduced by a new policy – called the Prescription Drug User Fee Act (PDUFA) – that was adopted in 1992. This new policy required the FDA to review a new application for a prescription drug within a limited time frame, effectively creating a deadline. Findings presented in this research demonstrate that reviews pile up at the FDA to right before the deadline. Moreover, regulatory decisions to approve a new prescription drug that are made right before the deadline are characterized by higher rates of post-approval and post-marketing regulatory actions (which are induced by concerns over the safety of a prescription drug) including market withdrawals, severe safety warnings, and safety alerts. Although the authors don't explicitly address the question of what triggers regulatory action, FDA statements in their communications about regulatory actions indicate that they are based on new information regarding the safety or efficacy of a prescription drug that was not taken into account when the prescription drug was approved. However, the very fact that regulatory action was taken implies that this new information is salient and should be used when deciding to start treatment with the drug or when deciding to continue treatment with the drug.

The positive association between approval right before the deadline and the number of safety problems that occur after approval is explained by suggesting that regulatory decisions just before the deadline were not based on the high quality evidence that would potentially have been taken into account by the regulator, had the deadline not rushed them into making a decision (Carpenter et al. 2012, 2008). Although these findings on deadlines and the quality of regulatory decisions are consistent with the plausible explanation that a thorough review allows the regulator to precisely determine the boundaries of efficacy and safety of the prescription drug, questions about several closely related and important issues are not yet answered. For example, it remains unclear why some drugs are rushed before the deadline while others are reviewed much ear-

lier. The mechanism causing the correlation also remains obscured. Do longer review times enable the FDA to study all the evidence that is available to them, do longer review times allow for additional studies to be conducted, or are extensively reviewed drug applications more likely to be approved[2] than drug applications that received less scrutiny, thereby introducing a selection effect?

In a related line of research, firm characteristics are linked to the review time of a new drug application. Olson (1997) analyzes all drugs approved between 1990 and 1992 and finds that firms that are less diversified and more R&D-intensive receive shorter review times for their drugs under review at the FDA. Using a similar type of research design Kim (2012) presents findings that demonstrate that the status of the sponsor of a prescription drug is associated with shorter review times. The argument advanced in these studies is essentially that some characteristic of the firm provides a signal to the regulator about the quality of the application which is then internalized by cutting review times. This idea is consistent with research on status and reputation conducted in other empirical settings that demonstrates that status fosters trust in an exchange partner, thereby reducing transaction costs (Podolny 1993). Both Kim (2012) and Olson (1997) aim to control the quality of the drug and sponsor characteristics that are likely to be correlated with application quality to isolate the effect of status and reputation. However, neither study links drug review times to post-approval regulatory action, making it difficult to reconcile their findings and the findings presented in Carpenter et al. (2008, 2012). If the status and reputation of pharmaceutical firms are indeed good proxies for the quality of the drug application, the induced variation in review time should not account for different levels of regulatory action. However, if status and reputation are poor proxies for application quality and if one assumes that differential treatment of

[2]Data on drug applications that were not approved are not publicly available and to date, no study has been able to effectively create a complete risk set of all drugs that have been reviewed by the FDA.

a drug application by the FDA is limited to shorter review times – and not to more intense scrutiny – then drugs sponsored by high status firms may well be characterized by higher propensities of post-approval regulatory problems.

Variation in foreign experience of a prescription drug has also been argued to account for differences in the likelihood that drugs will be the target of post-approval regulatory action (Olson 2013). Foreign experience is defined here as the lag between the date of first foreign approval (approval in any country outside the U.S.) and the date of approval of the prescription drug in the U.S. The study by Olson (2013) hypothesizes about the effect of the length of foreign experience of a prescription drug before approval on the U.S. market on the post-approval safety problems associated with the drug in the U.S. The findings presented in the study show that longer, post-approval exposure of a wide range of patients to a therapeutic substance in foreign countries reduces the number of post-approval safety problems. In interpreting this finding the author suggests that additional information about drug safety and efficacy obtained from experiences in foreign markets can be used to inform regulatory decision making (e.g. labeling). However, the study by Olson (2013) also demonstrates that drugs first approved in the U.S. are less likely to be associated with high levels of regulatory action, a finding that generates an odd functional form in Olson's model. A first coefficient in her model suggests that as the foreign launch lag diminishes, drugs are characterized by more post-approval safety issues. The second coefficient in her model, however, shows that if the launch lag is zero[3], the likelihood of post-approval safety issues abruptly reduces again. It is difficult to propose a hypothesis that captures the mechanism that could produce this functional form.

In addition to variation in firm characteristics and regulatory contexts, re-

[3]Or negative – neither does the paper of Olson (2013) describe how it handles drugs approved in the U.S. first., nor can it be inferred from the table with descriptive statistics.

search has also focused on variation in behavior by the FDA. Research by Moffitt (2010) scrutinizes the process of evaluating prescription drugs prior to approval by examining the use of FDA advisory committees. These committees are comprised of academic experts and are consulted by the FDA to guide the agency in decision making about the approval of prescription drugs. The distribution of advice given by advisory committees tends to approximate unanimity and the outcome of the voting process is very likely to be adopted by the FDA; a shift from 50% supportive voting to unanimity increases the approval probability from 0.64 to 0.92 (Lavertu and Weimer 2011).

Two interrelated questions are addressed in the research on advisory committees. First, what are the conditions under which the FDA decides to consult with an FDA advisory committee about the approval of a prescription drug? Second, are prescription drugs for which the FDA consults with an advisory committee more likely to be associated with post-approval safety problems? The findings from the analyses conducted by Moffitt (2010) indicate that the FDA tends to consult with advisory committees when there are high levels of uncertainty about the safety and efficacy of a prescription drug. The author interprets this result by suggesting that the FDA uses the advisory committee to gain further insights about those basic properties – safety and efficacy – of the drug under review. Moreover, she argues that by consulting with a group of experts the FDA builds a buffer to mitigate future blame that could potentially arise. Regarding the second research question, the results presented in the paper are somewhat mixed, depending on the dependent variables used in the model. This is not surprising; on the one hand, the argument can be made that drugs for which the FDA consulted with an advisory committee have undergone greater scrutiny and should therefore exhibit fewer post-approval safety problems. On the other hand, one may argue that since drugs associated with higher uncertainty are more likely to be subjected to advisory committee

review, the scrutiny of the advisory committee does not eliminate this uncertainty.

In summary, research that focuses on the social organization of drug safety has essentially studied two outcomes: review speed and volume of safety issues that occur after approval of the prescription drug. To explain the variation in these outcomes, researchers have looked at deadlines, sponsor credentials, prior foreign experience with the drug, and the use of scientific expertise in the form of FDA advisory committees. Reconciling the findings in this literature is not straightforward. For example, in the interpretations of their results, Carpenter and colleagues (Carpenter et al. 2008, 2012) demonstrate that shorter review times (drug applications piled up to right before the deadline) reduce the quality of the review. However, as mentioned earlier, it is unclear which types of drugs are piled up before the deadline. If the FDA is able to infer the quality of applications by observing the credentials of the sponsor of the drug, the FDA may have pushed the high quality applications towards the deadline. If that were the case, the second finding in the studies by Carpenter and colleagues (Carpenter et al. 2008, 2012) should be surprising and theoretically inconsistent. Turning to the studies by Kim (2012) and Olson (1997) does not alleviate this potential contradiction. Although both studies show that certain firm characteristics (mostly those describing some form of reputation or status) are correlated with shorter review times and suggest that this correlation could arise from inferring the quality of the application from firm status, none of those studies show that there are indeed differences in the quality of the evidence that supports a new drug application.

It is also not entirely clear how Olson's (2013) argument about the effect of foreign experience fits with other work on the social organization of drug safety. Even if one were to disregard the odd functional form presented in Olson's model, it remains unclear why some drugs are approved earlier in other

countries. Comparative studies on the approval of prescription drugs suggest that the FDA approves drugs fast compared to regulators in other countries (Kessler et al. 1996; Schweitzer et al. 1996). Therefore, one starts to wonder whether some drugs are submitted to different national regulators at different times (Olson (1997) does not account for variation in submission dates) and whether some firms devise variation in submission dates as a strategy to maximize their profits. For example, a potential hypothesis could be that if the approval success rate varies with foreign experience and if foreign approval benefits domestic approval, firms may exploit their experiences with different national regulators to develop a global submission plan.

Finally, the research on advisory committees does not provide evidence on whether consultation with advisory committees benefits the decision making process at the FDA. And neither does it show a difference in the safety problems associated with drugs for which an advisory committee was consulted to decide on its approval. However, it may still be the case that advisory committee consultation has had a beneficial effect in terms of decreasing the number of patients harmed by an unsafe drug. First, advisory committees may have kept the most harmful drugs of off the market, but since rejected applications remain unobserved, it is impossible to answer this question. And second, if the drugs discussed in advisory committees are indeed the ones that have the most risk and uncertainty associated with them, FDA advisory committees may have shaped the conditions under which their harm is minimized. The main problems in testing this hypothesis is that given the data constraints, establishing a counterfactual is very difficult.

1.2 The Salience of Understanding Post-approval Evidence Production

To understand why there is significant variation in the likelihood that prescription drugs will experience post-approval regulatory action, extant research has identified salient variation – albeit contradicting at times – in a range of conditions in the pre-approval stages of a prescription drug's life-cycle. While this research has provided insights into the question of *if* safety problems will occur, it has not been able to answer the question of *when* safety problems will be identified. In other words, while we now know some of the risk factors predicting safety problems, we do not know what accounts for variation in when the regulator learns about a safety problem associated with a prescription drug. This dissertation is a first step in answering this question.

Understanding *when* safety problems are identified – rather than *if* they are identified – is an important public health question. First, understanding why safety issues are detected later rather than sooner may save lives. Recent drug disasters such as those of Vioxx and Avandia provide clear examples of why early detection is important. Vioxx had been on the U.S. market for about five years before the FDA decided to withdraw approval for the drug. During those years on the market Vioxx is claimed to have caused 27,785 acute myocardial infarctions and sudden cardiac deaths.[4] Avandia is another case of a prescription drug that killed many people. Avandia was a blockbuster drug treating diabetes and was approved for marketing by the FDA in 1999. After being on the market for 7 years, the drug was shown to cause cardiovascular disease and the drug was withdrawn from the EU market and was put under significant restrictions in the U.S. FDA researchers have found that in its 7 years on

[4] This estimate is based on a relative risk of 1.5 to 3.7 – depending on the dosing. More information can be found in a document by David Graham, M.D., M.P.H. (1.usa.gov/1CbWQvh) who works for the FDA.

the market, Avandia has caused a total of 131,000 acute myocardial infarction, strokes, and deaths in the U.S. alone.[5] Thus, since the number of casualties can rapidly increase within a limited amount of time, early detection of safety problems may save lives.

A second reason why studying the "when" question is important is that doing so forces the researcher to take seriously the evidence that is produced about the safety of prescription drugs *after* a drug is approved by the FDA. Unless the regulator decides to revisit and reinterpret evidence that was already available prior to approval (which virtually never happens), regulatory action is based on evidence produced *after* approval of a prescription drug.

This second argument is especially salient because the sample population on which a drug is tested in pre-approval clinical trials and the population to which the drug is prescribed differ greatly (Psaty and Burke 2006). Prior to approval by the FDA, prescription drugs are tested in several stages of clinical trials and while some large trials have several thousands of human subjects in them, the population to which a drug is prescribed post-approval is typically much larger and so is the diversity of human bodies exposed to the prescription drug. Moreover, dosing and indication are no longer controlled once the drug hits the market and are therefore likely to introduce additional variation in treatment. This additional variation introduced once a drug is released on the market increases the probability that previously unknown pharmacodynamic and pharmacokinetic[6] processes will reveal itself during treatment with a prescription drug. Therefore, much of the evidence that will determine the safety profile of a prescription drug can only be constructed *after* the substance is ap-

[5]This estimate is based on a relative risk of 1.4 for cardiovascular events obtained from a Rosiglitazone (Avandia) meta-analysis and the DREAM trail. More information can be found in the slides of David Graham, M.D., M.P.H. (http://1.usa.gov/1c3fYU3), presented at the "Endocrinologic and Metabolic Drugs Advisory Committee and the Drug Safety and Risk Management Advisory Committee Meeting" on July 30, 2007.

[6]The term pharmacodynamics refers to the effect of a drug on the body while the term pharmacokinetics refers to the effect that the body has on the drug.

proved. In that sense, prior research on variation in drug safety adheres to the somewhat outdated belief that the safety profile of a prescription drug can be established prior to approval of the prescription drug (Psaty and Burke 2006). The study by Olson (2013) is obviously an exception. She acknowledges the idea that market experience of a drug is likely to generate information that has not been revealed in the clinical stages of drug testing. However, she does not hypothesize about the relation of foreign experience of a drug and the temporal pattern that describes when safety problems emerge.

Perhaps the need for serious attention for post-approval evidence about the safety of a prescription drug is most strikingly paraphrased by the authors of an influential study about COX-2 inhibitors (including Vioxx) published just before Vioxx was taken off of the market:

> *What are the implications of the COX-2 inhibitor story in terms of pharmacoepidemiology? It illustrates the fact that new risks will become apparent only when a drug is in widespread use. After a new drug is approved, the dosing and indications in routine use are likely to differ from those in the pivotal studies leading to registration. One question is whether proactive surveillance should begin with the first routine use of a new drug, with the aim of shortening the lead time before potential adverse effects are identified and minimizing the propagation of risk. In an era of blockbuster drugs, do we also need blockbuster pharmacoepidemiology? There is a growing disequilibrium between the efficiency of the drug approval process and that of post-marketing surveillance. Perhaps mandated, proactive, post-marketing surveillance, based on informed assessment of potential risks, should become a routine part of the drug approval process in the future (Ray et al. 2003)*

This quote indicates that besides the lack of scientific progress made to un-

derstand the post-approval evidence production processes, regulatory expertise on this stage of the drug life-cycle is also lacking. This idea is widespread; in 2006, the Government Accountability Office stated in one of its reports that "the FDA lacks clear and effective processes for making decisions about, and providing management oversight of, postmarket safety issues."[7]

Although the aim of this dissertation is not to understand the lack of attention for post-approval regulatory processes, several interrelated reasons are plausible. First, the number of actors in the pre-approval stage of a drug's life-cycle is limited and the level of standardization, regulation, and bureaucratization is high. As a result variance in those regulatory processes can be somewhat cleanly defined (Luke and Stamatakis 2012). Second, since the FDA regulates both the pre- and the post-approval processes, any post-approval regulatory action is essentially signals that prior regulatory behavior was not in the best interest of public health. This is one of the reasons why, in many European countries, pre- and post-marketing regulation is done in separate organizational entities. Third, the support that the FDA has received for proposed improvements on its post-marketing regulatory processes has been very limited. An example of this is a negotiated exception in the 1992 Prescription Drug User Fee Act (PDUFA). The PDUFA did not only include deadlines as a new provision; it also included the provision of user fees contributed by pharmaceutical companies that are supposed to make the FDA a better regulator. However, the PDUFA prohibited the FDA from applying user fees to improve post-approval drug surveillance (Psaty and Burke 2006).

[7] See http://www.gao.gov/products/GAO-06-402

1.3 Post-Approval Evidence Production as a Complex Social System

Besides the additional variation in patient characteristics and treatment, another major change going from the pre- to the post-approval setting is its evidence production process: while procedures, behavior, and conditions that guide evidence production in the pre-approval stage are standardized, controlled (e.g. protocols describe what study designs should look like, Randomized Clinical Trials (RCT) are the de facto method of evidence production, and the behavior of the sponsor of the drug is heavily regulated), and centrally managed by the sponsor of a drug, the post-approval stage is characterized by much less regulation and a much more diverse set of stakeholders who produce the evidence. For example, neither the behavior of patients and physicians nor the professional behavior of medical scientists writing about drug safety is strictly regulated. Moreover, whereas RCTs dominate the pre-approval evidence production process, the post approval stage is characterized by patients, physicians, and medical scientists contributing large amounts of case reports, observational studies, smaller RCTs, comparative RCTs, and adverse event reports.

Building on these differences, the argument that I will advance in this dissertation is that the post-approval evidence production system is a complex social system. Agents are heterogeneous, they interact with one another in the process of producing evidence, and jointly they produce emergent effects (Smelser 2011) that are different from the effects that the individual actors would have produced in isolation (Luke and Stamatakis 2012). As a result, each individual piece of evidence derives its meaning from its embeddedness in a larger collection of claims and in order to extract this meaning one needs to identify the structural patterns that guide evidence production. Such a setting in which heterogeneous agents interact with one another to collectively produce

a body of evidence warrants sociological inquiry that pays particular attention to structures that guide evidence production. Hence, this dissertation studies how patients, physicians, and medical scientists go about in producing the evidence about drug safety – based on which the FDA has to build its regulation. In doing so, this dissertation builds on recent literature that scrutinizes how the structure in which social actors are embedded affects their evaluation of an object, person, phenomenon, or context.

In recent years, an increasing volume of scholarly work in sociology has examined how people make evaluations and produce claims, especially in settings characterized by high levels of uncertainty. This diverse body of research embraces the inherently sociological question of how interactions and observations inform social actors in the process of attributing value. The contexts of these studies range from science (Collins 1998; Shwed and Bearman 2010) to symbolic goods (Lieberson 2000; Salganik et al. 2006) and finance (Beunza and Stark 2004; MacKenzie 2011) and they contribute theoretically to the understanding of diagnosis and social influence (Liu et al. 2010; King and Bearman 2009; Rossman 2014), status and honor (Gould 2002), popularity and fashion cycles (Lieberson 2000; Strang and Macy 2001), classification and categorization (Durkheim 1963; Zuckerman 1999; Hannan 2010), and expertise (Collins 1992; Eyal 2013).

Following this work, the aim of this dissertation is to demonstrate that in order to understand the evidence produced in the post-approval stages of a drug's life-cycle, one needs to identify the patterns that guide processes of evaluation. That is, I argue that evidence production is ultimately an act of evaluation and that extracting meaning from aggregated evaluations requires one to identify the mechanisms by which individuals construct their evaluations. As a result, this dissertation is a study about the sociology of evidence production, not a study about drug safety.

To be explicit about the structures that I am referring to and the way in which I think they can help to understand the meaning of an aggregate pattern, I will briefly introduce two papers, each of which sets a precedent for the two empirical chapters of this dissertation.

The first paper by Salganik, Dodds, and Watts (2006) describes an experiment in which subjects are asked to evaluate songs from unknown bands. The paper presents evidence that demonstrates that exposure to the evaluations of others who have listened to the same songs increases inequality in evaluations. In particular, positive evaluations concentrate in few songs while the majority of options receives a less positive evaluation. The authors argue that this outcome is observed because in conditions of uncertainty others are likely to serve as cues for one's own evaluation. Obviously, the tendency of individuals to align their evaluation with the evaluation of others, thereby reducing the perception of uncertainty associated with subjective judgments has been prominent in the sociological literature on inequality within reward systems since the introduction of the "Matthew Effect" (Merton 1968; Zuckerman and Merton 1971). Essentially, this effect implies that objects, people, organizations, or events of similar quality attract disproportionally positive judgments when its status is high and disproportionally negative judgments when its status is low, regardless of its underlying quality.

Although the literature on how relational cues are used to make evaluations is vast and further expanding, at least two questions remain unanswered. First, do individuals differ in the ways in which they use evaluations contributed by others? And if they do, how and why do they differ? Salganik et al. (2006) do not identify causes of variation in the degree to which individuals use relational cues and they therefore cannot answer the question of how this would translate into different outcomes of the evaluation process. And second, given that many social contexts are characterized by a variety of different relational cues,

how do individuals select the ones that guide them? For example, the findings by Salganik et al. (2006) have mostly been cited by research on the role of social influence in the diffusion of behavior. That body of research argues that individuals are embedded in social networks and that the relations that make up those networks serve as the conduits that enable diffusion. However, as Rossman (2014) recently argued, under some conditions behaviors do not diffuse because its adopters engage in social interactions, but because the behavior is embedded in a legitimized category. He argues that categorical legitimacy can alleviate the problem of uncertainty and could cause diffusion patterns that bear resemblance to the social influence diffusion pattern.

A second study, conducted by Shwed and Bearman (2010) sets a precedent for Chapter 4, which describes the second empirical study of this dissertation. In their paper, Shwed and Bearman (2010) argue that through the identification of structure in citation networks one can extract meaning from a body of evidence. In particular, using network analytic techniques, they show that network modularity in citation networks can reveal the consensus within a scientific debate. Building on Pinch and Bijker (1986) and their notion of closure, they quantitatively follow scientific debates to detect temporal variation in the modularity in the citation networks over time. They link this variation to expert reports on consensus and show that their modularity score detects scientific consensus even before the expert reports claim that there is consensus. Moreover, Shwed and Bearman (2010) use their strategy to identify three trajectories in which science is conducted: 1) spiral, in which substantive questions are answered and revisited at a higher level; 2) cyclical, in which similar questions are revisited without stable closure; and 3) flat, in which there is no real scientific contestation. Using their network modularity score and linking it to these different scientific trajectories should enable outside observers without deep, substantive expertise to assess the status of a scientific debate. The paper pro-

vides an important example of how the identification of a structure allows one to extract meaning from a body of evidence.

The paper also gives rise to a series of questions: why do debates differ in terms of the trajectory in which arguments are made? To my knowledge no research has shown why some scientific debates unfold by answering questions and revisiting them at a higher level, while in other cases similar questions are revisited without scientific closure. Another question that the research of Shwed and Bearman (2010) gives rise to relates to ways in which the consumption of science is affected by the patterns by which scientific debates unfold. For example, some academic disciplines are heavily drawn from to inform policy and regulation. If there is variation in the structure in which debates unfold, one expects that this affects the implementation of regulation and policy.

In Chapter 3 and Chapter 4, I will further build on the two papers briefly discussed above and I will demonstrate how the research presented in this dissertation starts to provide answers to some of the questions that those papers give rise to.

The structure of the dissertation is as follows: The next chapter first describes the emergence of the FDA as an expert regulator and it describes the salient components of the process of evidence production in the context of drug safety. I will describe how the FDA assumed its position as a strong pre-approval regulator and I will then show how the FDA's expertise in guiding evidence production pre-approval does not translate well to the post-approval stages of a drug safety debate. In Chapter 3, I will present an empirical study of the production of evidence by patients and physicians. I do so by analyzing data on reports filed by patients and physicians about ADRs that they have experienced. The temporal variation in the number of reports filed by patients and physicians provides useful information for the regulator to take regulatory action. Yet, this data is also strongly ambiguous. I'll explore how cycles of in-

creased reporting are produced and I'll explain how those cycles are generated through mechanisms other than an increase in the number of ADRs. In Chapter 4, I turn from evidence produced by patients and physicians to evidence produced by medical scientists. I explore how the academic debate about the safety and efficacy of prescription drugs changes over time and I examine how those changes relate to post-approval regulatory action. In Chapter 5, I synthesize the main findings from this dissertation and I discuss them.

Chapter 2

Structure as the Guiding Principle in Evidence Production

Since its founding in the early 1900s, the Food and Drug Administration has become the expert organization in regulating and evaluating evidence production in the pre-approval stage of a drug's life-cycle. The regulation and evaluation of post-approval evidence production, however, has proven to be an elusive quest. This chapter first describes the emergence of the FDA as an expert organization. It then discusses the discrepancies between pre- and post-approval processes of regulating drug safety and it shows that the tools that make the FDA an expert in regulating the drug approval process fall short when regulating drugs after market approval. Finally, this chapter advances the argument that evidence production is a complex social process and that a social theory of evidence production is needed to extract meaning from the evidence produced. In doing so, I propose that by

identifying social and cognitive structures that guide evidence production, ambiguity about the interpretation of evidence reduces.

2.1 The Emergence of Regulation and Jurisdiction

Although therapeutic substances have been regulated since the birth of the republic, early regulation was minimal, incoherent, and scattered across many different regulatory levels including the municipal, state, and federal level. The first real step towards a centralized national regulation came when the Biologics Act of 1902 and the Pure Food and Drugs Act of 1906 were adopted (Carpenter 2010). Provisions in the Biologics Act required manufacturers of vaccines to be licensed and they authorized inspection of manufacturing facilities by the U.S. Department of Agriculture (USDA).[1] The Pure Food and Drugs Act gave the USDA the power to financially penalize manufacturers of therapeutic substances for adulterating or misbranding those substances (Law 2006).

These provisions were deemed necessary given the total lack of control over those who were manufacturing and selling medicines to patients. Druggist's Circular, a commercial publication that targeted pharmaceutical retailers, listed about 50,000 therapeutic substances in 1906 and estimates suggest that the total market for those substance was worth between $75 million and $110 million annually (Adams 1907; Carpenter 2010) which would be between $1.9 billion and $2.8 billion in 2014 dollars[2]. The wide availability of this large number of substances was problematic for two main reasons: First, many of these drugs were toxic and they either made people sicker or they actually killed the patient. Second, the vast majority of these substances did nothing in terms of treating the condition that triggered the use of the substance – meaning that potentially

[1] See 1.usa.gov/1kBtBfr for other significant dates in the history of drug regulation in the U.S.

[2] Although these numbers are very large, they dwarf compared to the size of the U.S. pharmaceutical industry today. In 2014 the U.S. pharmaceutical industry was worth more than $377 billion. See theimsinstitute.org.

available and beneficial treatment was foregone by their consumption (Carpenter 2010). Figure 1 shows one of the advertisements for therapeutic substances as found in the Druggist's Circular. The pills in the second bottle from the left were recommended as a treatment for liver inflammation (and many other conditions). Note that the pills were sugar coated and that they did not come with a "stereotyped form for printing", because "each lot is distinctive in style."

Figure 2.1: Advertisement for therapeutic substance in Druggist's Circular

Note: This advertisement from the April 1892 Druggist's Circular shows some of the pills sold by Warner. Please note that Warner is one of the few manufacturers of therapeutic substances that survived well into the 20^{th} century (Warner-Lambert was acquired by Pfizer in 2000).

The problem of widespread consumption of many unsafe or ineffective drugs on the U.S. market was further accentuated by the fact that virtually none of the stakeholders in the domain of medicine – including physicians, patients, regulators, drug manufacturers, or scientists – could legitimately claim to hold expert knowledge. Virtually no stakeholder group would be granted the status of expert by any other stakeholder group (Parascandola 1992) and each interaction between members from different social groups or professions created conflicts over jurisdiction and expert knowledge (Abbott 1988).

The lack of hierarchy in the expertise distribution and the systematic disagreement over the boundaries of expertise is perhaps most clearly illustrated by the attitude and behavior of many consumers of therapeutic substances. They expressed serious concerns and heavily opposed the 1906 Pure Food and Drugs Act arguing that it was their right to engage in self-medication and autotherapy (Young 2015). Although these votaries of self-medication strongly opposed a government that limited their access to any therapeutic substance, at the same time they were actively lobbying to get more information disclosed on drug labels arguing that information would allow them to make a well informed decision. These patients strongly believed in the layman who could, given access to full information, make a well informed medical decision (Parascandola 1992; Carpenter 2010). Obviously, one of the main results of a general lack of expertise was that it was virtually impossible for patients to be guided by a reliable source to select only those medicines that were not toxic and treated their condition effectively.

Organizationally, the Biologics Act of 1902 and the Pure Food and Drugs Act of 1906 were enforced by the Bureau of Chemistry which was situated within the USDA. The bureau had been active in a wide range of activities within the domain of agriculture, but by the end of the 1920s the agency decided that drug regulation had become so differentiated from agricultural regulation – which

obviously was at the core of the USDAs activities – that a separate agency was established called the Food, Drug, and Insecticide Administration. Three years later, in 1930, the name of the agency was changed again to the Food and Drug Administration (FDA).[3]

Although the new regulation of the early 1900s provided the FDA with some legal power, in practice the two acts did little to redefine and clarify the boundaries of expertise and jurisdiction. The Pure Food and Drugs Act formally gave the USDA/FDA the power to seize fraudulent or falsely advertised substances, but the fines were small and the resources available to the USDA to prosecute those who falsely advertised were minimal. Moreover, both acts contained many loopholes and the majority of manufacturers of therapeutic substances were able to essentially continue operations in the same way as they had done for years. Rexford Tugwell, a former Professor of Economics at Columbia University who went on to become the Assistant Secretary of Agriculture under Roosevelt, explains the power that could be derived from ignorance under the 1906 Pure Food and Drugs Act:

> *The old bill of 1906 (...) had put on the Administration not only the obligation to prove that damage had been done by adulterated (...) or dangerous (...) drugs, but also to show that it had been done with malicious intent. Well, of course, this was impossible to do and it had the absurd effect of making the person who did it practically free of any obligation because all he had to prove was he didn't know he was doing it. The more ignorant he was, the safer he was, and this was a situation that naturally the Food and Drug people found frustrating, because every time they went into court the lawyers on the other side proved to a jury's satisfaction that what was done might have been damaging, but their client, they were sorry to say,*

[3]See John Swann's writing (1.usa.gov/18Z18fz) for a more elaborate discussion of organizational changes within the FDA in the early 1900s.

didn't mean it, didn't even know about it and had no intention of hurting anybody. So they went free.[4]

Meanwhile, officials within the FDA – who recognized their limited jurisdiction – developed several strategies to strengthen its position and to become more effective and credible as a regulator of medicines (Carpenter 2010): First, to get drug regulation on the national agenda at all, the FDA started to ally with key players in the Franklin D. Roosevelt administration. Although Roosevelt himself was not particularly supportive of the FDAs efforts, some officials in the administration certainly were. One of them, Rexford Tugwell, was closely tied to the main directors of the FDA and he was actively involved in drafting bills that would strengthen the 1906 Pure Food and Drugs Act.[5]

Second, the FDA launched a campaign to prevent the public from consuming just any therapeutic medicines citing its lack of proof of safety and efficacy. Although the FDA had been prevented from exercising publicity efforts by means of the Deficiency Appropriations act of 1919, the agency had found ways to reach the American public. Such efforts included the publication of a book titled *American Chamber of Horrors* by one of the FDA officials.

A third channel through which the FDA tried to strengthen its position as a national regulator was through initiating and strengthening ties to women's organizations and consumer unions. These organizations experienced ballooning member counts and their values and concerns showed strong overlap with the values and concerns within the FDA (Carpenter 2010).

While the FDA received support from an important set of stakeholders, there were strong counterforces too. For example, organizations representing the pharmaceutical industry – including the Proprietary Association (PA) and the United Medicine Manufacturers of America (UMMA) – were very active in

[4] 1.usa.gov/1ACVWFC.

[5] See 1.usa.gov/1ACVWFC for an interview with Rexford Tugwell about his close ties to the FDA and his activities to strengthen the regulation of therapeutic substances.

trying to maintain the status quo through advertisements and lobbying. Another important counterforce was the fact that the supporters of the FDA in the Roosevelt administration (mostly prominently Rexford Tugwell and Senator Royal Copeland) were unpopular figures in the regulatory domain. These two counterforces were so strong that several bills that would strengthen the position of the FDA and that were voted on in congress did not get majority support (Jackson 1970; Carpenter 2010).

Although it is not clear how and whether it affected the FDAs attempts to strengthen its position as a regulator, another organization was also actively trying to position itself as one of the main voices in the drug regulation domain: The Council on Pharmacy and Chemistry of the American Medical Association (AMA) was actively engaged in discussions about drug safety and explicitly portrayed itself as an expert organization. The AMA was strongly tied to physicians and manufacturers of therapeutic substances through their journal, the Journal of the American Medical Association (JAMA) and JAMA published articles that were read by physicians nationwide (Carpenter 2010). Moreover, drug retailers heavily used JAMA as an advertising outlet. The dual role of protecting the public interest and commercially publishing a journal created conditions that were difficult to navigate for the AMA. In particular, the AMA walked a fine line between striving for expertise and striving for profitability. In 1930 the AMA made an additional bid for being accepted as a main expert by creating a "Seal of Acceptance" for drugs that were deemed safe and effective according to the standards of the AMA. And while certification by an expert organization could certainly benefit the safety level of drugs available on the U.S. market, the seal also required its recipients to advertise their products in AMA publications.

In sum, the years following the newly introduced drug acts of 1902 and 1906 were characterized by bids for expertise. Patients tried to force manufacturers

to add more information to a label to be able to make well informed decisions, the AMA introduced a seal to certify the quality of a drug, drug manufacturers united in order to maintain the status quo, and the FDA sought allies to generate legitimacy for their claims about drug safety. Yet, the fact that the knowledge system underlying claims of expertise was largely underdeveloped and the fact that the development of scientific standards was nascent prevented an efficient evidence production process from emerging.

2.2 A Discontinuous Shift in the Boundaries of Expertise

The late 1930s represent an important episode in the strife over expertise about drug safety. In September and October of 1937 reports came in at several organizations including the FDA and the AMA about people in Tulsa, Oklahoma who were dying unexpectedly. Many of the reports[6] linked these deaths to treatment with a therapeutic substance called "Dr. Massengill's Elixir Sulfanilamide." Since sulfanilamides were widely used – without problems – for a variety of indications including colds and pneumonia, FDA inspectors suspected the specific substance to have caused patients in Tulsa to die. Laboratory tests confirmed the FDAs hypothesis: it turned out that "Dr. Massengill's Elixir Sulfanilamide" did not only contain sulfanilamide; it also contained the highly toxic substance called diethylene glycol. Massengill's chief chemist mixed up a solution containing 10 percent sulfanilamide, 72 percent diethylene glycol, and 16 percent water. The solution was also flavored with raspberry extract, saccharin, and caramel making it a popular drug (Geiling and Cannon 1938; Wax 1995; Barley 2007). The total number of deaths caused by exposure to the

[6]Since there was no standardized system in place yet to collect reports and complaints about therapeutic substances, these reports came in through traditional means of communications including telephone, telegram, and regular mail.

elixir is at least 76 while most sources argue that the number exceeds 100 (Geiling and Cannon 1938). Compared to more recent episodes of casualties caused by unsafe drugs these absolute numbers are not exceptionally high. However, given the fact that only 633 shipments of the elixir had been distributed across the U.S. and that many of them were still intact, the number is very high.

The tragic sulfanilamide episode exposed some interesting details on how drugs were regulated in the U.S. One of the outstanding features of the episode that was picked up in newspaper reports and other accounts of the tragedy was that the FDA handled the cases swiftly and strongly (Young 2015). The organization started investigations about the causes immediately after the first reports came in and it sent out field agents to track down all batches of the elixir that had been shipped around the country. In the media, these organizational interventions were portrayed as acts of strong and well-informed leadership on behalf of the FDA. Moreover, this image of the FDA was only strengthened when it became clear how large the crisis would have been had the FDA not intervened. Many reports about the episode contained details on the counterfactual outcome which suggested that if the FDA had not been as fast in recovering the elixir, the tragedy would have taken the lives of at least 4,500 Americans (Jackson 1970).

The episode also showed that despite the strong performance of the FDA, there was no legal mandate to act forcefully upon these new findings. According to the 1906 Pure Food and Drugs act, Massengill's only fraudulent action was calling the medicine an "elixir", a designation that was reserved only for compounds containing ethanol (Wax 1995). However, although the FDA was tied in its actions and many people had died, the fact that the episode was picked up by the media had created a strong and lasting support for increased regulation of drugs. Existing ties, such as those to women's organizations, were strengthened and prior competitors reached out to the FDA. For example, Mor-

ris Fishbein (the director of the AMA and editor of JAMA) put in strong words of support following the episode despite the fact that the relationship between the FDA and AMA had been one of friction and competition (Carpenter 2010). As a result, bills that had not been a priority for years in U.S. congress suddenly became salient following a public outcry about the abysmal state of the drug regulation system.

In addition to support for the efforts of the FDA, the sulfanilamide episode also provided the FDA with a platform to further expose stakeholders to the proposed changes that were seen as necessary to make the marketplace for drugs a safer one. Walter Campbell, the director of the FDA at the time, took the occasion to detail one of the main proposed changes:

> *In the interest of safety, society has required that physicians be licensed to practice the healing art. Pharmacists are licensed to compound drugs. Even steamfitters, electricians and plumbers are required to have licenses. Certainly a requirement that potent proprietary medicines be manufactured under license can be justified on the ground of public safety.*[7]

The idea underlying this statement was obviously to push legislation to move from a system in which the FDA had very limited post-approval jurisdictions to a system in which a drug's therapeutic effects were critically scrutinized before a drug became available to patients. Following this statement, Walter Campbell put out a report in which the idea of further regulating drugs before they could reach the market was laid out. The report contained four recommendations that were intended to be used to draft new regulation: First, drugs should be regulated by subjecting them to a pre-market review and notification. Second, the FDA should be given the power to prohibit the marketing of dangerous drugs already on the market. Third, manufacturers of prescription

[7] This quote appeared in Campbell to Copeland, October 29, 1937: NA, RG 46 HR75A.

drugs should be forced to disclose ingredients used in the drug both including the active and inactive substances. Fourth, claims about the therapeutic benefits of the drug should be regulated (Carpenter 2010). Although these changes in regulation had been proposed before and although the FDA had pushed the administration several times to adopt these rules, the context, support configuration, and urgency had never been as favorable. Adopting these rules would provide the FDA with significant jurisdiction. It would also hand the FDA a monopoly over legal claims about drug safety.

On June 25th 1938 Franklin D. Roosevelt signed the Food, Drug, and Cosmetics Act, which required manufacturers of therapeutic substances to get FDA approval before being allowed on the market. The other proposals were also more or less integrated into the bill, but a salient detail was that the criteria upon which the FDA could deny a substance access to the market was safety, not efficacy (Cavers 1939). In other words, legally the FDA could stop drugs from entering the U.S. market if they posed harm for patients, not if they were not effective as a treatment.

While the role of the FDA and its relations to other stakeholders in the drug safety debate had slowly taken shape before the sulfanilamide case, the tragedy is likely to have formally established the boundaries of expertise and jurisdiction. The result of this new hierarchy was that the stakeholders in the drug safety debate, led by the FDA, had to develop an entirely new system to process and evaluate evidence about drug safety (Carpenter et al. 2010). This showed when the first drugs were submitted for approval under the new law. The main problem was that there was no structured protocol that could be followed that would allow for the acceptance of safe drugs and the rejection of unsafe ones. Such standardization would provide the basis upon which claims could be built (Bowker and Star 1999; Timmermans and Berg 2003). Since this structure was missing, the early approval process could be described as highly experimen-

tal. It was characterized by very few rejections of submitted drug applications, but a large number of serious efforts to force the manufacturer of the drug to revise the label and the claims made on a label. For example, if the research done by the FDA investigator indicated that a drug would be unsafe in certain quantities, the FDA suggested that the manufacturer of the drug included this information on the drug label.

The years following the 1938 regulation were characterized by ambiguity, experimentation, and renewal. An entire agency, an entire profession, and an entire scientific community had to reinvent itself (Parascandola 1992). Historical accounts of this era are a good example of this ambiguity. While some authors portrayed the FDA as an organization that was too closely tied to industry, resistant to change, and unable to respond to the new complexities imposed by rational pharmacological evaluation, others stress the reflexive nature of the agency and see the work done in the years following the new law as necessary steps for the regulatory victories that followed (Carpenter 2010).

The years following the introduction of the Food, Drug, and Cosmetics Act also involved a great deal of precedent-setting work being done at the FDA. One of the divisions that contributed strongly to the emergence of a system of drug evaluation was the Division of Pharmacology. This division was founded just before the Food, Drug, and Cosmetics Act was adopted and it grew rapidly in its first few years. The division became known for its innovations in developing standards, tools, designs, and procedures for testing drugs for their toxicology. A major innovation initiated and further developed by the Division of Pharmacology was to change the meaning of the concept of toxicity. Whereas the debate had been mostly about acute ADRs, the Division of Pharmacology started started doing research on chronic toxicity (the toxic effects following the continuous or repeated exposure to a substance) (Parascandola 1992; Carpenter 2010). Another change in which the Division of Pharmacology at the FDA

played an important role was the erosion of the boundaries between the Division of Pharmacology and the Division of Medicine which allowed for pharmacological practices to be merged into the procedures guiding the medical evaluation of drugs.[8]

The Division of Pharmacology was also actively involved in creating standards to reduce the ambiguity involved in evaluating therapeutic substances. Geoffrey Woodard, one of the FDA pharmacologists at the time, credits Arnold Lehman, the head of the division, for his influence on these efforts:

> *Well, I think that this was probably Dr. Lehman's main contribution, well maybe not main, but a major contribution by Dr. Lehman. He recognized that there were all these various industrial groups coming in and asking for consultation and getting advice but, depending on who happened to be at the conference, the advice wasn't always very uniform. And he realized that there ought to be a book or a source that would have all these things spelled out. So he was instrumental in getting all of us to write about what was known as the "Bible" in the industry.*[9]

The "Bible" to which Woodard refers is the book "Procedures for the Appraisal of the Toxicity of Chemicals in Foods" which was published in 1949 by scientists from the Division of Pharmacology, with Lehman as principal author. This publication began to codify the practices that scientists of the Division of Pharmacology discussed with industry representatives when they met (Stirling and Junod 2002).

[8]The rise of pharmacology as one of the main disciplines involved in evaluation drug safety also created conflict over turf. FDA physicians and others who were not trained in pharmacology began to resent the dominance of pharmacologists in FDA decision making processes. For example, Robert P. Fischelis – the former Secretary of the American Pharmaceutical Association – wrote up an essay that was published in the *American Druggist* with the title "Who should decide to release a drug?" In the essay, Fischelis made the argument that it should be a *medical* decision (Carpenter 2010). However, pharmacology had assumed such a pivotal role in the drug evaluation system that without its expertise the carefully created dependence structure would most likely collapse.

[9]See 1.usa.gov/1BmrgaS.

Besides interaction with industry to discuss standards of drug evaluation, the late 1940 and 1950s were also characterized by the formation of network ties with other stakeholders. For example, Carpenter (2010, pp. 309-313) lists 59 external committees in important organizations (including – but not limited to – the World Health Organization, the American Chemical Society, and the American Medical Association) at which the FDA was represented by at least one FDA official in the mid 1960s. These memberships had been part of the effort of the FDA during the previous decade to engage stakeholders in its quest to improve the safety and efficacy of prescription drugs on the U.S. market. In addition to these committee memberships, the FDA became closely tied to pharmacology departments at prominent universities including the University of Chicago, Harvard, Johns Hopkins, and Cornell. Initially these networks were informal and ties were activated ad hoc, but later these ties were active constantly, not the least caused by the changed hiring practices of the FDA. Increasingly, the FDA became a popular employer among academically trained pharmacologists and mentor-student relationships often translated themselves into relationships of co-workers.

These hiring practices did not go unnoticed by pharmaceutical firms that were trying to get their products approved for marketing. In an attempt not to fall behind, pharmaceutical firms also built extensive ties with pharmacology departments in American universities and were responsible for hiring many of its students (Furman and MacGarvie 2012; Carpenter 2010). Furman and MacGarvie (2012) use a publication of the National Research Council, *Industrial Research Laboratories of the United States*, to describe the emergence and strengthening of network ties between industry and pharmaceutical companies. They show that pharmaceutical research labs set up without industry collaborators grew much less than those with industry collaborators in terms of the number of R&D workers, the number of patents, and the size of the laboratory. The

formation of these networks generated a structure in which sets of people and practices from the market, state, and science interacted.

While rules, procedures, and routines were carefully designed to provide a robust drug safety evaluation system, officials within the FDA increasingly discussed the provision of drug efficacy that was left out of the Food, Drug, and Cosmetics Act of 1938. Through discussions about dosing of a drug, FDA experts started to agree on the fact that the line between safety and efficacy was very fine. Obviously, at a dose that is too high, virtually all drugs are unsafe. This idea then led to the question of how much is needed for a drug to be effective? The main conclusions from these debates was that safety of a drug cannot be studied without paying attention to efficacy considerations. Since the formal mandate outlined in the Food, Drug, and Cosmetics Act of 1938 allowed the FDA to only to reject drugs on the basis of safety, the organization adopted another strategy that would allow them to take efficacy into account when evaluating drugs. By means of a paper, published in 1944 and co-authored by Walton van Winkle, Robert Herwick, Herbert Calvery, and Austin Smith – the first three who were FDA officials and the latter being the Secretary of the Council on Pharmacy and Chemistry (AMA) – the FDA essentially notified manufacturers of therapeutic substances that it would also take efficacy into account when evaluating new drug applications (Van Winkle et al. 1944).

Towards the end of the 1950s, review practices at the FDA changed. Since the approval of a drug by the FDA was increasingly seen by consumers as a signal about safety – not the least because manufacturers used approval to convince consumers of its quality – FDA officials started to internalize those expectations. In doing so, several changes were brought about. One of them was the publication of the Form FD 356 in the *Federal Register*. This was the form that manufacturers – seeking approval for their therapeutic substances – had to fill out when submitting their application. Once the form was submitted, a

case for a New Drug Application (NDA) was opened. While the FD 356 form had been distributed before, it had never been made public. The effect of publishing the form was that expectations about what it was that the FDA could provide were managed. A second major change in the review process was the increased scrutiny by FDA reviewers. Whereas prior to 1955 the majority of all drugs was approved within 3 months, in the early 1960 the approval time increased to more than 18 months for the majority of the drugs submitted to the FDA (Carpenter 2010). What gave the FDA the mandate to do so was the fact that regulation was clear in its goal – drug safety – but was ambiguous and broad enough for the FDA to exercise discretion in determining how companies were to comply with the regulation. One may argue that this discretion allowed the FDA to develop new and innovative practices that could be used to evaluate prescription drugs.

Despite the increased scrutiny during the drug evaluation process, the FDA did not have the authority to regulate what happened before an NDA was submitted. During the investigational stages of drug development companies where not bounded in terms of the type of evidence that they produced. What is more, increasingly accounts started to emerge claiming that pharmaceutical firms were commercializing the clinical testing process. These firms would essentially use the investigational stages of the drug approval process to let the drug penetrate the market well beyond the legitimate set of clinical investigators[10] (Carpenter 2010). Moreover, since the FDA had little to say about the stages prior to NDA submission, very few studies included randomized experiments. While randomized experiments had been proven in the statistical literature to provide a superior design in causes where causal claims were to be made (Peirce and Jastrow 1884), most evidence about drug safety that FDA re-

[10]Pharmaceutical firms would essentially give out the drug to physicians who would clinically investigate the drug in a clinical trial. However, since regulation was so weak, the drug was often given out to a large number of physicians.

viewers had to base their evaluation on in the 1940s and 1950s, came in the form of testimonials. Often, these testimonials were provided by physicians who had received a batch of the drug under investigation and had – non-randomly –prescribed the drug to the patient. Towards the end of the 1960, FDA reviewers and officials increasingly started to raise their voice over the poor quality of evidence presented to them in NDAs (Carpenter 2010).

2.3 The Thalidomide Episode, the Kefauver-Harris Amendments, and the Birth of Modern Drug Regulation

By the end of the 1960s, senator Estes Kefauver from Tennessee, commenced a series of congressional hearings that were intended to scrutinize drug pricing and monopoly profits within the pharmaceutical industry. Kefauver had not intended to touch upon other issues than pricing and profits, but inevitably the discussions hit the topic of drug regulation and drug safety. What stood out in the hearings was that drug regulation had changed substantially since 1938 and that the system had many features in place that were not incorporated in laws (Peltzman 1973). The hearings were one of the first highly publicized events in which the fact that drug regulation practices had changed was receiving so much attention (Carpenter 2010). One of the main changes discussed during the hearings was the inclusion of efficacy as one of the criteria upon which drugs were evaluated. Although the 1938 act mentioned only safety as the criteria that could be used by the FDA to review NDAs, it was widely acknowledged that efficacy was strongly entangled with safety and should therefore take a prominent role in drug approval processes.

The Kefauver hearings also exposed the strong connections between the

FDA and very well respected and high status academic pharmacologists. Pharmacologists from Harvard, the University of Chicago, and Cornell testified in the hearing and explicitly supported the FDA in its actions and behavior. Joel Podolny (2001) describes a social network, and in particular the ties that ego has to alters, as a prism through which ego can be evaluated. In this framework, outsiders seeking to evaluate some characteristic of ego may use ego's alters to inform the evaluation. Moreover, some alters are likely to provide beter quality information based on their status, credibility, or expertise. So while the FDA would have probably benefited from positive evaluations from any testimony during the Kefauver hearings, the positive evaluation of those with high status can be argued to be even more beneficial to an agency who's actions and behaviors are under scrutiny.

During the Kefauver hearings, a tragic but pivotal episode unfolded at the FDA. The story started when the William S. Merrell Company of Cincinatti, Ohio submitted an NDA for a drug called Kevadon (generic name: thalidomide) in the first half of 1960. The FDA reviewer who was assigned to the case, Frances Kelsey, had just joined the organization after having previously been employed as a Professor of Pharmacology at the South Dakota State University (Kuehn 2010). Thalidomide was argued by Merrell to be a mild sleeping pill that would effectively help pregnant women combat symptoms associated with morning sickness. Thalidomide had already entered a number of European markets and Joseph Murray, Merrell's scientific officer, hoped that European approval would make a strong case for fast approval for marketing on the US market. Like the other NDA reviewers in her cohort, Kelsey's work style was one that was characterized by a scientific approach; questioning assumptions and claims and comparing results to prior results published in the medical literature. She thoroughly scrutinized the statistics and research design of the studies in the application, she consulted the medical literature in detail, and she

was critical of any friction between the claims of the company and the results of the studies (Stephens and Brynner 2009).

Kelsey's first response letter is a good example of this careful scrutiny as it questions nearly every claim and argument made by Merrell in its NDA. What further reinforced Kelsey's negative stance towards the NDA of Kevadon was the response of Merrell officials. Some negative evidence was clearly and purposefully withheld from the application and when Kelsey met with Merrell officials early 1961 she "had the feeling throughout the day that they were at no time being fully frank (...) and that this attitude has obtained in all (...) conferences etc. regarding this drug."[11] The fact that the organizational practice of careful scrutiny was well accepted, legitimized, and widespread was shown when Merrell officials refused further interaction with Kelsey and instead approached higher FDA officials. Their response was perfectly in line with Kelsey's (Stephens and Brynner 2009; Carpenter 2010).

While the NDA for thalidomide was further delayed by the additional demands that Kelsey forced upon Merrell, reports came in from European countries – and from Germany in particular – about the potential relationship between treatment with thalidomide and birth defects in children of mothers treated during pregnancy (Stephens and Brynner 2009). Many of the babies of mothers that had used thalidomide during pregnancy were born with deformities including missing or seriously deformed limbs. Current estimates suggest that pregnant women in about 48 countries were treated with thalidomide, resulting in the live births of more than 8,000 babies with deformities. Of the babies exposed between days 35 and 48 after the last menstrual period, 20% to 30% had severe limb defects and other organ defects (Annas and Elias 1999). As the evidence about the link between drugs and defects became stronger, Merrell quietly withdrew the NDA for Kevadon in March 1962 (Carpenter 2010).

[11] nyti.ms/1ukFtYA.

So while the Thalidomide episode was one of absolute tragedy for patients worldwide, the episode had also revealed the strengths and weaknesses of drug regulation systems worldwide. For example, Frances Kelsey's critical stance towards the NDA of Merrell and the fact that her behavior was in line with FDA norms, rules, and practices caught widespread attention. Moreover, the fact that the episode took place during the Kefauver hearings completely changed the topic of the conversation. No longer were the hearings about pricing and monopolies; the focus of the hearings was now drug regulation (Lasagna 1989).

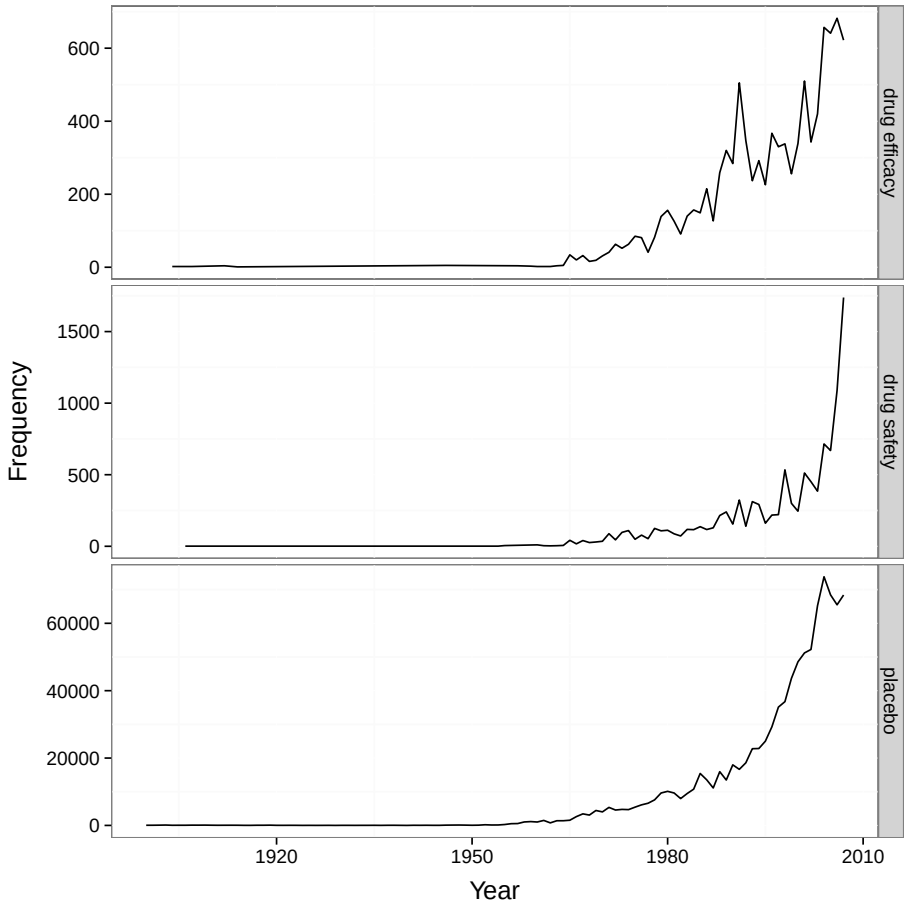
In October 1962, Congress passed the Kefauver-Harris Drug Amendments, requiring companies to provide evidence of efficacy, in addition to safety, for drug approval. Moreover, in 1963 an additional bill was passed that introduced an Investigational New Drug (IND) license. This license was necessary for pharmaceutical companies to start testing the safety and efficacy of a prescription drug and it gave the FDA more authority over the pre-approval stage of the therapeutic substance. For example, the IND amendment included the provision that drugs should go through three stages of trials. In phase I trials the drug is introduced in humans (only animal and in vitro studies are available) with the purpose of determining: human toxicity, metabolism, absorption, elimination, preferred route of administration, and safe dosage range. In phase II trials the number of patients is increased plus the methods, research design, etc. are identified. Finally, in phase III subjects that have a certain disease or condition are treated with the drug to evaluate its safety and efficacy. Pharmacologists were hired to conduct those trials, statisticians were hired to analyze the data, and engineers were hired to provide the digital infrastructure to seamlessly connect the dots. More changes happened; pharmaceutical firms trying to get their substance approved were no longer "manufacturers" but "sponsors" of drugs. They would only become manufacturers of medicines once the FDA approved an NDA.

While the thalidomide case can be seen as the episode that generated urgency, virtually all of the newly adopted laws where proposed explicitly in earlier years by FDA representatives. So the regulatory changes from 1962 and 1963 did not appear out of thin air; rather, the salience, boundaries, and jurisdiction that had been in place as fuzzy concepts for years, now became crystallized.

With the 1962 and 1963 amendments in place, a new era of drug regulation had arrived. Many of the formal structures that allowed the FDA to operate as a regulator of the pre-approval process were laid out. This new and robust structure also allowed stakeholders to further carve out the definitions of the main concepts upon which their work was focused. For example, in the early 1960s there was still little consensus over the meaning of terms such as "new drug", "efficacy", "adequate", "well-controlled", and "scientific training and experience" (Carpenter 2010, p. 269). A simple plot (shown in Figure 2.2) of the use in American English literature of three of the most salient terms is consistent with the hypothesis that the new amendments had provided stakeholders in the drug regulation debate with a platform to start a meaningful dialog.

In the years following the adoption of the new amendments the meaning of these terms started to crystallize and it did so in interactions between a diversity of people and professions. For example, efficacy became understood as a positive effect of treatment as shown in a Randomized Clinical Trial (RCT). However, FDA officials recognized that in some instances RCTs were impossible either because they were infeasible or ethically inappropriate and they would turn to its *experts* to determine appropriate testing for the prescription drug. In those cases the quality of the drug review was not so much captured by the quality of the study but by the expertise and status of the people carrying out the drug review. As a result of the powerful position of the FDA, in which FDA employees had the jurisdiction to carve out the details of pharma-

Figure 2.2: The salience of "drug safety", "drug efficacy", and "placebo" in American literature



Note: This plot is based on data from the Google ngram project and shows that the terms "drug safety", "drug efficacy", and "placebo" started to emerge in the early 1960s in American literature. After a gradual increase up and until the 1990s, the use of the terms "drug safety" and "placebo" have increased exponentially since the early 2000s. The use of the term "drug efficacy" has increased steadily since the early 1960s. Data to replicate the plot can be downloaded from bit.ly/1gfTzNG.

ceutical evaluation, scientific standards changed and in a way, the FDA shaped the meaning of scientific expertise, it shaped hiring practices at major pharmaceutical companies, and it shaped the curricula of prominent medical and pharmacology departments.

In sum, the adoption of the Food, Drug, and Cosmetics Act in 1938 marked a big step in extending jurisdiction and in removing ambiguity about boundaries of expertise. The FDA slowly started to change from an organization with limited jurisdiction to police drugs that were sold on the U.S market to an organization with directive authority. By forming ties to other groups and organizations that provided them with political robustness, the FDA was successful in generating legitimacy for its cause. At the same time, the FDA formed ties with scientific pharmacological experts from prominent universities. Through these ties, the FDA slowly and informally diffused its expectations regarding the quality and quantity of evidence needed to get an NDA approved. Those organically grown sets of expectations, behaviors, and interactions crystallized during the Kefauver hearings. The fact that the thalidomide episode coincided with those hearings allowed the FDA to formalize some of the innovations that had been developed and a new era of modern regulation had arrived. Although some changes in the regulatory process have happened in the 70s, 80s, and 90s, none of those changes were nearly as drastic as the amendments that had passed in the early 1960s.

2.4 The Science of Post-Approval Regulation

Amidst all the changes in the regulation of pre-approval practices, a discrepancy was borne between the pre- and post-approval regulatory expertise and behavior of the FDA. The FDA had been very successful in expanding its expertise and jurisdiction to gain control over the pre-approval evidence production system, yet it also failed to develop a coherent set of practices to continue to evaluate drugs once they hit the market (Ray et al. 2003; Psaty and Burke 2006). The introduction of this dissertation provided a set of possible explanations for why the main focus of regulatory expertise shifted to the early stages in a

drug's life-cycle. However, regardless of its cause, the salient consequence is that an increasingly complex post-approval evidence production system needs to be monitored and interpreted with a limited set of tools.

The production of post-approval evidence is essentially accounted for by two sources. The first source comprises the patients who are treated with a drug and the physicians who prescribe a drug. If patients or physicians think that a drug has been responsible for an ADR they can report the event to the FDA. The decision to report is entirely voluntary; neither physicians, nor patients are obliged to report an ADR. The FDA has developed a system called the Adverse Event Reporting System (AERS) to collect those reports and it uses the reports filed in the system to detect unsafe drugs. The second source that generates evidence and evaluations about the safety of prescription drugs is the medical literature. Each year thousands of research papers are published that study the relationship between a drug and some safety aspect of that drug and the results of those studies are used by the FDA to guide its post-marketing regulatory behavior.

There are two main reasons why the evidence produced after the approval of a prescription drug introduces complexity that the regulator has to understand and account for when making interpretations. First, the evidence that is produced after approval of a prescription drug is massive. Each day the FDA receives more than 2,000 reports of ADRs that are submitted to the AERS. Moreover, the rate at which academics publish on the safety of the 1,500+ drugs that are available on the U.S. market is staggering too. For some drugs, more than 15,000 publications have been published in academic journals and assessing and interpreting the differences between large number of scientific claims about drug safety is a daunting task. Second, whereas prior to approval the FDA has to interpret the claims of one actor whose behavior can – to a certain extent – be directed and disciplined and who is incentivized to comply with

FDA requests, the post-approval setting is characterized by a much larger set of constituents over which the FDA has much less control. Moreover, these constituents produce a much greater variety of evidence that builds on a diverse set of beliefs, methods, and theories. Not only does this introduce uncertainty for the regulator but it also introduces uncertainty for those who produce evidence, especially when claims about safety are build on prior claims about safety.

To deal with this complex and uncertain context, the medical profession has introduced a set of heuristics aimed at providing guidance for action in the healthcare domain. For example, the emergence of "evidence-based-medicine" (EBM) has brought about a series of standards that are used to evaluate medical evidence. Evidence-based-medicine has been defined as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (Sackett 2005, p. 1) and it has been a dominant force in guiding healthcare practices over the past decades (Timmermans and Berg 2003; Timmermans and Mauck 2005). EBM also has had its impact on how evidence about drug safety should be judged. For example, a highly cited article in the *British Journal of Medicine* (Harbour and Miller 2001) proposed a ranking of the quality of the types of evidence that should serve as a guideline when evaluating the validity of the evidence. This ranking – which is shown in Table 2.1 – is intended as a guideline for the evaluation of post-approval evidence. The ranking shows that the more control is exerted on the design of the study the higher the validity of the claim that is advanced.[12]

This ranking also makes explicit some of the complexities and uncertainties associated with evidence of a certain type. For example, there are several reasons why case reports are placed at the bottom of the hierarchy. First, it is

[12]While EBM has gained tremendous support throughout the world, it remains contested by many (Timmermans and Mauck 2005). Timmermans and Mauck (2005) explain the support and contestation of EBM using the sociology of professions (Abbott 1988). Supporters of EBM, on the one hand, argue that the legitimacy of a profession is eroded if there are high levels of variation within the practices that define the profession. Critics of EBM, on the other hand, claim that large scale adoption of guidelines undermines the expertise of healthcare providers.

Table 2.1: The evidence hierarchy

Level	Criteria
1++	High-quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+	Well conducted meta-analyses, systematic reviews of RCTs, or RCTs with a low risk of bias
1-	Meta-analyses, systematic reviews of RCTs, or RCTs with a high risk of bias
2++	High-quality systematic reviews of case-control studies or cohort studies; or high-quality case-control or cohort studies with a very low risk of confounding, bias, or chance
2+	Well conducted case-control or cohort studies with a low risk of confounding, bias, or chance
2-	Case-control or cohort studies with a high risk of confounding, bias, or chance
3	Non-analytic studies (e.g. case report, case studies)
4	Expert opinion

Note: This table first appeared in Harbour and Miller (2001). It classifies different types of evidence based on the method by which that evidence is produced and then ranks them according to the amount of control that the researcher had on the design of the study.

difficult to identify a causal relation between taking a drug and experiencing an ADR. Drugs are not administered randomly, they are not administered in a controlled environment, and the number of cases to which a physician prescribes a drug is typically too low to generate a precise estimate of an effect. Second, not all patients and physicians recognize an ADR as being the result of treatment with a specific drug.[13] Moreover, even if patients or physicians are able to identify a cause-and-effect relation, few of them file a report of the ADR with the FDA (Inman 1985).

What the evidence hierarchy shown in Table 2.1 does not account for are the relationships between the claims in a body of evidence. Sociologists have

[13]Despite these shortcomings, AERS contains a valuable collection of evidence. One of its strengths is its size; each day thousands of reports are filed into the system and since 1997 more than 4 million reports have been filed. A second big advantage of AERS over RCTs is that they contain evidence about the treatment with prescription drugs outside a controlled setting. AERS contains a large variety of patients – each of which is exposed to a drug under different conditions.

demonstrated (Gould 2002; Podolny 1994; Salganik et al. 2006; Shwed and Bearman 2010) that when social actors evaluate objects, people, or phenomena, they do so by using relational cues. Especially when uncertainty is high individuals are argued to increasingly rely on those relational cues to form their opinion and guide their judgment. As I have made clear in the introduction, the argument advanced in this dissertation is that by identifying those structures that inform the evaluations made by individuals, uncertainty in a body of evidence can be reduced and meaning can be derived from a body of evidence. In the following chapters, I will identify some of the social and cognitive structures that guide evidence production and I will show how they are instrumental in guiding the evidence that is produced about the safety and efficacy of prescription drugs.

2.5 Concluding Remarks

This chapter has explored the emergence of the laws, rules, practices, and standards that have allowed the drug regulation systems in the U.S. to become a state-of-the-art bureaucracy. In particular, the FDA and the web of stakeholders with which it interacts have transformed from a loosely coupled system characterized by ambiguity in interaction to a tightly coupled, highly structured system. The ultimate goal of developing this system is and has been to define what a safe and effective drug looks like and how one should go about in terms of making a drug accessible to the American public. Yet, while this system has proven to be successful in defining safe drugs before they hit the market, that same system has been less successful in rapidly identifying potential errors ex-post. Generating an understanding of why it is so difficult to be a powerful regulator in the post-approval context and why it sometimes takes a long time before drugs are taken off of the market is at the core of this dissertation.

This chapter has also argued that evidence about the safety of prescription drugs is created at the intersection of multiple social groups including patients, regulators, physicians, scientists, etc. The evidence about safety, advanced by these groups, varies in content and the content of this evidence also varies within actors over time. I have argued that, in order to make sense of the evidence produced by the stakeholders in the drug safety debate, making evaluations is and inherently social process and that we need to better understand how evaluations are made. Since the comprehension of medical evidence is often assumed to require deep medical knowledge, few social scientists have immersed themselves into a field. However, I will show in the following chapters that social scientists can develop methods and propose designs, such as those developed by Shwed and Bearman (2010), that allows one to evaluate evidence without a need for the context specific knowledge.

Chapter 3

Cognitive Structures and the Effect of Events on the Reporting of Adverse Drug Reactions

This chapter develops and tests a theory that can account for the lengthy market presence of unsafe drugs. The analyses are based on one of the primary data sources used to identify drugs as unsafe – the set of complaints about drugs filed by patients and physicians. The main argument developed in this chapter holds that reporting behavior is guided by the cognitive structure in which the drug implied in the complaint is embedded. This structure stems from similarity in the health condition that a drug treats and is invoked when new information about the safety of drugs is released. The fact that physician behavior is guided by this structure while

patient behavior is not and the fact that this structure is temporarily invoked are sources of ambiguity for the regulator, responsible for monitoring drug safety. The findings in this chapter suggest that an understanding of how and when patients and physicians attribute an effect to a cause may reduce the lengthy market presence of unsafe drugs.

3.1 Introduction

The average drug that has been withdrawn from the U.S. market as a result of serious safety problems has been on the market for more than 10 years, while the median drug that was withdrawn is more than 6 years old. How is it possible that drugs, killing and injuring many people, are on the market for years without the problem being detected earlier? This chapter takes up that question and points to ambiguity in the data that the FDA uses to detect safety problems. The argument developed in this chapter holds that a social theory of how this data is generated is needed to understand why regulatory action is slow to materialize.

The primary data used by the FDA to detect unsafe drugs are the complaints that it receives from patients and physicians about an adverse health effect[1] that they have experienced and that they attribute to a prescription drug. The FDA monitors this data to detect drugs that are linked to ADRs in disproportionately high numbers.

While the set of complaints of patients and physicians is one of the richest resources for studying drug safety[2], it comes with several drawbacks. First, the data is known to be plagued by high levels of underreporting (Martin et al.

[1] Throughout this dissertation I will use the term Adverse Drug Reaction (ADR) for all instances in which an actor claims that a drug has caused a negative effect on one's health. The term "adverse health effect" is used to describe the case in which drugs are not (yet) seen as the cause of the negative effect on one's health.

[2] The data have been used in hundreds of scientific studies and it is the primary resource used by the FDA to detect unsafe drugs (Wysowski and Swartz 2005).

1998). Several studies suggest that only 5% of all ADRs are reported and that this percentage is likely to vary by drug and over time.[3] A second major problem is the potential of misattribution. Patients and physicians who report a complaint *think* that a given drug caused an ADR, but it is difficult to establish that the causal relation truly exists. Patients often take multiple drugs, and the lack of formal training in pharmacology for both patients and physicians makes it difficult for them to identify the true cause (Tavory and Timmermans 2013). Finally, every drug causes ADRs and there is no commonly agreed upon threshold for the number or the severity of ADRs that is considered acceptable. The FDA has the discretion to move from case to case to decide beyond which level a drug is considered to be unsafe. These three drawbacks are further accentuated by the fact that the number of patients that takes a given drug is often unknown. Although drugs are sold through licensed pharmacies, there is no centralized system that records the number of patients to which a drug is sold and in which quantities.

The drawbacks of the data, joined with the absence of information on the population exposed to the drug, create high levels of ambiguity when identifying patterns from complaint reports and when formulating subsequent regulatory action. In this chapter I will show that the crux in interpreting disproportionality in complaint data is to distinguish between disproportionality caused by *an increase in the number of ADRs* for a given drug and disproportionality caused by *an increase in the number of complaints* filed for a given drug and for a fixed number of ADRs.

I will do so by showing that through the identification of the social process by which patients and physicians report their complaints, ambiguity in the interpretation of the data can be reduced. In particular, I develop and test a theory

[3]Obviously, if this 5% was accurate for each drug, and stable over time, one could simply infer the true population of ADRs for each drug. Unfortunately, there is no way of knowing when and for which drugs this 5% rule holds.

about how and when patients and physicians attribute an adverse experience to a prescription drug.

For the analyses in this chapter I identify 12 events – the recalls of 12 prescription drugs – and I observe how patients and physicians respond in terms of their reporting behavior to the new information about drug safety that the recall reveals. In developing hypotheses about the behavior of patients and physicians following the recall of a prescription drug, this chapter builds on research on the sociology of cognition and social psychology which shows that individuals use cognitive schemata to make sense of their experiences (DiMaggio 1997; Cerulo 2010; Negro et al. 2010). These schemata are knowledge structures that represent objects or events and the relations between these objects or events. I contribute to this literature by showing how the effects of a recall spread beyond the drug implied in the recall and by accounting for the differences between patients and physicians in terms of the schemata that guide their behavior. Informed by research on medical sociology I operationalize the schema that guides the attribution process as a network of drugs structured by disease categories (Rees 2011).

The results from my analyses show that immediately following the recall of a prescription drug, a sharp increase can be detected in the number of reports that are filed for the drug implied in the communication of the recall, as well as for a set of drugs that treat the same disease. This increase in the number of reports comes from patients and physicians who would not have reported, had the recall of the focal drug not been communicated. Moreover, the findings indicate that the second pattern, an increase of complaints filed for drugs that treat the same disease, is only found for physicians. Based on the differences in the attribution patterns observed for patients versus physicians, this chapter shows that audiences can be partitioned based on the cognitive schemata to which they are tied. Finally, I examine the time it takes for the FDA to take

regulatory action after a signal has been detected. In particular, I estimate the size of this lag and compare the lag in periods where a recall had taken place to the lag in periods where no recall had taken place. Preliminary findings show that the lag is larger in the period where the recall had taken place. These results are consistent with the argument that recalls produce noise in reporting which delays the evidence interpretation process at the FDA.

The stakes of detecting safety issues earlier are high. Every day, the FDA receives more than 2,000 reports from patients and physicians about the adverse effects of prescription drugs. And estimates suggest that these 2,000 complaints represent only 5% of all ADRs. Moreover, the problem of high levels of ADRs extends beyond the patient: Costs of ADRs for national health care systems are tremendous (Rodriguez-Monguio et al. 2003), physicians that fail to identify them become uncertain about their expertise (Wears and Wu 2002), sponsors of drugs associated with disproportionately high numbers of ADRs may be held responsible, potentially leading to severe financial consequences (Sarkar and de Jong 2006), and the Food and Drug Administration (FDA) may suffer serious reputational damage when a drug that was approved is found to cause large numbers of ADRs (Carpenter 2010).

By identifying the social process of attributing an effect to a cause, this chapter shows how noise in a set of data points can be transformed into a predictable pattern. In doing so, it reduces ambiguity when interpreting signals contained in the reports of complaints filed by patients and physicians. This problem of making sense of data that is voluntarily contributed and that is based on interpretations of an object or event is very general (Langley 1999). Examples include data used for the early detection of food-borne illnesses[4] or technical defects in cars[5]. Besides its substantive contribution, this chapter contributes to

[4]See for example bit.ly/1uJOZjV.
[5]See for example nyti.ms/1qCq43B.

the literature on cognitive sociology and organization theory by showing how cognitive schemata guide evaluation processes and how audiences can be partitioned based on the schemata that they use. While this literature provides many examples of how variation in evaluations can be explained by characteristics of those who are evaluated, relatively little is know about the audience that provides the evaluations. This chapter addresses that issue.

In section 2, I describe the context of post-approval evidence production in further detail and I explain why accounting for the social conditions under which patients and physicians report their complaints is important. In section 3, I introduce and develop a theoretical framework that provides guidance in the understanding of ADR reporting patterns and in section 4 I develop and state the hypotheses tested in this chapter. In section 5, I provide a rationale for the analyses conducted in the chapter. Section 6 to 8 present the data, the empirical strategy, and the results respectively. In Section 9, I summarize and discuss the findings.

3.1.1 Post-Approval Learning About Safety and Regulating Prescription Drugs

In order to assess whether a prescription drug is safe, it goes through many rounds of evaluation. The first rounds of evaluation are comprised of a series of clinical trials conducted by the sponsor of the drug. If these trials indicate that the drug is an effective and safe treatment for the disease it is developed for, the drug is likely to be approved by the FDA and admitted access to the market.

Once a drug is admitted to the market, physicians start prescribing the drug and patients start using the drug. The conditions under which patients are treated with a drug in the pre-marketing stage are radically different from the

conditions in the post-marketing stage. First, heterogeneity among patients exposed to a given drug in the post-approval stage is much higher than heterogeneity among patients in the clinical trial stages of drug development (Epstein 2007; Timmermans and Epstein 2010). As a result, the newly approved chemical compound is released on a much more diverse set of biomedical human bodies and it is likely that some of these new combinations of chemicals on the one side and human bodies on the other will result in ADRs. Although various public health advocacy groups have successfully lobbied for increasing heterogeneity in sex, race, ethnicity, and age among subjects in biomedical research (Epstein 2007), including the full range of heterogeneity among human bodies in clinical trials is financially and practically infeasible. A second complication that is typically not accounted for in a clinical trial is the concomitant use of prescription drugs. In order to increase the treatment efficiency or to treat diseases occurring simultaneously, drugs are prescribed concomitantly and often the chemical interaction between the multiple drugs has not been studied in a clinical trial. Evidence suggests that the negative consequences of these interactions are substantial and some estimates suggest that they account for about 30% of all ADRs (Tatonetti, Fernald, and Altman 2012). Finally, clinical trials are conducted under controlled conditions. For example, food, temperature, physical exercise, and use of the medicine are controlled by the medical staff. This controlled environment allows researchers to isolate mechanism of action, but it fails to account for interactions between the drug and external conditions. These three differences make it virtually impossible to detect or identify all potential ADRs before a drug is released on the market. Therefore, the FDA aims at the early detection of post-marketing signals that indicate a relation between a drug and an ADR.

Given the fact that virtually all drugs cause ADRs once they are released on the market (Lazarou and Pomeranz 1998; Pirmohamed et al. 2004), the FDA

aims to intervene whenever drugs become associated with disproportionately high numbers of ADRs. The FDA uses various strategies to minimize the likelihood that drugs will seriously compromise public health. One of the most often used strategies is the request for a label change. In such cases, the FDA finds that there is enough evidence to warrant a change in the label of the drug that indicates a new side effect of the drug or that provides further detail to a side effect that is yet known. A far more serious regulatory tool is the request for a Boxed Warning. This tool also involves a label change but rather than adding or adjusting some text on a multi-page label, the Boxed Warning appears at the beginning of the label and is accentuated by a black box. This regulatory tool is reserved for cases in which the drug, under certain conditions, can cause ADRs that are a fatal, life-threatening, or permanently disabling (Murphy and Roberts 2006). In some cases, the Boxed Warning is directed at specific demographic characteristics of the patient and does therefore exclude that patient from treatment with the drug. Finally, the most severe regulatory action that the FDA can take is the withdrawal of a prescription drug. In such cases, the FDA decides – together with the drug sponsor – to take the drug of the market indefinitely. These most severe interventions (Boxed Warnings and drug withdrawals) pose a challenge for the FDA; each regulatory action should be aimed at minimizing the risk of ADRs without denying access to the drug for patients that benefit from it (Eichler et al. 2013). Despite the large numbers of ADRs, there is ample evidence that the FDA is a strong regulator and that their actions benefit public health tremendously (Carpenter 2010).

The primary source of evidence that the FDA uses to detect unsafe drugs after approval is the Adverse Event Reporting System (AERS). This system is maintained by the FDA and allows patients and physicians to file their complaints about prescription drugs directly or via the manufacturer of the drug who is then obliged to submit the information to the FDA. As mentioned ear-

lier, while the data in AERS is a valuable resource for detecting safety problems with prescription drugs, it suffers from underreporting, misattribution, and an undefined baseline of acceptable ADRs. Recent years have been characterized by efforts of the FDA to increase the quality of the data by increasing awareness and ease of reporting and by providing clear guidelines of what and when patients and physicians should report (McClellan 2007). But while these efforts are likely to have improved the quality of the data in AERS, they have not eliminated its problems.

To understand why the drawbacks in AERS are such a problem for the regulator using the complaints from patients and physicians to identify unsafe drugs, I will briefly review how the FDA uses the data in AERS to detect signals.[6] The FDA monitors complaint data to detect disproportionality in the occurrence of a given drug-ADR pair vis-a-vis all other drug-ADR pairs. They do so through a case/non-case methodology. In a case/non-case approach the population of reports is split into two samples; one that contains all reports that name drug i as the potential cause of an ADR (cases) and one that contains the complement of the reports that name drug i (non-cases). These two subsets can then be further partitioned into the reports that report ADR a – for example arrhythmia – and the reports that report all other ADRs – all ADRs but arrhythmia. They then employ various statistical methods to compute the proportionality of the occurrence of an ADR in those being treated with drug i and the occurrence of that same ADR in those treated with other drugs. If the number of reports for a focal drug increases disproportionately relative to the number of reports for other drugs, the FDA "detects a signal". The assumption is that this increase in the number of reports for a focal drug relative to the number of reports for other drugs is an indication that the focal drug might be unsafe.

[6] Appendix A in this chapter contains a more detailed description of the methods used by the FDA to detect unsafe drugs from data in AERS.

This chapter shows that this assumption does not always hold. The argument advanced in this chapter is that the social process by which patients and physicians are induced to report into AERS may both cause true safety problems to be masked and "socially constructed" signals to be detected. The next section further develops this argument by theorizing the processes through which patients and physicians identify the cause of an effect.

3.1.2 Sense Making, Relational Structures, and Heterogeneity within Audiences

Despite the importance of understanding the data used to detect these unsafe drugs, very little research has been done on the process by which patients and physicians decide to file a complaint about a prescription drug. The fact that AERS is characterized by high levels of underreporting makes this process especially salient. If each ADR that ever occurred could be perfectly identified and if reporting of ADRs was mandatory, one would not need to theorize about the social factors through which the data was shaped. However, the drawbacks of the data leave ample room for social processes to work their way in to the data. Research on other domains of the health care practice has identified a battery of social factors that predict increased participation. For example, the use of medical services and the beliefs held about illness and disease has been shown to be predicted by ethnicity, race, and gender (Jenkins 1966; Suchman 1964; Landrine and Klonoff 1992). This chapter further builds on those ideas and it stresses the salience of understanding the process by which patients and physicians end up contributing a complaint to the data.

One of the factors that can account for underreporting is the failure to recognize an adverse health effect as an ADR. Medical research suggests that both patients and physicians are often unable to identify an adverse health effect as

an ADR (Tegeder et al. 1999). Literature in social psychology offers a theoretical framework that helps us understand how patients and physicians transition from failing to recognize an ADR to attributing an adverse health effect to a prescription drug. In trying to understand how people make causal explanations, social psychologists have theorized and tested processes of attribution (Kelley 1973; McArthur 1972). In particular, this literature aims to understand which information individuals use to answer causal questions and how they transform information to a causal interpretation. Three classes of potential causes are typically identified: persons, entities, and times (Kelley 1973). To illustrate how these three classes may be used to understand the process of attributing adverse health effects to a prescription drug, it is important to stress the scope conditions under which patients and physicians make their causal interpretations. A first important condition is that patients who take prescription drugs are ill in the first place and are therefore likely to experience adverse health effects that are *caused by the illness* rather than by the prescription drug. A second scope condition that is important for the interpretation of adverse health effects is that when a prescription drug is approved by the FDA, it has been shown by experts that the drug is safe and effective (Gieryn 1999; Temple and Himmel 2002).

Given these two conditions, the most common scenario in which patients or physicians make sense of adverse health effects involves a patient who is administered a prescription drug and experiences an adverse health effect. Based on research in other contexts, social psychologists (McArthur 1972) suggest that if the patient takes a prescription drug and experiences an adverse health effect, he or she is unlikely to attribute the effect to the prescription drug. Since the drug was argued to be safe by credible experts and since the patient was ill in the first place, it is unlikely that the adverse health effect is caused by the drug. The adverse health effect is much more likely to be caused either

by some characteristic that is particular to the patient or by the temporal context or circumstances in which the effect was experienced. However, Kelley's work (Kelley 1973) suggests that this line of reasoning changes if the patient gets exposed to other patients who have taken the same drug and experienced the same adverse health effect. Such an exposure may happen either through social interaction or through a broadcast event that reveals that more patients taking the drug have experienced the health effect. A similar line of reasoning applies to the interpretation made by the physician. Physicians may observe more than one case in which a patient is administered a drug and experiences an adverse health effect and it is not unlikely that the co-occurrence of multiple adverse health effects within the same medical practice causes physicians to report into AERS. However, the increased likelihood of observing multiple similar cases only increases the baseline rate at which physicians are expected to make causal claims. It does not eliminate the effect that a broadcast event may have on the attribution of adverse health effects to a prescription drug.

Given the fact that patients and physicians are likely to attribute an effect to an external entity if they are exposed to others who link the same effect to the same external entity, highly publicized broadcasts about drug safety may induce patients and physicians to report an ADR. The scenario laid out in the previous paragraph suggests that such a broadcast may even cause individuals to revisit past experiences and make sense of them using the new information that highly publicized events reveal. Literature outside the realm of medical science also theorizes about how the release of new information may affect one's evaluations of an organization, person, or phenomenon. For example, recent research shows how evaluations of an organization or product are altered when negative information on the organization or product are publicized (Roehm and Tybout 2006; Jonsson et al. 2009). Audiences, comprised of consumers or other exchange partners, use the new information to review the

status, quality, or morality of the target of the new information and they adjust their behavior accordingly. Using similar arguments, Adut (2005) shows that even if the newly released information is not new – but only brings publicity to commonly held beliefs – repercussions for the target of the negative exposure may be severe.

This same line of research has shown that newly released information does not only have an effect on the interpretation of the object or person targeted in the information but also on others. For example research on stigma has provided detailed descriptions on how stigma's associated negative consequences spread through an entire population even when only a handful of targets are directly stigmatized (Goffman 1986; Pontikes, Negro, and Rao 2010). Various modes through which negative consequences – associated with a given stigma, identity, or status – may spread have been proposed. Pontikes et al. (2010) show that stigma resulting from adherence to communist ideology was readily transmitted through casual professional associations. Other research has shown that individuals, organizations, or objects belonging to the same category as the target of negative information are likely to experience repercussions. For example, Legewie (2013) shows that attitudes towards immigrants are influenced by terrorist attacks attributed to a group that identifies itself as an Islamist group. The mechanism responsible for creating and enhancing anti-immigrant sentiments following such an event involves making associations between the main actor in the event and immigrants that bear no responsibility for the event. Thus, observers extract certain critical and meaningful pieces of information from the behavior of a single social actor and internalize them followed by a (temporal) revision of their attitudes and perhaps behavior with regards to a large group of other social actors. Another example is provided by Roehm and Tybout (2006) in their research on scandals that argues that scandals are – under certain conditions – likely to spill over. In their work, the authors hypothesize about the

typicality of the target of a scandal for the category that it belongs to. Findings indicate that negative externalities are more likely to spill over when the target of the scandal is typical of the category. Conversely, they argue that when the evaluators of the object are primed to differentiate between two or multiple objects, the contagion effect will be limited.

Despite the fact that the literature on how evaluations spread typically uses these evaluations as dependent variables, little is know about those who contribute evaluations and the heterogeneity among these individuals. Recently, Kocak et al. (2014) have argued that audience members – those who observe and make evaluations – vary in the heuristics that they employ. Audience members hold different sets of prior beliefs, they differ in their vested interests, and they vary in the level of expertise that they have about the object or organization that is to be evaluated. Kocak et al. (2014) argue that the heterogeneity among audience members affects how consensus about the meaning of an object, person, or organization is formed and how this consensus is spread among a wider audience. By making this distinction the authors raise the salient idea that the outcomes of evaluation processes depend on heterogeneity among audience members. This chapter further builds on the idea about heterogeneity among audience members. It argues that heterogeneity in expertise is associated with the ability to generalize from a single case to a more extensive domain of evaluation. This research builds on the idea that heterogeneity in the formal expertise of audience members creates variation in the way in which the outcomes of events may spread. The next section further translates these ideas to hypotheses and it links theory on attribution, evaluation, and categorization to the case of drug safety.

The Release of New Information and the Spread of Increased Reporting

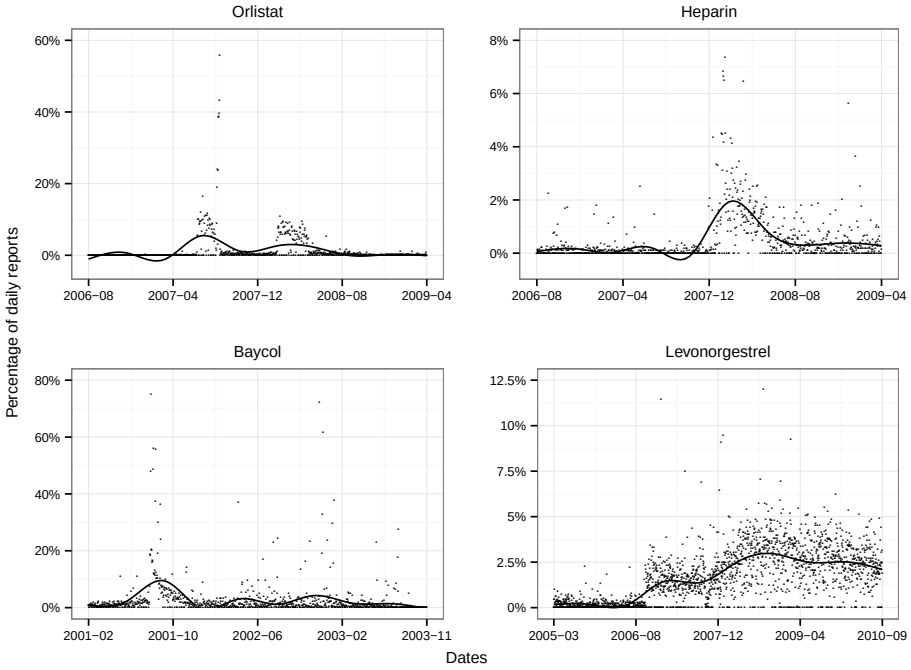
The hypotheses tested in this chapter are concerned with questions of how patients and physicians attribute an effect to a cause, what the extent of this attribution pattern is, and how this attribution pattern becomes clustered in time. To understand why the temporal and categorical clustering of attribution patterns is important, I first show some of the complaint data that the FDA has had to analyze and I explain the main challenges. After that, I will bring the literature on attribution, categories, and contagion together with the empirical case at hand and I will briefly discuss the hypotheses that will be tested in the remainder of the chapter.

Figure 3.1 shows four episodes – one for each of four prescription drugs – characterized by a rapid increase in the reports filed for a specific drug. Episodes like these are what FDA researchers look for when conduction signal detection studies (Poluzzi et al. 2012). These data come from the Adverse Event Reporting System (AERS) and each panel shows the times series data for one specific drug. The main question that the data plotted in each of these four panels gives rise to is whether the relative increase in the number of reports is caused by an increase in the number of ADRs or by an increase in the number of reports for a given drug.[7] This chapter suggests that an understanding of the social process by which effects are attributed to causes allows one to distinguish between these two explanations.

The research in social psychology on causal attribution would predict that the release of and exposure to new information that indicates that other patients have also experienced an ADR after taking a specific drug should induce patients and physicians to revisit past experiences and revise their attribution of

[7] Please note that in this second case, the number of ADRs that occurs remains unchanged.

Figure 3.1: Temporal Changes in Reporting Ratios of ADRs



Note: This graph plots the temporal changes in the number of reports that list a specific drug as the primary suspect of an adverse event. The y-axis represents the percentage of total reports per day that indicate the specific drug as primary suspect and the x-axis represents time, measured by day. Orlistat is a drug that treats obesity, Heparin aims to prevent the clotting of blood, Cerivastatin is prescribed to lower cholesterol and prevent cardiovascular disease, and Levonorgestrel is an emergency contraceptive pill.

an effect to a cause. Moreover, there are at least two other pathways that should induce patients and physicians who would otherwise not have reported to report. First, the release of new information about the safety of a prescription drug makes patients and physicians aware of the possibility to report ADRs to the FDA. As mentioned before, many patients and physicians are not aware of the existence of AERS and many potential reports do therefore never materialize. The second pathway leading to increased reporting builds on the idea that the fact that others have expressed that a drug has caused an ADR should serve as a legitimation device. If patients and physicians were aware of the potential

causal link between drug and ADR and the option to report to into AERS, but felt that their report was not warranted, the official statement by the FDA may have legitimized their claim. In this chapter, I will use the recall of a prescription drug as the event that reveals new information about drug safety. A recall is the most severe tool in the regulator's toolkit and recalls are often widely publicized.[8] I test the following hypothesis:

H1: The communication about a recall of drug *i* will cause patients and physicians who would otherwise not have reported to report complaints for drug *i*.

Although a recall of a prescription drug targets only one drug, it is not unlikely that its effects will extend beyond the drug implied in the recall. As suggested by research on stigma, audiences often use social ties or categorical similarity between the target of negative attention and other non-targets as the structure by which inferences about quality, status, or identity can be made. For prescription drugs, one of those structures is the similarity in the disease(s) that two prescription drugs treat. Timmermans and Buchbinder (2012) demonstrate the salience of disease categories in their analysis of newborn screening. They argue that diseases are prominent categories along which understanding of conditions and treatments is shaped. Moreover, in communicating a recall the FDA consistently mentions the disease that the recalled drug is approved for. In doing so, the FDA focuses attention to the disease category to which the drug belongs.

Although I expect the increase in the number of reports for drugs that treat

[8] A recall is unlikely to be the only source that is expected to cause temporal variation in reporting. Other events, such as widely publicized academic research, may also force patients and physicians to re-evaluate prior experiences. Moreover, some social behavior may cause the the incidence of certain ADRs to go up. For example, David et al. (2010) find that post-marketing promotional activity may involve the risk of inappropriate drug prescriptions, leading to regulatory actions against the firm. They provide some evidence that increased levels of promotion and advertising lead to increased reporting of ADRs for certain conditions.

the same disease to be much smaller than the increase in the number of reports for the drug that is recalled, similar micro processes may underlie the increase. The fact that a drug that is used to treat the same disease has been shown to be unsafe may cause some to use this information and revisit the prior causal interpretation of an experience. This is especially true if the type of ADR that is caused by the recalled drug is similar to the ADR that is caused by the drug that treats the same disease. Along similar lines, the legitimation effect may also extend to drugs that treat the same disease. Mere association to the recalled drug in terms of disease similarity, plus the fact that the disease is mentioned in the communication about the recall may result in the legitimation of potential complaints. Finally, it is unlikely that the awareness effect can account for the relative increase in the number of reports filed for drugs that treat the same disease. Although awareness may certainly cause an increase in the absolute number of complaints filed for drugs that treat the same disease, awareness will simultaneously increase the absolute number of complaints filed for all other drugs, thereby keeping the ratio stable. To examine the extent to which the meaning of newly released information is extended beyond its original context, I test the following hypothesis:

H2: The communication about a recall of drug *i* will cause patients and physicians who would otherwise not have reported to report complaints for drug *j* if and only if drug *j* treats the same disease as drug *i*.

Although making a connection between two drugs that treat the same disease may seem straightforward, research on the "health literacy" of patients suggests that this is not per se the case. Research has shown that despite the large volume of freely available information about diseases, drugs, and other health related topics, a large number of patients suffers from "health illiteracy" (Berkman et al. 2011). Broadly speaking, "health literacy" refers to the set of

skills that people need to understand and navigate the health care environment. Examples of skills include the ability to read and understand patient labels of prescription drugs (print literacy); interpret quantitative information contained of drug or food labels (numeracy); and effective communication about health topics with health care providers (oral literacy) (Berkman et al. 2011). Health illiteracy may cause the effect of a recall in the reporting ratios of drugs that treat the same disease to be stronger for physicians than for patients. Other findings in the medical scientific literature suggest that physicians are primed to reason based on disease classifications. For example, Barabasi et al. (2011) argues that while physicians lack the proper training to understand the chemical dependencies of drugs, their formal training and clinical experience provides effective guidance in navigating dependencies between diseases. To test the effect of differences in the use of cognitive schemata between patients and physicians, I propose the following hypothesis:

H3: The contagion effect will be stronger for physicians than for patients.

The empirical strategy employed in this chapter aims to allow for a causal interpretation of the observed effects. I therefore briefly review some of the potential alternative explanations and I show what can be done in the analysis to eliminate these alternative explanations. Essentially, an increase in the number of reports received by the FDA could arise from three main sources: First, the sudden increase in reporting of both the focal drug and its neighboring drugs can be caused by a sudden increase in the number of patients to which the drug is prescribed. Given that a certain percentage of patients will experience an ADR from taking the medication, the number of ADRs will go up and this rise could account for the increase in the number of filed reports. Several mechanisms can be identified that would lead to a sudden increase in the number of prescriptions. For example, a new scientific insight or the im-

plementation of new guidelines for physicians might cause the pool of patients eligible to be treated with the drug to go up. The increase would be further strengthened if the rise in prescriptions was accounted for by individuals that have demographics in common. If the new consumers of the drug had a higher likelihood of experiencing an ADR and the effect-reporting interval[9] remained stable, one might expect to see a sharp increase in the number of reports filed for a given drug. This source of increased exposure could both cause the number of reports for focal drugs and the number of reports for drugs that treat the same disease to go up. Another mechanism that would lead to a sudden increase in the number of patients to which the drug is prescribed, is physicians switching patients from the withdrawn drug to the related drug. Obviously, this source of increased exposure would only lead to an increase of the number of reports for neighboring drugs. To rule out this alternative explanation, I conduct an analysis that assesses the variance in the interval between experiencing the ADR and reporting it to the FDA. If the increase in the number of prescriptions was sudden and the mean interval would remain unchanged, the variance of the interval would have to reduce discontinuously right at the time of the announcement of the recall. A gradual increase in the number of prescription would not cause a discontinuity right at the time of the event given that the mean interval is continuous at that same time.

Second, the sudden increase in reporting of both the focal drug and its neighboring drugs can be caused by a sudden increase in the number of patients experiencing an ADR. For example, both the focal drug and the drugs that treat the same disease could see a sudden increase in the number of reports filed if the number of ADRs rapidly grew following exposure of patients to external strain such as stress at specific days of the year. Some ADRs may

[9]The number of days between the date at which the ADR occurred and the date that it was reported.

be more likely to present themselves when the human body is under stress. As a result, those ADRs have incidence rates that are higher during some days of the week, and, given a fixed effect-report interval, are more likely to be reported during some days of the week. If such ADRs are closely related[10] to the use of certain prescription drugs, reports for these drugs are likely to go up. Another external source of variation in reporting may be seasonal weather changes. If certain ADRs are more likely to occur during warm weather, the first heat wave of the summer may increase the incidence of that ADR. And again, given fixed effect-report intervals, ADRs caused by these seasonal whether changes may be cause sudden shocks in reporting. To evaluate the second alternative explanation, which holds that an increase in the number of ADRs, leading to more ADRs being reported could be the cause of increasing reporting ratios, I test whether the mean event-reporting interval remains stable right at the time of the cutoff.

The final explanation, and the explanation advanced and hypothesized in this chapter, holds that the sudden increase in reporting is neither caused by an increase of prescriptions, nor by an increase in ADRs, but by the reinterpretation of patients and physicians of past experiences.

3.1.3 Roadmap

The aim of the analyses in this chapter is to determine whether safety communications by the FDA about the withdrawal of a prescription drug caused the number of ADR reports filed for the withdrawn drug and the drugs that treat the same disease to increase relative to the full sample of reports. I do so by conducting several tests – each of which attempts to disentangle the various alternative explanations described in the previous section.

[10] "Closely related" here refers to a co-occurrence of adverse effect x and a drug y that is higher than chance.

The first set of analyses focus on identifying the effect of the safety communication on the number of reports filed for the drug that is withdrawn. I first estimate the parameters for several Poisson models in which the independent variables include a treatment dummy, indicating whether the reports are filed before or after the warning, day dummies, month dummies, the total number of reports filed at a given day, and the number of days to or from the safety communication. Each model is estimated for a different window around the announcement of a recall, ranging from seven days before and after to 35 days before and after the announcement date.

Limiting the sample to include observations from a relatively narrow window is important because it helps disentangle the effect of the safety communication from the effect of other time-varying factors that influence the reporting ratios into AERS. However, even within a relatively narrow window, unobserved factors that change over time could be a source of variation. These, in turn, can cause the error term in a Poisson regression to be correlated with time, allowing for the potential of biased estimates of the treatment coefficient. A Regression Discontinuity (RD) design can be used to overcome problems caused by confounding variables. It does so by considering an arbitrarily narrow window of time around the safety warning. Within this window, the unobserved factors influencing the reporting ratio are assumed to be similar so that observations before the event provide a comparison group for observations after the event. For the case at hand, this implies that patients' and physicians' pre- and post-communication reports are not drawn from samples with different distributions of key variables. One of the major advantages of the RD design is that this core assumption can be evaluated by examining whether the distribution of the covariates other than the treatment effect is similar right before and right after the event. Therefore, I then check whether patients and physicians reporting into AERS before and after the event differ in their key characteristics. In

sum, the reports that form the basis for calculating the reporting ratios are the same before and after the event, meaning that the characteristics of patients and physicians do not change in response to a safety communication (Davis 2008).

After estimating the Poisson coefficients, conducting the RD analysis, and evaluating the other key variables, I probe whether there is a difference between patients and physicians in the way in which they respond to a safety communication. I do so by conducting the RD analysis for patients and physicians separately.

Although the main interest of the chapter is in changes in the reporting ratio, I use the RD design to estimate the effect on the ratio and the absolute number of reports. Obviously, an increase in the ratio could be caused both by the decrease in the denominator, as well as by an increase in the numerator. And, although theoretically interesting, an increase in the absolute number of focal drug reports that is accompanied by the same percentage increase in the absolute number of other drug reports does not cause scientists to detect a signal (disproportionality in the composition of drug reports).

After analyzing the effect on safety communications on the reporting ratio of the focal drug, I turn to the effect on the reporting ratios of drugs that treat the same disease. Essentially, the analyses for the related drugs are replications of the analyses for the focal drug. They include the Poisson regressions, the RD design, the evaluation of key variables, and the split into the patient versus physician strata. In the next section I describe how the data used in this chapter was obtained and how the raw data was transformed into a workable database. Then I will briefly return to the details of the empirical strategy before I describe the results.

3.2 Data and methods

3.2.1 Study population

3.2.1.1 AERS

The main data used in the chapter is the FDA Adverse Event Reporting System (AERS). This reporting system is a vital source of information for the FDA in their efforts of promoting high standards in public health (Robb et al. 2012). AERS is also used in hundreds of scientific studies and its content has recently become more accessible through websites such as "adverseevents.com" and "fdable.com". The data in AERS from 2003 to 2013 is freely available from the FDA website, while the data from 1997 to 2003 can be purchased from the NTIS.

AERS is a longitudinal database comprised of ADR reports of patients, physicians, and other healthcare providers. Reports on ADRs can be submitted to the FDA either directly by the patient or physician or through the manufacturer of a drug to which patients and physicians have reported. Manufacturers are obliged by law to submit the reports that they receive with the FDA. Multiple reports, each of a different ADR for the same patient, can be linked through a case number, but it is still possible that one ADR instance is reported multiple times. To circumvent this problem, the FDA recommends a de-duplication strategy to limit the bias that can be caused by duplication of reports. I follow Poluzzi et al. (2012) and remove reports that are likely to be duplicates of other reports from the dataset (a full description of the de-duplication strategy is available from the author). Another implication that follows from multiple source reporting is that one has to be careful in assigning a date to the report. Since patients file directly to the FDA and physicians file directly on behalf of their patients, I code the day at which the report is sent to the FDA from these

two sources as the date of interest. In the case of reporting by manufacturers, the day that the report was received by the manufacturer is coded as the day of interest. Typically, there is a one or two week lag between the date at which the manufacturer received the reports and the day that the manufacturer forwards to report to the FDA.

A major challenge in using the data is to correctly assign each report to an FDA approved prescription drug – with a standardized drug name. The AERS data contains 20,061,582 fields in which free text is entered to describe the drug that the reporter has associated with the reported ADR. To meet this challenge I constructed a dataset of all drugs approved for marketing by the FDA since 1950. The next section further explains how this data was constructed. Using highly restrictive matching criteria to minimize the number of false-positive matches I was able to match the drug field to a standardized drug name in 98.5% of all cases – a total that lies roughly 9% higher than the number of matches in Poluzzi et al. (2012). A total of 97.4% of all fields was matched against drugs in the database of FDA approved drugs[11]. The discrepancy between the sizes of these subsets as shares of the entire set of fields can be traced back to the fact that, although the FDA clearly states that reporters should limit their reports to adverse events likely to be associated with U.S. approved drugs, some reports contain references to drugs that were approved in other countries but not in the US. In accordance with other studies that use AERS to detect safety issues, the data was limited to reports that were filed from within the US. Moreover, I only focus in "primary suspect" drugs. These are the drugs that reporters name as the likely cause of the ADR.

[11] Each field should contain only one drug name. However, in some instances a drug field was matched against multiple drugs. For example, one drug field names "tipranavir + ritonavir coadm" as the suspect drug. The FDA has never approved a combination drug that has "Tipranavir" and "Ritonavir" as its ingredients. However, it has approved both drugs individually. The matching algorithm therefore splits this drug field up in two drug fields: "Tipranavir" and "Ritonavir". On the other hand, the drug field "abacavir sulphate+lamivudine+zidovudine" contains three drugs that were individually approved by the FDA but also in combination with one another. In such an instance, the standardized drug name is the combination drug - Trizivir in this case.

Table 3.1 lists three panels of descriptive statistics for AERS. Panel A describes the counts and percentages of the main variables in the data. The data contains more than 3 million cases and some cases are characterized by multiple reports of ADRs. The majority of patients that experienced an ADR that was reported to the FDA was female (58%) and the mean age of the patients for which a report was filed was 52. Patients are the most common reporters, while healthcare professionals including physicians and pharmacists jointly report about as often as patients. Some health outcomes are extremely severe, but most of the reports are accounted for by the least severe cases such as "Hospitalization" and "Other". Finally, Table 3.1 shows how much missing data the sample is characterized by. It shows that Age, Reporter Type, and Health Outcome are the most common fields to have missing data.

Panel B and panel C show two contingency tables based on the AERS data. The contingency table in panel B shows the multivariate frequency distribution of gender by health outcome. Although the distribution of health outcomes look very similar for men and women, reports for female patients contain disproportionally high numbers of "Other" health outcomes. It is not clear where this stark difference comes from. In panel C I show a similar contingency table but rather than subsetting by gender, I subset by reporter type. It becomes clear from the table that when the health outcome is missing, chances are that the reporter is the patients him or herself. Patients dominate the reporter type count, but physicians more often provide information about the health outcome of the ADR. Since there is a category "Other", this difference seems to result from precision in filling out the report rather than a difference in expertise. Finally, panel C shows that if the reporter type is missing, the health outcome is unlikely to be missing too.

Table 3.1: Descriptive Statistics for AERS

Database Size	
Unique Cases/Unique Reports	3,257,696/4,184,707
Demographic Data	
Female	2,414,006 (57.7%)
Male	1,474,467 (35.2%)
Gender Missing	296,234 (7.1%)
Mean Age (SE)	52.29 (0.012)
Age Missing	1,436,473 (34.3%)
Reporter Type	
Consumer	1,296,218 (31%)
Physician	758,972 (18.1%)
Pharmacist	188,199 (4.5%)
Other Health Professional	453,936 (10.8%)
Reporter Role Missing	1,308,949 (31.3%)
Health Outcome	
Death	438,511 (8.9%)
Life-Threatening	170,868 (3.5%)
Disability	164,125 (3.3%)
Hospitalization	1,225,324 (24.9%)
Other	1,664,299 (33.8%)
Health Outcome Missing	1,255,407 (25.5%)
Other Missing Data	
Unmatched Drug Names (PS)	109,250 (2.6%)
Unmatched ATC Codes (PS)	119,972 (2.9%)

Panel A: Observations by Category

	Male	Female	N/A
Death	209,445	185,788	43,278
Life-Threatening	78,052	88,119	4,697
Disability	64,295	94,916	4,914
Hospitalization	509,151	681,533	34,640
Other	579,390	974,961	109,948
N/A	349,320	786,039	120,048

Panel B: Contingency Table – Gender by Health Outcome

	Consumer	Physician	Pharmacist	Other Health Professional	N/A
Death	68,692	128,421	19,923	62,240	137,384
Life-Threatening	22,743	37,224	14,747	20,841	72,754
Disability	24,146	34,676	4,214	13,661	55,519
Hospitalization	237,299	281,474	75,187	161,259	417,114
Other	357,906	320,261	67,309	177,593	611,713
N/A	700,889	138,524	42,006	110,130	257,386

Panel C: Contingency Table – Type by Health Outcome

Note: This table presents descriptive statistics for the demographic variables that are found in AERS reports. Each report is linked to a case and one case may have links to multiple reports if a patient experienced multiple adverse events. The demographic variables are counted per report rather than per case. Also, the baseline for computing the percentages of unmatched drug names and unmatched ATC codes is the total number of drug fields – which is similar to the total number of unique reports. The count of unmatched drug names or unmatched ATCs includes drug fields that are no drugs or drug fields that are ambiguous.

3.2.1.2 NDAs

In order to match all free text fields in AERS to a standardized name of a drug approved by the FDA, I constructed a list of all drugs approved for marketing in the U.S. market since 1950. Although it seems like a trivial task to collect this data, it is not. Essentially, I have built the dataset from four main resources: Drugs@FDA[12], a 1989 Center for Drug Evaluation and Research publication, NDA Pipeline, and the 1999 - 2011 Drug and Biologic Approval Reports[13]. Drugs@FDA is a database that is freely available from the FDA website and contains information on approved drugs, including their New Drug Application (NDA) number, their trade and generic names, their approval date, their sponsor, and their histories of regulatory actions associated with the specific drug. Unfortunately, the Drugs@FDA database does not contain *all* drugs approved by the FDA and some of the more problematic cases (those that were withdrawn from the U.S. market) are missing. Therefore, I developed a strategy to compare the drugs from Drugs@FDA with the list of approved drugs from at least one other source in each year since 1950. The first comparator source consulted was a publication of the FDA (Center for Drug Evaluation and Research, Office of Management) issued in 1989. This publication lists all drugs approved by the FDA from January 1950 to December 1989. The second source was the NDA pipeline publication. NDA pipeline is a yearly publication by FDC reports (Chevy Chase, MD) and is based on the Pink Sheet, a trade journal that is also published by FDC reports. NDA pipeline lists all drugs approved by the FDA in a given year. Through the university library, I was able to access the 1984, 1986 - 1989, 1991 - 1992, and 1994 - 1998 editions of the NDA pipeline. Finally, I compared the Drugs@FDA data with the 1999 - 2011 Drug and Biologic Approval Reports found on the FDA website. This leaves 1990 and 1993

[12] 1.usa.gov/1pCpJaZ
[13] 1.usa.gov/1vr3lHr

uncovered. For 1990, I manually compared the list of drugs to the drug list in Kaitin et al. (1994), while for the NCEs approved in 1993 I manually compared the drug list to the drug list in Kaitin et al. (1994). The final dataset contains 1,341 unique prescription drugs making the coverage higher and the number of false positives lower than the leading list of prescription drugs (Carpenter et al. 2010; Carpenter 2010).

3.2.1.3 ATC

After matching the free text against standardized drug names I linked each drug name to an Anatomical Therapeutic Chemical code. The Anatomical Therapeutic Chemical (ATC) classification system, initiated and maintained by the WHO Collaborating Centre for Drug Statistics Methodology, organizes active substances found in drugs into different groups according to the organ on which they act and their therapeutic, pharmacological and chemical properties. The systems is hierarchically organized and consists of 5 levels. The first level comprises fourteen groups and indicates the anatomical main group, with therapeutic subgroups (2nd level), pharmacological subgroups (3rd level), chemical subgroups (4th level), and lastly the chemical substance further demarcating the similarities and differences between active substances.[14] The complete classification of Insulin Lispro in Table 3.2 illustrates the structure of the system.

Table 3.2: Levels in ATC

Level	Description
1	A - Alimentary tract and metabolism - anatomical main group
2	A10 - Drugs used in diabetes - therapeutic subgroup
3	A10A - Insulins and analogues - pharmacological subgroup
4	A10AB - Insulins and analogues for injection, fast-acting - chemical subgroup
5	A10AB04 - Insulin Lispro - chemical substance

[14] http://www.whocc.no/atc/structure_and_principles/

A total of 97% of all fields could be linked to an ATC code. The reason for the lower number of matches of ATC codes versus U.S. approved drugs is that the ATC classification system was first initiated in the early 1980s and some older drugs do not have ATC codes. Moreover, since the assignment of ATC codes lags behind the approval of prescription drugs, newly approved drugs often do not yet have an ATC code. The final reason why not every drug can be found in ATC is that in order for a drug to be included, the WHO requires an application, typically from the manufacturer of the drug.

In order to define a relational structure of drugs, I move from level 5 (at which each drug in the dataset is identified) to level 2. Timmermans and Buchbinder (2012) in their analysis of newborn screening show the salience of disease categories. Diseases are the most prominent categories along which understanding of conditions and treatments is shaped. For the current analysis, this implies that if the main hypothesis is confirmed, patients and physicians act upon communications about the withdrawal of a drug that are used to treat the same disease as the disease for which the patient is being treated. That is, meaning is extracted from the communication about the withdrawal and and put into action by reporting an ADR for drugs in the same disease class as the withdrawn drug. This is not to say that the structure of the ATC classification system is known by physicians and patients, but rather that the ATC classification system is meaningful in that it resonates with the understanding of drugs.

By linking prescription drugs that treat the same disease, a network of drugs can be created. This network is shown in Figure 3.2. Drugs are tied to one another if they treat the same disease and since some drugs treat multiple diseases, various clusters are connected through one or multiple multi purpose drugs. Figure 2 also shows the names of the drugs that were withdrawn. In defining drug *i*'s neighbors, it must be noted that I exclude drugs that are in the same chemical subclass. Although unlikely, there is the potential that the

chemical group is associated with some unobserved confounder that causes the number of reports for the group to go up. In order to rule this option out as an alternative explanation, I limited the sample of neighboring drugs to those that treat the same disease but are in a different chemical group.

3.2.1.4 Drug Withdrawals

The focus is on drugs recalled or withdrawn by the FDA or the manufacturer[15] because of safety reasons. The list used in this chapter is constructed by going through records of regulatory actions taken by the FDA and by identifying instances in which a drug is withdrawn. The records of regulatory actions are accessible through Medwatch and can be found through the FDA website. I code the day at which the FDA communicated (in an FDA talk chapter or a Public Health Advisory) about the withdrawal as the day of the communication. In some cases, the manufacturer sent out a "Dear healthcare professional" letter before the FDA communicated about the withdrawal, but the gap was never more than a day and given that US mail takes at least a day to be received by the recipient, it will not interfere with the exposure of the healthcare professional to the new information contained in the communication. Figure 3.3 contains an overview of the drugs that were withdrawn between 1997 and 2013.

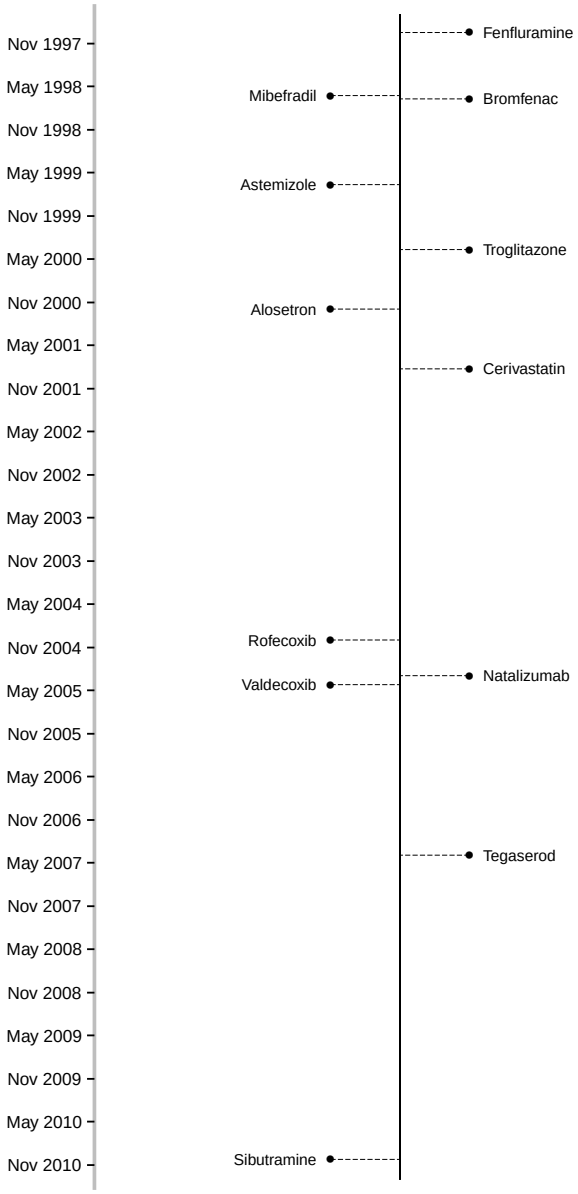
3.2.2 Empirical Strategy

The main outcome variable of interest in this chapter is the relative increase in reporting, rather than the absolute increase. While identifying the absolute

[15]Despite the desirability for clear and transparent categorization of safety issues, many drug safety communications leave room for multiple interpretations. For example, the difference between a recall and a withdrawal is not always clear (see this discussion on a consumer advocacy website). Perhaps as a result of this ambiguity and in an attempt to reduce it, the FDA recently revised its Regulatory Procedures Manual (RPM) and updated its definition of withdrawals and recalls.

Figure 3.2: The Network of Drugs that Treat the Same Disease

Figure 3.3: Timeline of drug withdrawals



Note: For each withdrawn drug, I looked up the precise date at which the FDA communicated the withdrawal. This is also the date that will be used in the analyses that follow.

increase in reporting is interesting, it is essentially meaningless for the detection of a signal, *if all other drugs also experience an increase*. However, by only studying the relative increase in reporting of a specific drug a signal may be detected that is solely due to the decrease in *one or multiple other drugs* that are aggregated in the denominator. Therefore, the analyses in the chapter report both the effects on the absolute number of reports and the effects on the relative number of reports.

As noted earlier, I first regress the reporting ratios for the withdrawn drug and for drugs that treat the same disease as the drug that is withdrawn on the effect of the treatment variable – being exposed to the safety communication (1) versus not being exposed to the safety communication (0). I employ a Poisson regression with the total number of reports filed per day as the exposure variable, the number of reports filed for the drugs of interest as the dependent variable, and the treatment variable as the main predictor. I also include dummy variables for the day and month and I include a variable that captures the number of days from the event.

To estimate the causal effect of communications of withdrawals on the reporting ratios of drugs, I employ the timing of these communications as the continuous forcing variable X while the date of the communication is used as the cutoff point that defines the treatment and control group (Davis 2008). The control group includes daily counts of reports sent to the FDA prior to the communication of the withdrawal and the treatment group includes daily counts for the same set of drugs after the communication was sent out.

$$T_i = \begin{cases} 1, & \text{if } x \geq c. \\ 0, & \text{if } x < c. \end{cases} \tag{3.1}$$

The actual modeling of the data depends on the size of the window around

the discontinuity (for which values of X do we drop data points from our sample?) and the statistical model that we use to estimate the treatment effect (Green et al. 2009). We follow Green et al. (2009) and use local regressions in combination with the Imbens-Kalyanaraman estimate in order to obtain the optimal bandwidth (Imbens and Kalyanaraman 2011). To fit the local regression, the data is sampled to include only observations within a bandwidth around the cutoff point. Moreover, observations that are closer to the cutoff are weighted more heavily. Defining a bandwidth poses a trade-off: a narrow bandwidth minimizes the chance of bias in the estimated treatment effect, but it also reduces the number of observations and increases the uncertainty in the estimated coefficients (Green et al. 2009). While there are various strategies to estimate the optimal bandwidth, the algorithm in Imbens and Kalyanaraman (2011) has been shown to outperform alternatives (Green et al. 2009). Therefore, following Imbens and Kalyanaraman (2011), I use a triangular kernel to weigh the observations closer to the cutoff more heavily so that the weight assigned to each observation increases linearly from the boundaries of the bandwidth to the cutoff point[16]

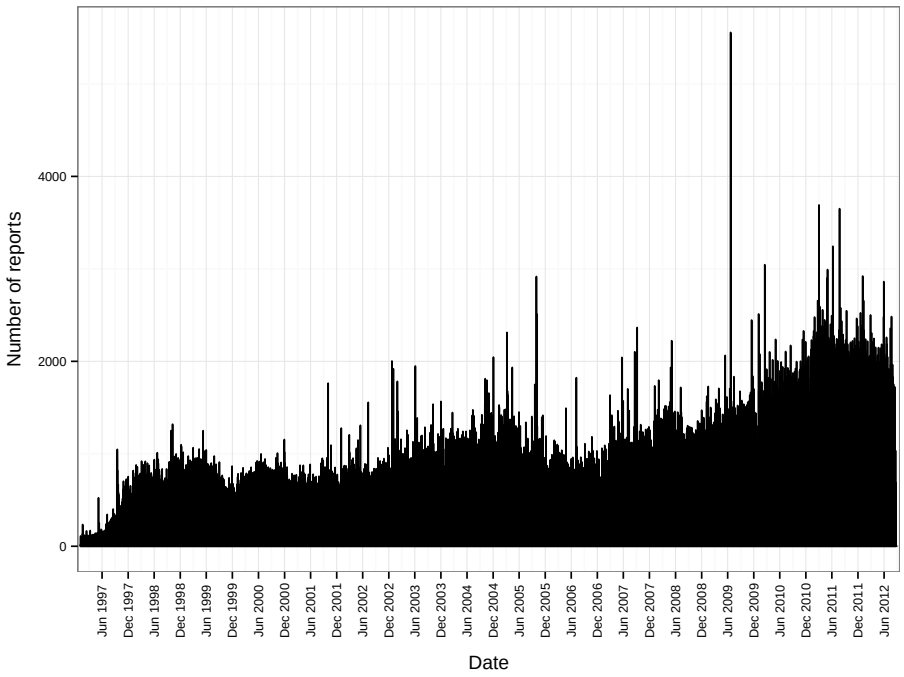
3.2.3 Descriptive Statistics

Before turning to the analyses, I briefly review a few descriptive statistics for the data analyzed in this chapter. AERS is a rich and complex dataset and I therefore think that it is useful to share some of its characterizing properties. Figure

[16] One of the most salient choices when analyzing data in a RD design is finding an appropriate regression specification. There are three commonly used approaches, all of which have their pros and cons. The first is to simply fit a parametric linear regression. The problem whit this approach is that there is no reason to believe that the true relationship is linear (Lee and Lemieux 2010). The second commonly used approach is to fit a non-parametric regression with polynomials for X. While this model allows for more flexibility than the linear regression, it provides estimates of the function at all levels of X. RD designs, in contrast, are built on the logic that causal effects can be identified *close to the cutoff* (Lee and Lemieux 2010). A third approach is to use non-parametric kernel regressions. Although this approach captures local estimates of Y it runs into problems close to the cutoff, because to estimate of the local value of Y on one side of the cutoff one cannot be used to estimate the local value of Y on the other side of the cutoff.

3.4 shows the number of reports filed per day. It shows that over time the daily number of reports filed with the FDA have increased substantially. This also explains the needs for sophisticated algorithms that continuously monitor the data for signals that indicate the potential for unsafe drugs. A second interesting feature is that there is considerable variance in the number of reports filed from day to day. Some of this variation is seasonal or related to the day of the week, but much of the variation is left unexplained.

Figure 3.4: Number of reports filed to AERS per day



Note: This graph plots the number of reports filed by patients, physicians, and other healthcare providers on a daily basis. All reports are included in the graph, including those that have missing data in one or multiple of the demographic variables.

Table 3.3 shows the descriptive statistics of the reporting intervals per day. The reporting intervals refer to the difference in the number of days between the occurrence of the ADR and the day that the patient or physician reported the ADR. The table shows that in some instances it takes a long time before an

ADR is reported. Moreover, weekends truly stand out as indicated by the much shorter intervals.

Table 3.3: Event - report interval per day

Day	Min	Max	Median	Mean	Std. Err.
Monday	0	1533	64	207.08	0.42
Tuesday	0	1533	63	199.51	0.42
Wednesday	0	1533	60	195.33	0.43
Thursday	0	1533	59	198.79	0.45
Friday	0	1533	58	196.71	0.46
Saturday	0	1533	30	141.75	1.35
Sunday	0	1533	24	132.95	1.55

Table 3.4 tabulates the average number of reports sent to the FDA per day. The table reveals substantial variation in the number of reports sent throughout the course of a week. The pattern clearly shows that Saturday and Sunday are off days and that people are most active early in the week in reporting ADRs. The standard errors are fairly low and, while not shown in the table, the pattern of decline in the number of reports throughout the week is stable over time.

Table 3.4: Reports per day

Day	Mean	Std. Err.
Monday	1081.83	20.66
Tuesday	1077.57	17.74
Wednesday	999.79	15.85
Thursday	952.1	16.41
Friday	882.47	15.59
Saturday	68.05	1.84
Sunday	51.91	1.53

Note: This table shows the average number of reports that are filed per day of the week. The means are calculated over the pooled data from 1997 to 2012. While the averages in recent years are certainly higher than the pooled averages shown in the table, the weekly trends are essentially the same.

3.3 Results

The regression estimates for the first set of regressions are shown in Table 3.5. An estimate of the coefficient of 2.54 implies that the number of reports filed for a drug that was withdrawn from the market were, on average, 251% higher in the post-removal period than in the pre-removal period, in proportion to the daily reporting rates. The table shows that the estimate is quite stable, even if the window around the announcement of a recall is extended to 42 days before and 42 days after the communication of the withdrawal.

Table 3.5: Point estimate for treatment effect on focal drug using Poisson regression

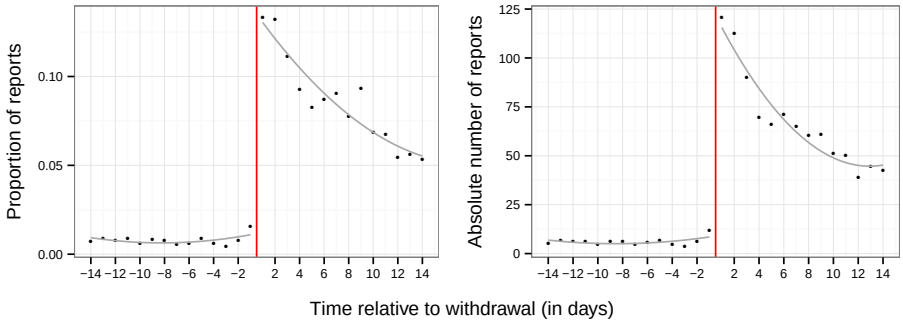
Interval width	Point estimate	Lower CB	Upper CB
$\pm$ 7 days interval	2.54	2.43	2.64
$\pm$ 14 days interval	2.55	2.47	2.64
$\pm$ 21 days interval	2.57	2.50	2.64
$\pm$ 28 days interval	2.28	2.22	2.34
$\pm$ 35 days interval	2.19	2.13	2.24
$\pm$ 42 days interval	2.16	2.11	2.21

Note: This table shows the point estimates and the upper and lower bounds for the 95% confidence interval of the estimate. The regressions include three sets of variables: day dummies, month dummies, and a variable indicating the number of days from the event.

Figure 3.5 shows the graphic representation of the RD analysis. The effect is large; the effect at the discontinuity for the reporting ratio is 0.12 and is statistically significant at the 0.001 level. For the absolute number of reports, the effect is also large – 107.31 – and statistically significant at the 0.01 level. This implies that – if the assumptions of the model hold – the direct increase in both the relative ratio and the absolute number is more than 1000%. So, besides its statistical significance, the effect of a communication of a withdrawal seems to be economically salient too. These results conform hypothesis 1.

Figure 3.6 shows the regression lines for two control variables. The first variable, which is found in the upper left panel of the graph shows the interval

Figure 3.5: Regression Discontinuity graphs



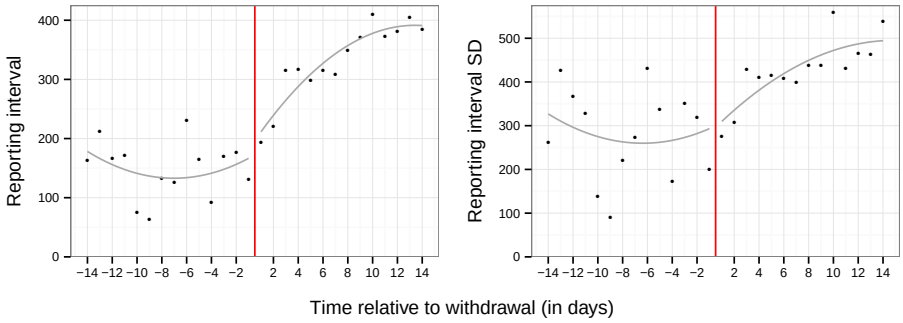
Note: The left panel plots the RD graph with "days from event" as the forcing variable and the daily proportion of reports that indicate the withdrawn drug as the primary suspect as the response variable. The graph in the right panel shows the alternative response variable: the absolute number of daily reports.

between experiencing and ADR and reporting it. One of the important questions is obviously whether an increase in the number of reports comes from (1) people that would otherwise not have reported or from (2) people that would have reported at a later day. One way of testing which of the explanations holds is by studying the effect-report interval. If, at the cutoff (so at the day of the safety communication), there is also a discontinuous change in the effect-report interval (there could be a strong decrease if people who would have reported in at a later stage would start reporting at the data of the communication), it is likely that people that would have reported at a later day but decided to report today are the cause of the increase. If there is no discontinuous change at the cutoff, it is likely that the increase comes from people that would have otherwise not reported. This is precisely what the graph shows.

The graph also provides a check for the argument that the number of ADRs suddenly increased in a period before the withdrawal. Let's assume that n days prior to withdrawal of the focal drug there was a heat wave causing many ADRs. If all these ADRs were reported at the day of the cutoff, one would expect an upward discontinuity in the regression line if the heatwave was much

earlier than the cutoff minus the mean of the interval right before the communication. The opposite would be observed for the case in which that increase happened more recently. In case that the ADRs happened right at the mean interval, one would expect a discontinuous reduction right at the cutoff in the left hand panel of the graph.

Figure 3.6: RD controls graphs



Note: Each panel represents one control variable. The left panel plots the daily reporting interval (the number of days between the the event date and the reporting date), the right panel plots the standard deviation of the daily reporting interval.

Although I have not explicitly stated any hypotheses regarding my expectations about the differences between patients and physicians in terms of how their reporting ratios for the drug that is withdrawn change as a response to the recall, table 3.6 shows the results from such a comparative analysis. In particular, the table presents a comparison of the effect of the withdrawal for the focal drug for healthcare providers – including physicians and other health care providers – and patients separately. Although the effect is somewhat larger for patients, both groups are characterized by a significant increase. Please note that the estimates shown in table 3.6 are generated from AERS data on *all* reports, not just the reports for which reporter type is non-missing. To fill the data for the reports on which the reporter type is missing, I build a model that uses a battery of other variables in the data to predict whether a report is likely

to be submitted by a physician or a patient. Then, for each report with missing data, I predict the likelihood that that report was submitted by a patient versus a physician and fill the missing cells accordingly. The same strategy was used for the analysis about the differences between patients and physicians in terms of their reporting behavior for drugs that treat the same disease as the drug that is withdrawn. Although the estimates (mostly in terms of their precision) differ between the analyses conducted for the subset excluding missing data and the full sample including the inferred data, the interpretation is similar (same direction, both statistically significant).

Table 3.6: RD Estimates

	Coefficient	SE	Z-score	P-value
Healthcare Provider – ratio	0.03	0.01	2.17	0.03
Healthcare Provider – absolute	30.47	10.74	2.83	0.00
Patient – ratio	0.07	0.03	2.42	0.02
Patient – absolute	70.24	25.26	2.78	0.01

Note: This table shows the point estimates and the upper and lower bounds for the 95% confidence interval of the estimate. They show that the increase in the relative and absolute reporting for "related drugs" is significantly different from 0 for healthcare providers and patients.

I now turn to the analyses for drugs that treat the same disease. In line with the analyses for the focal drug, I first estimate a Poisson regression. The regression estimates for these regressions are shown in Table 3.7. The coefficient for a 14 day window is 42 which implies that the number of reports filed for a drug that was withdrawn from the market grew by 42% as a result of the safety communication. As expected, the effect is substantially lower than the effect for the withdrawn drug. The table also shows that the estimate is decreasing as the window around the announcement of the recall grows.

Figure 3.7 shows the graphic representation of the RD analysis for drugs that treat the same disease. Similar to the results from the Poisson regressions, the analyses show that the effect is positive and significant; the effect at the discontinuity for the reporting ratio is 0.0035, and is statistically significant at

Table 3.7: Point estimate for treatment effect on related drugs using Poisson regression

Interval width	Point estimate	Lower CB	Upper CB
$\pm$ 7 days interval	0.42	0.36	0.48
$\pm$ 14 days interval	0.33	0.28	0.37
$\pm$ 21 days interval	0.22	0.18	0.26
$\pm$ 28 days interval	0.07	0.03	0.10
$\pm$ 35 days interval	0.13	0.10	0.16
$\pm$ 42 days interval	0.15	0.12	0.18

Note: This table shows the point estimates and the upper and lower bounds for the 95% confidence interval of the estimate. The regressions include three sets of variables: day dummies, month dummies, and a variable indicating the number of days from the event. The "related drugs" include those drugs that treat the same disease, but are categorized in a different chemical subclass.

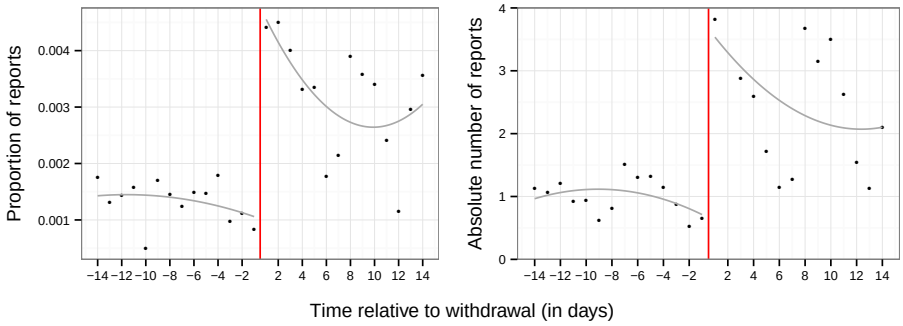
the 0.05 level. For the absolute number of reports, the effect is 2.07 and statistically significant at the 0.01 level[17] Although the estimates for the neighboring drugs show much smaller effect sizes they still show that the increase in both the relative ratio and the absolute number is more than 200%. So, besides its statistical significance, the effect of a communication of a withdrawal seems to be economically salient too. These results conform hypothesis 2.

Similar to the control variables for the withdrawn drugs, the control variables for the neighboring drugs are all continuous at the cutoff.

In the final analysis presented here, I compare the effect of the withdrawal for the neighboring drugs for healthcare providers – including physicians and other health care providers – and patients separately. The hypothesis stated that the effect for healthcare providers is expected to be more pronounced because they differ from patients in terms of their health literacy and because their formal training has prepared them to observe and recognize relations between drugs. The table shows that the effect for physicians is positive and significant, but that the effect for patients is not significantly different from zero. Moreover,

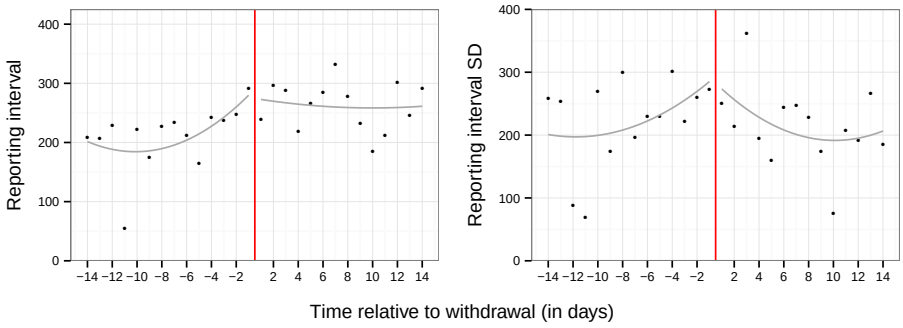
[17]Since there is considerable variance around the regression line, I also employed another empirical strategy that aims to test whether the observed effect is the result of random variation. Essentially, I simulated 100 placebo events and checked how the effects lined up. They suggest that the effect shown in table 3.7 is exceptional and is unlikely to be the result of random variation only. More details on the estimations can be found in Appendix B.

Figure 3.7: Regression Discontinuity graphs



Note: The left panel plots the RD graph with "days from event" as the forcing variable and the daily proportion of reports that indicate the withdrawn drug as the primary suspect as the response variable. The graph in the right panel shows the alternative response variable: the absolute number of daily reports.

Figure 3.8: RD controls graphs



Note: Each panel represents one control variable. The left panel plots the daily reporting interval (the number of days between the the event date and the reporting date), the right panel plots the standard deviation of the daily reporting interval.

the two estimates are significantly different at the 0.05 level. This finding is consistent with the argument that physicians use the relational structure between drugs to guide their ADR reporting behavior while patients do not. These results confirm hypothesis 3.

Finally, to understand whether the increased reporting of prescription drugs that treat the same disease as the drug that is withdrawn leads to a slowdown in regulatory action, I identified all significant increases in the reporting of an

Table 3.8: RD Estimates

	Coefficient	SE	Z-score	P-value
Healthcare Provider - ratio	0.01	0.00	2.77	0.01
Healthcare Provider - absolute	2.77	0.56	4.95	0.00
Patient - ratio	0.00	0.00	1.46	0.14
Patient - absolute	0.40	0.59	0.68	0.50

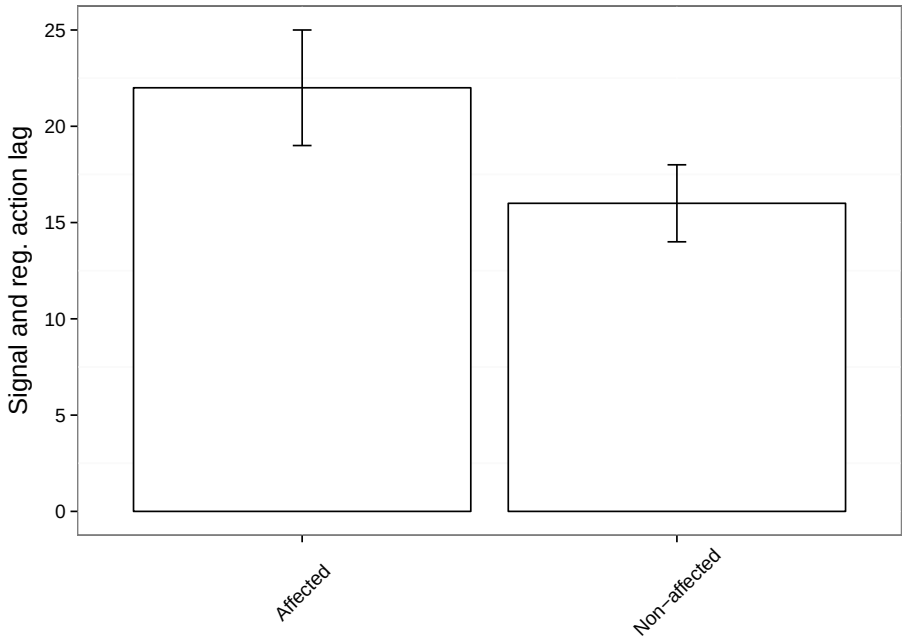
Note: This table shows the point estimates and the upper and lower bounds for the 95% confidence interval of the estimate. They show that the increase in the relative and absolute reporting for "related drugs" is significantly different from 0 for healthcare providers. The effect is not significant for patients.

ADR for a prescription drug and the regulatory actions taken by the FDA that are associated with those signals. I then compared those signals that were observed in a window of six months around the time of the withdrawal (three months before and three months after) and to the signals in a randomly chosen window of six months that had no drug withdrawn in it. As shown in Figure 3.9 the lags in the affected period are larger than the lag in the non-affected period. Although this finding is somewhat speculative, it suggests that the FDA is delayed in its regulatory decisions making during periods in which noise enters the AERS.

3.4 Discussion and conclusion

The results presented in this chapter are consistent with the idea that communications of regulators about unsafe drugs cause an increase in the number of reports about Adverse Drug Reactions filed to the FDA. I have shown that the number of ADR reports for a specific drug filed into AERS increases instantaneously after the announcement of a withdrawal of a prescription drug. The increase occurs both for the drug that was originally withdrawn and for drugs that treat the same disease as the drug that was originally withdrawn. The increase in the number of reports filed for drugs that treat the same disease is ac-

Figure 3.9: The effect of noise on the speed of regulatory action



Note: The left bar in the graph represents the drugs that were affected by an increase in reporting as a result of the withdrawal of a prescription drug while the right bar represents unaffected drugs. The bars and the error bars indicate that the lag between signal and regulatory action for the affected drugs is significantly larger than the lag for the unaffected drugs.

counted for by additional reports from physicians, not patients. Moreover, the analyses presented in the chapter indicate that this increase is caused by additional reports filed by those patients and physicians who would have otherwise not reported. I employ several tests to show that the results in the chapter are inconsistent with the idea that the increase in the number of reports is caused by an increase in the number of ADRs that occurred.

These findings have two major substantive implications. First, the large effect on the number of complaints filed for the drug that was withdrawn and the drugs that treat the same disease could mask other signals in the data that identify unsafe drugs. If the large number of additional reports are included

in the denominator when calculating the disproportionality of other drugs in the set of complaints, large effects may seem small and could therefore go undetected. It is not unlikely that this masking effect could seriously delay the detection of some drugs that are later found to be unsafe. A second major implication of the results presented in this chapter is that the increase of 200% in the number of reports filed for drugs that treat the same disease should not be interpreted as an increase in the number of ADRs. The increase is caused by physicians who respond to the release of new information rather than by additional ADRs. By accounting for high impact events, such as the communication of new information about drug safety, the FDA can filter out additional reports, thereby removing ambiguity in the data.

More broadly, the findings presented in this chapter are especially salient given the prominence of Evidence Based Medicine (EBM) in the medical discipline. EBM has been a dominant force in guiding healthcare practices over the past decades (Timmermans and Berg 2003; Timmermans and Mauck 2005) and can be seen as a set of guidelines and hierarchies that guide healthcare professionals in making decisions about how to improve a patients' health. EBM is built on the notion that – given the vast and increasing body of evidence – best practices can be identified and outcomes in healthcare can be optimized. My research suggests that the making of evidence is also very much a social process and that social processes should be accounted for when identifying best practices.

This chapter also has implications for research in cognitive sociology and organizational research. First, the findings confirm prior research that shows that negative attention directed at an individual, organization, or object results in repercussions through the expression of lower evaluations. The process by which this happens, however, is different from what was previously observed in other contexts. Rather than a fixed audience that acts upon an actor's status

or stigma by providing a lower evaluation, this research shows that audiences of evaluators revisit past experiences and reinterpret them based on the newly released information. In doing so, the audience grows which raises questions on how the evaluations of these "new" audience members should be interpreted. A second theoretical contribution of this chapter is that it shows that relational structures may guide the behavior of audiences that are presented with new information about the quality of a product. It demonstrates that negative attention for one object may contaminate the status and reputation of other objects *if those objects are categorically related*. Given the increased salience of all kinds of relational structures in social life (Bowker and Star 1999; Timmermans and Berg 2003), the findings from my research suggest that individuals and organizations should take seriously their position in a multitude of relational structures and consider the status and behavior of their "relational alters". Finally, this research indicates that differences between evaluators in terms of the cognitive schemata that inspire their behavior lead to heterogeneity of the patterns in the data that they produce. Organizational research concerned with evaluation processes is fairly homogeneous in terms of the research design that it adopts. The most common strategy is to take a pool of evaluations and explain the variance in these evaluations by accounting for characteristics *of those who are evaluated*. However, my research shows that audiences can be partitioned into subsets of evaluators who adhere to different evaluation processes. Individuals and organizations concerned with obtaining positive evaluations can exploit these differences between the subsets of audiences members.

This research also draws attention to the question of how new information induces actors to alter their behavior. Although the data used in this chapter does not allow me to identify the mechanism by which the release of new information induces audience members to contribute an evaluation, there are essentially three explanations. First, the release of new information may legit-

imize the contribution of an evaluation. If patients and physicians were aware of the potential causal link between drug and ADR and the option to report to into AERS, but felt that their report was not warranted, the official statement by the FDA may have legitimized their claim. A second explanation for the increase in the number of reports may stem from awareness. Several studies have shown that more and more patients and physicians are aware of the possibility to report into AERS. Although much of this increased awareness may come from effective campaigns designed by the FDA to inform patients and physicians about how they can contribute to increasing public health, safety communications that are highly publicized may also cause awareness to increase. The third explanation, which I term realization, was introduced earlier in this chapter and advances the idea that increased reporting is due to patients and physicians who revisited past experiences and realized that an ADR was caused by a prescription drug. There is merit in studying the different mechanisms by which new information induces patients and physicians to contribute. For example, a regulator interested in improving the quality of consumer contributed data may use this information to design strategies increase participation. Moreover, research on legitimation may benefit from a detailed analysis of how micro-processes cause individuals, organizations, objects, or practices to gain legitimacy. Currently, these micro-processes remain under exposed in organizational research on legitimacy.

Although this chapter did not explicitly test how patients and physicians learned about the recall of an unsafe drug, an interesting question is whether this happened through exposure to the media communication of the FDA or through social influence among patients and/or physicians. Similar to the parents of children diagnosed with autism in (Liu et al. 2010), patients and physicians may become aware of, realize, or legitimize the fact that they experienced a side effect through their peers. While I have not set up a formal test to an-

swer this question, the analyses indicate that the second explanation – social influence – is unlikely to be the sole mechanism to account for the increase. The regression discontinuity method used in this chapter suggests that the increase in reporting is instantaneous. If social influence were to be solely responsible for the increase one would expect a more slowly growing increase[18]

This chapter applied a social theory of cognition to a serious problem in the health care domain. In doing so, it showed how social theory can be used to advance problems in public health. It also provided a detailed account of how individuals attribute an effect to a cause and how accounting for the cognitive schemata that audiences use allows one to understand variation in evaluations.

[18]In an additional analysis not shown here, I proxied the cohesiveness of patients by counting the number of patient groups organized around the disease that the drug targeted. The hypothesis that I developed for the collection of this data was that diseases that have tightly organized patient groups would exhibit different patterns of increasing reports following the discontinuity. I therefore divided the drugs up into two groups; one group with a cohesive patient base and one group with a dispersed patient base. The analyses did not show differences in the patterns of response.

Appendix A Using AERS to Detect Unsafe Drugs

Regulators and scientists typically use various statistical techniques to detect disproportionality. A commonly used technique in the medical sciences to detect potential safety issues by using AERS is through a case/non-case methodology. In a case/non-case approach the researcher splits up the population of reports into two samples; one that contains all reports that name drug *i* as the potential cause of an adverse event (cases) and one that contains the complement of the reports that name drug *i* (non-cases). These two subsets can then be further partitioned into the reports that report adverse event *a* - for example arrhythmia - and the reports that report all other adverse events - all adverse events but arrhythmia. Table 1 shows how the reports in AERS can be divided up into subsets that allow researchers to detect disproportionality in reporting.

Table A3.9: Subsets in AERS

	Drug$_i$	**Drug**$_i^C$	**Total**
Adverse Event$_i$	A	B	A + B
Adverse Event$_i^C$	C	D	C + D
Total	A + C	B + D	A + B + C + D

A = number of reports where suspect drug is Drug$_i$ and adverse event is Adverse Event$_i$
B = number of reports where suspect drug is Drug$_i$ and adverse event is Adverse Event$_i^c$
C = number of reports where suspect drug is Drug$_i^c$ and adverse event is Adverse Event$_i$
D = number of reports where suspect drug is Drug$_i^c$ and adverse event is Adverse Event$_i^c$

Note: This cross table is used by medical scientists to compute disproportionality in the reports filed for a specific drug/adverse event combination.

Although there are various techniques (including frequentist and Bayesian approaches) that employ different formulas to capture the level of disproportionality, the basic intuition behind all of those measures is that they capture the proportionality of the occurrence of a specific adverse event in those being treated with drug *i* and the occurrence of that same event in those treated with other drugs. These techniques are non-parametric versions of a bivariate

logistic regression where the outcome variable is a dummy that equals 1 if a report reports the adverse event of interest and the explanatory variable is a dummy variable that equals 1 if the patient in the report is being treated with the drug of interest. Typically the distribution of adverse events for a given drugs is skewed with many reports coming in for few adverse events and only few reports for other adverse events. It becomes apparent that if the number of reports for a specific drug increases and the confidence bands around the estimate of the signal become less wide, the likelihood of detecting disproportionality goes up.

Appendix B Robustness Checks

In order to test whether the observed effect is not the result of another temporal effect that I did not theorize, I have simulated – for each event – 100 placebo events. In other words, I randomly selected 100 dates for each observed event date and I replicate the analyses presented in the chapter. Please note that it is unlikely that the withdrawal of a prescription drug is the only event to alter the reporting proportionality. Other events such as extensive media coverage for a drug, or a widely publicized lawsuit could also trigger strong effects on the reporting behavior of patients and physicians. Therefore, I expect there to be at least a few instances in which random selection of placebo dates yields results that show comparable effects on the the reporting.

Both for the effect on the focal drug, and for the effect on the neighboring drug, the findings of the placebo analyses indicate that in less then 5% of all placebo events, the increase in the reporting ratio is significantly different from 0. Moreover, for those placebo events that generate a significant effect on the reporting behavior of patients and physicians, none is characterized by an effect on reporting that is as large as the effect of the observed event.

Chapter 4

Scientific Evidence Production and Regulatory Action

This chapter explores the role of industry involvement in the production of scientific evidence. The analyses are based on the scientific publications discussing one or more of the 200 prescription drugs approved between 2000 and 2010 and data on the citations between those publications. The argument developed in this chapter holds that the structure in which scientific publications become embedded can be used to understand the debate about the safety of a prescription drug. The analyses conducted and presented in this chapter show that there is substantial variation in the degree to which pre-approval scientific evidence becomes integrated within the debate that follows the approval of a prescription drug and that this variation is associated with the cohesion among pre-approval publications. Finally, a series of regression models show that there is a substantive and significant association between the pre-approval cohesion of the scientific debate and the frequency and speed of regulatory action. The findings presented in

this chapter suggest that the unfolding of the scientific debate can account for the variation in the regulatory actions taken for prescription drugs.

4.1 Introduction

How is the direction and content of science affected by involvement of social actors affiliated with industry? This question has become increasingly prominent in the social sciences and has sparked a debate among some who argue that industry involvement makes academic research more strategic and technologically relevant and others who have argued that industry involvement distorts the development of new knowledge (Evans 2010). However, while there is an abundance of evidence that shows that the direction of science is altered as a result of industry involvement (Shane 2000; Washburn 2006; Etzkowitz 1998; Azoulay et al. 2009), studies that have elaborated the mechanisms by which this change occurs are scant. Moreover, few studies have explored the implications of change in the direction and content of science on its consumption. This chapter identifies a mechanism that can account for variation in the direction of scientific work about the safety and efficacy of prescription drugs and shows that the regulatory pathway that unfolds after approval of a prescription drug is adjusted by this variation.

In this chapter I build on data about all scientific publications that discuss at least one of the 200 prescription drugs approved by the FDA between 2000 and 2010. For each of these 200 drugs, I distinguish between studies that were published prior to approval of the prescription drug and those that were published after approval. The way in which publications about the safety and efficacy of prescription drugs are produced looks roughly as follows. The studies that are published prior to approval are sponsored by the pharmaceutical firm in an attempt to produce evidence that will get its new drug application approved for

marketing in the U.S.[1] The research design upon which the analyses in these studies are built is implemented by physicians who are not employed by the drug sponsor – they are called clinical investigators and are contracted to study the drug that the sponsor is aiming to market. However, rarely do these investigators have any influence on the design of the study and rarely do they have access to the raw data that the study produces (Baer et al. 2011; Suvarna 2012; Davidoff et al. 2001). And while about 70% of clinical investigators were affiliated with academic organizations in the early 1990s the proportion had decreased to about 35% in 2001 when independent hospitals, private practices and for-profit, dedicated clinical research sites replaced the academic sites as the dominant player in the market of clinical trials (Azoulay and Fishman 2006). The study implemented by the clinical investigator and the analyses produced by the drug sponsor are used by the FDA to evaluate the safety and efficacy of the new drug application. After approval of the new drug application, the newly marketed substance becomes available more widely and research is carried out by academic organizations without contracts with the sponsor of the drug.

This instantaneous shift from industry-directed science to science produced by a much more diverse set of actors provides an opportunity to study how industry science differs from normal science, how the two become integrated, and how the quality and content of industry science affects how it is consumed. I exploit this opportunity by analyzing all citations between scientific publi-

[1] Obviously, official communications issued by drug sponsors will contain somewhat different language for describing the function of clinical trials. For example, Otsuka – one of the leading Japanese pharmaceutical firms – states on its website that "Otsuka, as a company focused on innovation, recognizes that access to clinical trial data is valuable for the advancement of public health and science. The benefits and associated responsibilities of broadly available clinical trial data are considerable, including wider communication of the safety and efficacy data for our medicines, of how clinical trials are designed and conducted, and of the diseases for which we, and many others, are seeking to meet unmet medical needs. Otsuka has been sharing clinical trial data and results through clinical trial registries in public databases, such as the clinicaltrials.gov website and through publication in peer-reviewed journals and will continue these efforts." Please see bit.ly/1HdZ65l for more information.

cations in the discourses that discuss the 200 prescription drugs approved between 2000 and 2010 and I develop hypotheses about how the structure of these citation relationships emerges and how those same structures can be used to understand the regulatory process after approval. In doing so I use the image discussed in Latour (1987) about the construction of facts though the process of tying knowledge claims to objects (people, claims, devices, etc.) that support that specific claim. In particular, I study the way in which claims become embedded in a scientific support network and the implications that has for how consumers of science (including scientists themselves and regulators) make sense of a body of literature containing a multitude of competing claims.

In this chapter I will develop hypotheses about how the transition from an exclusive to an inclusive discourse can be used to understand the production of evidence. Throughout this chapter I will compare the pre- and post-approval debates and hypothesize about how internal structures are associated with integration. Moreover, I will show how these structures can be used to understand post-approval outcomes, using data on the regulatory actions taken by the FDA. First, I will show that the pre- and post approval debates are truly different. I do so by analyzing the content of the papers published before and after approval. Second, I study how the discussion network forms. Using Exponential Random Graph Models I explore the extent to which pre-approval publications become integrated in the post-approval debate and whether this is a function of the way in which the pre-approval debate is organized. Finally, I examine the association between variation in the way in which pre-approval debates are organized and the regulatory action that follows approval. In doing so, this chapter aims to understand the extent to which scientific trajectories are affected when the building of facts shifts from an industry controlled activity to an activity that is open to a much wider group of scientists.

To anticipate the main findings from this chapter: I show (1) that science

published prior approval pays less attention to the potential of ADRs than studies published after approval, (2) that pre-approval publications do not become integrated into the core of the debate if they are initially highly cohesive, and (3) that drugs characterized by high levels of network modularity in the pre-approval citation structure receive fewer label changes and it takes longer for those drugs to be targeted by larger numbers of regulatory actions.

4.1.1 Industry Science, Academic Science, and the Integration Discourses

The interaction between university and industry has been widely studied in the past couple of years and much of the research has demonstrated that from these interactions, new social and cultural institutions have emerged. An overview of this literature also makes clear that – although there is some variation in the understanding about the character of academic and industrial science and the roles that their interactions play in bringing them closer together or creating a stronger separation between the two (Evans 2010) – industry and science have moved closer together over the past decades. For example, Owen-Smith (2003) has argued that universities and firms are increasingly evaluated based on similar criteria. His work shows that universities have become increasingly active in commercial activities including patenting and that success in the scientific arena and success in the commercial arena are mutually dependent. Along similar lines, Vallas and Kleinman (2008) demonstrate that values of scientists in universities and firms have blended. They argue that academic scientists increasingly engage in activities to commercialize their research and that industry scientists are increasingly embracing the notion of "basic science". Yet another stream of research makes similar claims and argues that interests of government, universities and industry have become tightly coupled in a "triple helix"

(Etzkowitz 1983). In sum, evidence suggests that science and industry have grown similar over the past decades.

Despite this seeming convergence between industry and academic science, much of the research that describes how industry and science have become similar along a wide range of dimensions focuses only on the forms of production rather than on its content. For example, Owen-Smith (2003) shows that science has become increasingly active in patenting its inventions. He argues that the adoption of this commercial activity is evidence of the growing similarities between science and industry. However, his research does not demonstrate if and how the content of the patents taken out by industry versus science are similar or whether they differ substantively. A recent study by Evans (2010) shows that although collaborations between science and industry are increasingly common, differences in incentives still persist and that these differences translate in different types of science. His study shows that research conducted by teams that involve more scientists affiliated with industry are more likely to produce science that builds less on theory.

This change in the direction of science that is produced as a result of industry involvement is seen by some as a serious form of contamination of science (Proctor and Schiebinger 2008) – thereby adding a negative connotation to the change in direction. This negative stance towards the change in direction of science is likely to be related to studies that find that industry sometimes tends to devise strategies that prevent science from promoting policies that would harm the constituents of that industry (Proctor and Schiebinger 2008; Oreskes and Conway 2010; Abraham and Ballinger 2012). An example of a debate in science that is argued to have suffered from such strategies is the debate about tobacco carcinogenicity (Shwed 2015).

The context of scientific evidence production about the safety of prescription drugs provides a testing ground to explore how science produced by sci-

entists affiliated to different types of organizations varies. Given the fact that the incentives to publish critically about a prescription drug are different for university scientists than they are for scientists affiliated with the organization that sponsors the product, publications are expected to differ between these two contexts. In particular, drug sponsors are expected to be less likely than university researchers to publish about the potential ADRs associated with a prescription drug. Therefore, I hypothesize:

H1: Studies published prior to approval of a prescription drug are less likely to discuss the ADRs potentially associated with the prescription drug.

Embedded in the claims about the convergence of industry and academic science is the observation that industry and academic science have become more integrated and that this is achieved partly as a result of the increasing collaborative integration. That is, scientific collaborations increasingly include both academics and industry scientists (Evans 2010). What is missing from this conversation, though, is how science produced by industry and science produced by academics becomes integrated. If industry science and academic science emerge as two disconnected discourses, normal science in which knowledge is accumulated and builds on prior knowledge is hampered. To understand how industry science and academic science are integrated, I argue that the networks of citations between published papers from both sources can provide a telling picture. A good example of a study that builds on a similar logic is Shwed and Bearman (2010) who show that a network analysis of citations allows one to identify levels of consensus. In particular, building on Latour's actor-network theory (ANT) and Pinch and Bijker's (1984) idea of consensus as closure Shwed and Bearman (2010) show that networks of citations provide meaningful sociological objects that allows one to assess the level of consensus between arguments made within a scientific debate. Broadly speaking, the idea

that they propose and test is that modularity in the structure of the scientific citation network is likely to proxy for the level of consensus about the question that has sparked a scientific debate. That is, cohesive citation structures are associated with high levels of consensus while citation structures with loosely or unconnected regions are likely to be associated with disagreement.

Implicit in their account is that on average, when authors cite prior work, they do so because they agree with one or more of the claims made in that study. A citation between two papers also means that some of the content of the cited paper has been relevant for the current study. Translating these insights to the context of drug safety, I argue that industry science becomes integrated in a scientific debate if the publications produced by industry are unconstrained. In particular, if the science produced prior to approval of a prescription drug covers non-redundant questions, that science will be built upon by science approved after approval of the drug. To operationalize redundancy, I compute the modularity in the structure of citations between studies published prior to approval. Since cohesive pre-approval citation networks are less valuable for post-approval science, cohesive structures in the pre-approval citation network inhibit those papers from becoming embedded in the post approval citation network. Therefore, I hypothesize:

H2: Cohesive structures in the pre-approval citation network inhibit those papers from becoming embedded in the post approval citation network.

How does science inform practice? In their discussion on how the role of science has been discussed in the sociological literature, Collins and Evans (2002) point out the different waves of science studies, each of which contains a body of fairly consensual literature about the role and function of science in society. They argue that between these waves the consensual view on the meaning and value of scientific knowledge has shifted. The first wave of science studies –

which ended around the 1970s – largely agreed that good scientific training provided scholars with high levels of authority and decisiveness that often extended beyond their discipline. Consensus was that decision-making in contexts that involved science and technology should be top down and the bases of scientific argumentation were by no means subject of debate (Collins and Evans 2002). One of the first to question the broad and unambiguous authority of the sciences and the scientists that occupied the most prominent positions was Thomas Kuhn in "The Structure of Scientific Revolutions" (Kuhn 2012). His work gave rise to a second wave of science studies that re-conceptualized science as a social activity. In doing so, this body of literature started to question the bases of expertise and the authority linked to the label of expert. For example, comparisons between scientific expertise and other forms of expertise sparked debates about the boundaries of jurisdiction.

The third wave in the sociology of science described by Collins and Evans (2002) seeks to understand how science informs policy. In the case of the medical domain, a prominent task of science has been to evaluate policy and to protect citizens from unintended harm. That is, the medical scientific enterprise does not only develop the basic building blocks of the drugs that patients consume every day; medical science has also become an active evaluator of its own inventions. However, when building on science the regulator faces the problem of which claim is most valid and which claim should therefore be used to inform policy. Eyal (2013) shows that this uncertainty is often reduced by a priori classifing stakeholders as experts or non-experts, but that by doing so valuable information can be ignored. In translating the idea of stratification by expertise to the context of prescription drugs, one would – for example – only pay attention to those claims advanced by high status actors. An alternative to stratification by expertise is assessing the level of consensus among a body of claims. Kogut and Macpherson (2011), for example, studied economists and

has shown that agreement or consensus within a set of professionals affects the diffusion of policies over which they agree.

Developing strategies to make sense of a body of scientific claims is further complicated if the production of science is controled by one organization. If commercial parties control the production of science they may refrain from producing science that contradicts a claim made by the commercial party outside the realm of science. For example, various studies have shown that industry sponsored science is constrained in some sense and may not be in the interest of promoting public health (Proctor and Schiebinger 2008; Oreskes and Conway 2010; Abraham and Ballinger 2012). Industry science is therefore expected to be most valuable if it is not constrained by financial incentives and if it provides a set of non-redundant insights about the safety and efficacy of prescription drugs. In line with hypothesis 2, I argue that pre-approval citation structures that are highly cohesive provide less information about the safety and efficacy of a prescription drug and that drugs characterized by such evidence are therefore more likely to experience regulatory action after approval. Therefore, I hypothesize:

H3: Scientific debates characterized by high levels modularity (low cohesion) will be targeted by fewer regulatory actions and slower unfolding of regulatory action patterns.

4.2 Roadmap

The aim of the analyses in this chapter is to examine patterns by which scientific debates about prescription drugs unfold. In doing so, I pay particular attention to the ways in which the debate differs between the pre-approval stage and the post-approval stage and how and whether the evidence in the two stages

becomes integrated. Moreover, I also explore the impact of the way in which scientific evidence making unfolds on regulatory action. I do so by conducting several tests – each of which attempts to obtain a more crystallized description of the process of scientific evidence production.

In the first set of analyses, I will show how the debate about drug safety and efficacy differs between the pre- and the post-approval period. By analyzing the content of the scientific literature published about a prescription drug, I test whether the pre- and post-approval stages differ with respect to their attention for the potential of ADRs. I use several Logistic regression models in which I specify the likelihood of a paper discussing ADRs as a function of the stage in which it was published and a set of control variables.

In a second set of analyses, I move to the data about citations between papers in each of the 200 prescription drugs in the sample. The goal of the analyses is to model the process by which citations are created between papers in the scientific debate about a prescription drug. I do so by using Exponential Random Graph Models (ERGMs) in which I treat papers as the nodes in the network and the citations as directed network ties. A correctly specified ERGM allows me to recover the mechanisms that have guided the process of tie formation. The mechanism of interest can be described as the likelihood of a post-approval paper citing a pre-approval paper conditional on the cohesion within the pre-approval citation structure.

In the final set of analyses, I build on the insights generated by the ERGMs and specify sets of Poisson regressions and Cox Proportional Hazard models to capture the effect of pre-approval citation cohesion on the number and rate of regulatory actions targeting a prescription drug. The main goal of these analyses is to understand how the level of cohesion within the publications prior to approval has an impact on the regulatory trajectory through which a drug goes once the drug is approved.

Finally, I will summarize the outcomes of the statistical analyses and I will interpret the outcomes to describe how the production of scientific evidence unfolds.

4.3 Empirical Strategy

I conducted three sets of analyses. The first uses data at the paper level and models the likelihood that papers will explicitly and prominently address the ADRs associated with a prescription drug. Hence, the response variable is binary and equals 1 if the paper explicitly discusses the adverse effects associated with a prescription drug and 0 otherwise. I model this binary dependent variable as a function of the hypothesized effect of publication stage and a set of control variables by using a series of logistic regressions. I estimate models in which I pool the observations for all drugs, but I also fit multi-level models to account for the nested structure of a paper in a scientific debate about a specific prescription drug.

The second set of analyses focus on understanding the evolution of citation structures in scientific debates. I do so by employing Exponential Random Graph Models (ERGMs). ERGMs are widely used by researchers interested in different substantive contexts (Wimmer and Lewis 2010; Srivastava and Banaji 2011; Papachristos et al. 2013; McFarland et al. 2014) and they allow the researcher to test hypotheses about the social mechanisms that generated a set of (social) interactions between nodes[2] in a network. In contrast to widespread standard regression techniques such as the Ordinary Least Squares (OLS), Probit, or Logit regression[3], ERGMs accommodate various types of dependencies

[2]When I discuss general topics related to network analysis, I will used the terms nodes and edges to refer to the actors in a network and their relationships, respectively. When I discuss the empirical case of scientific debates, I will refer to the actors in the network as papers and to the relationships between papers as citations.

[3]Modeling network formation using standard regression techniques violates the independence assumption and – as a result – errors will be correlated with the structure of the network. Standard

between observations within a sample. In particular, rather than assuming that each dyad (i.e. a tie between two actors) is independently formed, ERGMs build on the idea that the probability of dyad formation is a process that is conditional on the structural properties of the network in which a particular dyad is embedded. Moreover, the ERGM framework allows the researcher to specify the precise structure of the dependencies between the actors in a network that are hypothesized to guide the process of dyad formation.

Similar to standard regression, the equation for an ERGM includes a dependent variable and a set of independent variables. The dependent variable is the observed network (i.e. the network that the researcher is interested in explaining) and this network represents one realization from the set of possible networks given the set of nodes. Since there is no variance in a single network, the ERGM exploits the hypothetical variation over the set of possible networks (Cranmer and Desmarais 2011). The goal then is to specify an equation for the ERGM that includes independent variables that capture theoretically motivated social processes that are likely to have caused network ties in the network of interest to be formed. Under the model specified by the researcher, each network in the set of possible networks can be assigned a probability. If the model specification assigns a high probability to the observed network (which is one network in the set of all possible networks) the model is argued to fit the data well. That is, to generate a good model fit the values of the parameters are set in such a way that the most probable network statistics computed on the networks are those that describe the observed network (Robins et al. 2007).

ERGMs can be fitted on relational data that comprise nodes, the attributes of those nodes, and alternative relationships between the nodes. The independent variables are the local network configurations (e.g. triangles, stars, etc.),

regression techniques also lack a design that can accommodate structural terms. For example, standard regression techniques are unable to model network formation as a function of triadic closure and will therefore – in the presence of a triadic effect – be characterized by an omitted variable that could potentially introduce bias in the model estimates.

exogenous dependencies (i.e. spatial distance between actors), and actor characteristics (e.g. age, size, etc.) that are hypothesized to influence the process of tie formation. The observed network is regarded as a random draw from the probability distribution of networks and the ERGM uses Markov Chain Monte Carlo (MCMC) maximum likelihood estimation to determine which parameters of the model create a probability distribution of networks in which the observed network has a high likelihood of being the outcome of the modeling process. The ERGM is specified as follows:

$$Pr(\boldsymbol{Y} = \boldsymbol{y}) = \left(\frac{1}{k}\right) exp\left[\sum_{A} \eta_A g_A(y)\right], \tag{4.1}$$

where the probability that a network $\boldsymbol{Y}$ generated by an ERGM equals the observed network $\boldsymbol{y}$ is a function of the sum of all network configurations A[4], where $g_A(y)$ is the network statistic corresponding to configuration A, equal to the frequency at which configuration A is observed in network $\boldsymbol{y}$[5], and η_A is the estimated parameter associated with $g_A(y)$. This parameter captures the importance of configuration A for generating the network that is modeled and is assumed to be homogeneous for the entire network. The normalizing constant k ensures that equation 4.1 is a proper probability distribution and is essentially equal to the sum over all possible networks given N nodes in the network.

While equation 4.1 has a global orientation because it models the probability of an entire network, one may also write an ERGM in a way that is more akin to a standard regression equation. By writing the ERGM in this alternative

[4] A configuration is a subgraph structure for which there is a parameter in the model. Examples of such configurations are mutual dyads (reciprocity), triangles (triadic closure), or stars (popularity).

[5] Each configuration refers to only a small set of nodes in the network. For example, if reciprocity is one of the hypothesized social mechanisms for dyad formation a configuration is included for each possible dyad in the network. Since estimating parameters for all these sets is infeasible, they are aggregated if the configurations they refer to are of the same type (e.g. reciprocity) and one thereby ignore the labels on the nodes. This constraint is termed the *homogeneity of isomorphic network configurations*. As a result of this constraint, the $g_A(y)$ term can be viewed as a count rather than as a binary term that indicates whether a configuration is present or not.

form one can directly probe the probability of a single tie (Y_{ij}) being formed conditional on the rest of the network (Y_c):

$$logit\ Pr(Y_{ij}|Y_{ij}^c) = \sum_A \eta_A \delta g_A(y) \tag{4.2}$$

In this representation of an ERGM, the parameter estimates η_A generated by fitting ERGMs can be expressed as conditional log-odds and can therefore be interpreted as the change in the log-odds of a tie being present in response to an increase in the variable of interest. Hence the parameter corresponding to a specific covariate (e.g. one of the included network structures) tells us if the log-odds of observing a network tie increase or decrease as the tie is embedded in the the network configuration for which the covariate value is computed.

Similar to standard regression techniques, the Goodness-Of-Fit (GOF) of an ERGM can be computed and one way of doing so is by using standard statistics including BIC and AIC. Besides these measures known from standard regression, ERGMs also allow for another class of GOF statistics. After estimating each of the models presented in this chapter, I inspect its GOF by comparing counts of network statistics[6] in the observed network to the counts of those same statistics in a set of simulated networks (these networks are simulated from the estimated coefficients in the ERGM). If the statistics for the observed network lie in the center of the distribution of statistics for the simulated networks (i.e. there is no significant difference between the observed and simulated network) the model is arguably a good representation of the data generating process.

Before describing the third set of analyses, I would like to note that the use of ERGMs in this chapter may seem inappropriate at first. Since the questions that I am trying to answer are clearly about the unfolding of social ties over

[6]Typically a set of network statistics is chosen that was not included in the original model.

time, one may argue that some of the newly developed methods that allow for the analysis of social networks over time are more appropriate. Two of the most popular methods are Stochastic Actor Oriented Models (SAOMs) which are implemented in the RSiena R-package and TERGMs which is essentially ERGMs version of a panel regression. While these methods are statistically advanced and provide a wide range of options to customize the model for many different contexts, they also come with several shortcomings. For example, SOAMs do not handle networks of the size used in this chapter. They are simply too computationally intensive. A shortcoming that both models share is that they require panels of network relations, where each panel represents a slice of the network in time. Since network ties in a citation network never dissolve, it is not quite clear how to practically implement this empirical setting. Given these limitations, I argue that SOAMs and TERGMs do not provide the methodological framework that could accommodate the analyses of the data in this chapter. However, by building on an innovative set of variables I will show that questions about the temporal unfolding of network ties can be modeled within the classic ERGM framework.

The third set of analyses link the structure of the pre-approval citation network to: (1) the number of regulatory actions and to (2) the duration to a given number of regulatory actions.

The number of regulatory actions is the dependent variable and has the following characteristics: 1) its value is non-negative and discrete and 2) it exhibits overdispersion which is a result of its skewed distribution. Since certain assumptions of the Ordinary Least Squares (OLS) regression are violated, non-negative discrete outcome data is most commonly analyzed by assuming that the data-generating process follows Poisson distribution. However, various factors including overdispersion violate the assumptions of the Poisson process. Since Negative Binomial regression models do not carry these assump-

tions I employ negative binomial regression models to study the effect of structures in the pre-approval citation network on the total number of regulatory actions.

To test the hypothesis about duration until regulatory action I estimate a set of Cox Proportional Hazard models (Cox PH models). These are semi-parametric regression models that are commonly used to model the survival of humans in medical settings, but have – in recent years –increasingly been used to accommodate the analysis of duration until an event. The Cox PH model describes the hazard rate of a prescription drug receiving a given number of regulatory actions as a function of covariates and a baseline hazard rate. The hazard rate at age t indicates the rate at which drugs receive the x^{th} regulatory action given that that did not happen up until t and can therefore be referred to as the conditional instantaneous rate at which events occur. In particular, the hazard rate in the Cox PH model employed in this chapter can be specified as a function of the number of months of a drug on the market and the independent variables described in the following section. I define the hazard rate as $h(t, X(t), \beta)$ for a drug that has been on the market for t with independent variables $X_1(t)$, $X_2(t)$ ..., $X_k(t)$ – which are collected in matrix $X(t)$ – and a vector of regression parameters β. Please note that $X(t)$ allows for the inclusion of time-dependent covariates in the model used to estimate the model parameters. The Cox PH model employed in this paper can now be described as:

$$h(t, X(t), \beta) = h_0(t) exp(\beta X(t)) \tag{4.3}$$

The first expression in the function is the baseline hazard and can be estimated non- parametrically. The regression coefficients in the β vector are obtained through maximum partial likelihood, a technique that exploits the order in which objects exit rather than a specific time scale (days, months, years, etc.).

4.4 Data Collection

4.4.1 Study Population

To address the questions outlined in the introduction of this chapter, I collected data on the names, approval dates, and sponsors of all 200 drugs approved by the Food and Drug Administration between 2000 and 2010. Essentially, this data is a subset of the data described in Chapter 3. The data is built from four main sources: Drugs@FDA[7], a 1989 Center for Drug Evaluation and Research publication, NDA Pipeline, and the 1999 - 2011 Drug and Biologic Approval Reports[8]. The main reason for focusing on the subset of drugs rather than the entire population is twofold (and mostly pragmatic): First, left truncation is needed because citation data for older publications (mostly pre-1996) is of much lower quality than more recent citation data. Second, right truncation is used because for each prescription drug a window is needed after approval of the substance to test the hypotheses that are the focus of this chapter.

After identifying all drugs approved between 2000 and 2010, I matched each of the generic names of these drugs to a unique MeSH term[9]. MeSH is a classification scheme and controlled vocabulary and it is mainly used to classify academic literature through the matching of the primary content of a publication with one or multiple MeSH terms. MeSH is maintained by the National Library of Medicine and can be downloaded for free from PubMed.[10] In some cases a drug was not indexed as a MeSH term but only as a Supplementary Concept Record (SCR). SCRs are used to index chemicals, drugs, and other substances that are not included in the MeSH vocabulary. The main reason why these drugs are not included in the MeSH tree is simply that there are too many:

[7] 1.usa.gov/1pCpJaZ.
[8] 1.usa.gov/1vr3lHr.
[9] MeSH is the acronym for "Medical Subject Headings".
[10] ncbi.nlm.nih.gov/mesh.

while there are about 26,000 MeSH terms, there are about 200,000 SCR records with over 505,000 SCR terms. The major difference between a MeSH term and an SCR is that the former has its own MeSH tree number. However, each SCR is linked to one or more MeSH terms[11] and like MeSH terms, SCRs are searchable through PubMed.

After matching each of the generic names of the prescription drugs in the sample to a MeSH term, I extracted all publications from PubMed that were assigned at least one of these MeSH terms (or SCRs). PubMed is National Library of Medicine's database of academic publications in the fields of medicine, nursing, dentistry, veterinary medicine, health care systems, and pre-clinical sciences and has been show to be one of the most complete resources for querying publications in these fields (Falagas et al. 2008). The extraction of the publications from PubMed was done through a Python script that interacts with The Entrez Programming Utilities, the API of the National Center for Biotechnology Information (NCBI). The script downloads the PubMed ID (PMID) of each publication and a series of characteristics of the publication including title, journal title, authors, associated MeSH terms, language, etc. I subset the data to only include publications that were in the English language.

After identifying all PMIDs of the publications associated with the 200 prescription drugs, I used the Scopus API to download all the references of those papers. Scopus is a commercial product from Elsevier and has been argued to be one of the most detailed resources for retrieving citation data (Falagas et al. 2008). Although this strategy to download citations is superior to any other currently available strategy, it comes with one caveats. First, some of the academic journals indexed in PubMed are not indexed in Scopus. As a result, the citations for those publications cannot be retrieved. This problem is minor

[11] The MeSH terms to which SCRs are linked typically describe a drug class which is defined either on the basis of chemical similarity or on the basis of similarity in the mechanism of action.

for the analyses conducted in this chapter because less then 1% of the PMIDs downloaded from PubMed is not found in Scopus.

4.4.2 Variable Descriptions

4.4.2.1 Logistic Regressions

Data for the first set of analyses come mostly from PubMed. The risk set comprises all papers that discuss one of the 200 drugs in the sample and the response variable in the models is binary and equals 1 if the MeSH terms on a paper include the MeSH qualifier "adverse effects." The independent variables in the model include a Post-Approval Publication dummy which is equal to 1 if the paper was published after approval of the drug by the FDA and 0 otherwise. The second independent variable, Previous Article Counts, is the sum of all articles about a given drug published prior to the focal paper. This variable controls temporal variation in the likelihood of publishing adverse effect papers that can be accounted for by the stage in which the debate is in. If papers about adverse effects need a critical basis of scientific work to become meaningful, one would expect that the early stages of a scientific debate are characterized by low levels of papers about adverse effects. I include a similarly constructed variable, Previous AE Article Count, that captures the number of publications prior to the focal paper that have been assigned the adverse effects MeSH qualifier. This variable controls the variation in the dependent variable that relates to opportunities to publish a specific type of paper. One may argue, for example, that first publications on a specific topic suffer from a first mover penalty and that the risk of publishing an adverse effect paper increases with the number of adverse effect papers. Conversely, there might be a premium for those publications that pioneer uncharted territory which would reduce the number of publications about adverse effects if the number of adverse effect papers

increase. The fourth independent variable, Number of Authors, controls variation introduced by the type of publication. Different types of publications vary in how much time and resources are needed to carry out the research (Wuchty et al. 2007), and the number of authors of a publication represents a crude proxy size of the project.[12] Debate Embeddedness captures the share of the references citing another paper within the debate. Obviously, some of the references on the papers within the scientific debate about a prescription drug are citing papers that are not embedded within the debate so those citations are coded as external citations. The variable is computed then as the sum of citations within the debate over the total number of citations. Finally, the Missing References dummy captures whether a publication has no references. That can have three reasons: 1) the publication is too recent and is not yet added to Scopus, 2) the publication contains no references, and 3) the publication contains no references to publications within Scopus.

4.4.2.2 ERGMs

Before describing the variables used to model the ERGMs, let me briefly review the definitions used to construct the network. Nodes in the citation network are represented by papers, while the ties or edges between papers are represented by a citation from one paper to another. The relationship between two papers is binary – it indicates the presence or absence of a citation going from node i to node j – and it is directed. Therefore, by definition, an edge that runs from node i to node j implies that the paper represented by node j must have been published before node i was published.

The network mechanisms that I will introduce below aim to account for the actual tie-formation processes. The first variable included in ERGM is the Edges

[12] I experimented with this variable. For example, I included a squared term for the Number of Authors, but decided not to include the variable in the model because it was statistically not significant and it did not alter the effect size or significance of the variable of interest.

term. This variable is akin to a constant in a standard regression model and if all other variables are removed from the model it is equal to the density of the network. It therefore captures the average tendency of actors in the network to form ties with other actors in the network.

I constructed an Adverse Effect variable and included it in the model as a "nodemix" term.[13] This variable builds on a dummy variable that is equal to 1 if the paper was assigned an adverse effect MeSH qualifier and 0 if not. What a nodemix term allows the researcher to do is to model the effect of all possible combinations of the dummy variable while leaving one of the possible combinations out to serve as the reference group. Since the network in this chapter is directional, there are four possible combinations: 1-1, 1-0, 0-1, and 0-0. I leave out the 0-0 combination and the log-odds for the remaining three combinations can be interpreted as differences from the reference group.

The third variable included in the models also builds on a dummy variable: it captures whether a paper was published prior to the approval of the prescription drug addressed in the paper or whether if was published after approval. This variable is modeled through the inclusion of a nodemix term. Since, a citation from a paper published prior to approval to a paper published after approval is logically impossible, the term only leaves three combinations. Moreover, since the hypothesis outlined earlier builds on the concepts of pre-approval cohesion and post-approval integration, I could leave out the combination $0 \rightarrow 1$ or $0 \rightarrow 0$, where 1 represents a pre-approval paper and 0 represents a post-approval paper. The $1 \rightarrow 1$ must be included because it captures the cohesion within the pre-approval publications. The two realizations of the term essentially capture whether citations within a stage are more or less likely to occur between stages (i.e. in the pre- or post-approval stage).

[13] The independent variables in an ERGM can come in a variety of forms. See **?** for a detailed overview of the available terms.

I also include a "nodeofactor" variable. This term captures the average tendency of a group (the two groups of pre- and post-approval publications in the current context) to send ties – hence the "o" in the term which stands for outdegree. This term can be seen as the main effect of the pre-post nodemix variable. While the model by definition corrects for the sizes of the groups for which mixing is evaluated, the model does not account for different levels of networking tendencies. A nodefactor variable does precisely that. I have included a nodeofactor variable because the opportunities for pre-approval publications to send out ties within the debate are much fewer (i.e. there are fewer within-debate papers available). It is not advised to include a nodeofactor variable for each group because, because the sum of all such statistics equals twice the number of edges and hence a linear dependency would arise in any model also including edges (Morris et al. 2008). Therefore, one of the groups is omitted and serves as the reference group.

The fifth variable is included as a control variable. If science evolves cumulatively and if new findings replace older findings, citation patterns are likely to be characterized by short lags between the publication date of the focal paper and the cited paper.[14] However, this would also increase the likelihood that papers within stages cite one another. To control this temporal effect I include it in the ERGM as an absdiff term. Such a term requires the researcher to assign some value to all nodes and the model then computes a matrix in which the differences between these absolute values are stored. Although are data are time-stamped at the day level I include the variable as the difference in weeks to generate a larger size for the estimated coefficient. Including the term allows one to test whether papers published closer together in time are more likely to

[14]Please note that I am using the publication data here, not the submission date of a paper. While a number of journals provides the submission date of a paper on their website, it is not always available in PubMed. However, manual inspection of the time difference between submission and publication in the medical sciences is short. Rather than months or years, the lag between submission and publication in the medical sciences tends to be a couple of weeks at most.

cite one another than papers published further apart.

The sixth variable captures similarity in the content of two publications. Based on the MeSH data described above, I generate a set of possible MeSH terms for each scientific debate about a prescription drug. From this set of terms, I construct a vector that captures whether a publication was assigned a specific MeSH term (i.e. I set the value in the vector to 1) or not (i.e. I set the value of the vector to 0). I then compute the cosine similarity between the vector of each publication in a debate and I store the values in a square matrix. This matrix is essentially a distance matrix (i.e. rather than capturing spatial distance, it captures distance in content) and is has the same dimension as the matrix of observed citations. Since in many of the larger networks, the inclusion of the distance matrix inhibits convergence, I transform the valued matrix to a binary matrix by assigning a 1 to cases that exceed the median and 0 to cases that are equal to or lower than the median. The estimated coefficient for the term indicates whether more similar papers are more likely to be linked through a citation from the more recent to the older paper.

The seventh variable in the model is also an edgecov term. It builds on a matrix of dummies indicating whether there is overlap in the author sets of two papers. If two papers share at least one author, the cell in the matrix is equal to 1 and it is equal to 0 if there is no overlap between the two. The estimated coefficient for the term indicates whether papers with overlap are more likely to be linked through a citation than papers that do not share authors.

Finally, the model includes two network statistics. The first statistic – GWESP, or "geometrically weighted edgewise shared partner" – captures triadic closure. In the case of citation networks it allows one to test whether a citation from paper A to paper B (A $\rightarrow$ B) is more likely to occur if A was also citing paper C and paper C had also cited paper B (A $\rightarrow$ C and C $\rightarrow$ B). An alternative term in the model would have been the triangle term, but inclusion of such a

term increases the likelihood that the ERGM is degenerate. The GWESP variable is essentially a parametric version of the triangle variable and it reduces the potential for degeneracy.[15] The GWESP term adds a network statistic to the model that equals the (geometrically weighted) edgewise shared partner distribution – the number of times each connected pair shares ties with a common third.[16]

The second network statistic included in the model is the "geometrically weighted dyadic shared partner" (GWDSP) statistic, which is a function of the number of dyads which have k neighbors in common (where k ranges from 1 to $N - 2$). The GWDSP statistic models the distribution of shared partners of nodes that may or may not be tied themselves. Since the GWESP statistic is included in the model, the GWDSP statistic can be thought of as a measure of structural imbalance, representing situations where A does not cite B despite having one or more citations in common. Evidence of endogenous clustering in a network would be supported by a positive GWESP term and a negative GWDSP term in the same model.

Given these data, the resulting model looks as follows:

[15] Model degeneracy stems, in part, from the cascading tendencies of the transitivity term. That is, in closing one triad another triad involving adjacent nodes will typically be opened leading to an infinite regress) (Hunter et al. 2008).

[16] Compared to the triangle term, the GWESP term implies a decreasing marginal return to each additional shared partner that would create a transitive triad.

$$
\begin{aligned}
Pr(Y = y|X) = \beta_{0c}(Edges_{ij}) + \\
\beta_{1c}(Adverse\ Effect_{ij}) + \\
\beta_{2c}(Pre - Approval_{ij}) + \\
\beta_{3c}(Time\ Difference_{ij}) + \\
\beta_{4c}(MeSH\ Similarity_{ij}) + \\
\beta_{5c}(Author\ Overlap_{ij}) + \\
\beta_{6c}(GWESP_{ij}) + \\
\beta_{7c}(GWDSP_{ij}).
\end{aligned}
\tag{4.4}
$$

4.4.2.3 Negative Binomial and Cox Proportional Hazard Regressions

The risk set in the analyses predicting the volume and speed at which drugs are targeted by regulatory action comprises all 200 drugs in the sample. The dependent variable for the Negative Binomial Model captures the number of label changes that a drug has been characterized by since approval of the substance. The dependent variable for the Cox PH model is the duration to the second label change, which is the median of the number of label changes in the first 4 years since approval. Appendix A of this chapter further describes the details of duration until label changes using Kaplan-Meier plots.

The independent variables in the model include a dummy variable for Priority Review which is equal to 1 if a drug was approved through Priority Review and 0 otherwise. A Priority Review designation means that the FDA reviews the drug application within 6 months (compared to 10 months under standard review). Drugs that the FDA believes would be significant improvements in the safety or effectiveness of the treatment of serious conditions are eligable to receive Priority Review designation. The second variable included in the model

captures the number of approved drug applications by the sponsor of the focal drug prior to submitting it for approval. As I have mentioned before, new drug applications that were submitted but not approved are unfortunately not publicly available. The third variable captures the number of months it took to get the drug approved by the FDA. For each submission, I recorded the submission date and the approval date, calculated the days in between those dates, and transformed the variable from days to months. I also included variables that describe the pre-approval scientific debate. First, I include a count of the number of academic papers that have been published about a given prescription drug *before* the drug is approved for marketing in the U.S. Finally, I also include a variable that described how cohesive the citation structure is among the papers that were published before approval of the prescription drug. Building on the citation network of the pre-approval publications, I compute community membership based on a community detection algorithm developed by Newman and Girvan (2004). This algorithm essentially captures the idea that it is likely that edges connecting separate communities must have high edge betweenness (meaning that those edges serve as high traffic routes if one were to find all shortest paths between all nodes in the network). By gradually removing edges with the highest edge betweenness one can create a set of communities. Based on this detection method, I compute the modularity score.

4.5 Results

The results of the analyses aiming to describe the differences between the pre- and post-approval debate are shown in Table 4.1. Each of the four models presented in this table shows the coefficient estimates of a Logistic regression. In the first column all observations across all drugs are pooled; the second column presents a drugs fixed-effects model; the fourth column presents a multilevel

model with random intercepts; and the final models is a multilevel model with random intercepts and random slopes. The results show that scientific studies published prior to approval of a prescription drug differ significantly from studies published in the post-approval stages of a drug's life-cycle. In particular, the pre-approval debate is characterized by significantly lower numbers of publications that are specifically about ADRs potentially associated with a prescription drugs. Despite the fact that the coefficient estimates vary across all four models, the Post-Approval Publication variable is consistently positive and significant. The most conservative estimate – shown in column 2 – implies that the odds of a post-approval publication discussing ADRs are 1.5 times the odds of a pre-approval publication discussing ADRs. The least conservative estimate – shown in column 1 – suggests that the odds ratio is 2.5. Other coefficients indicate that publications are likely to be explicitly about ADRs if the number of previous articles on the specific drug is higher, if the number of previous ADR articles is lower, if the number of authors on a publication is higher, and if the publication is more engaged with other publications about the same drug.

The main finding can be interpreted in two ways. First, the difference can be explained by arguing that the resources to identify the effect of treatment with a drug on the incidence of ADRs are increasing after approval. Examples of such resources include increasing data volumes, allowing for the statistical identification of the effect, and the increase in human resources which allows for a wider variety of hypotheses to be constructed. In the absence of other explanations, this interpretation should hold across essentially all drugs and the effect size should be fairly homogeneous – meaning that the lower likelihood for publications prior to approval to discuss ADRs associated with a drug should be of similar size across all drugs.

A second explanation points in the direction of the influence that the spon-

Table 4.1: Predicting the probability that a paper discusses ADRs

	Pooled	Fixed Effects	Random Intercept	Random Intercept-Slope
Intercept	-2.462^{***}		-2.099^{***}	-2.164^{***}
	(.021)		(.068)	(.077)
Post-Approval Publication	$.907^{***}$	$.400^{***}$	$.420^{***}$	$.480^{***}$
	(.020)	(.024)	(.024)	(.053)
Previous Article Count	$.000^{***}$	$.000^{***}$	$.000^{***}$	$.000^{***}$
	(.000)	(.000)	(.000)	(.000)
Previous ADR Article Count	$-.002^{***}$	$-.001^{***}$	$-.001^{***}$	$-.001^{***}$
	(.000)	(.000)	(.000)	(.000)
Number of Authors	$.012^{***}$	$.020^{***}$	$.020^{***}$	$.020^{***}$
	(.002)	(.002)	(.002)	(.002)
Debate Embeddedness	1.470^{***}	1.766^{***}	1.764^{***}	1.758^{***}
	(.032)	(.035)	(.035)	(.035)
Missing References Dummy	$.202^{***}$	$.553^{***}$	$.545^{***}$	$.553^{***}$
	(.021)	(.022)	(.022)	(.022)
AIC	151777.502	143151.319	143751.805	143437.783
BIC	151848.681	145246.024	143833.153	143539.468
Log Likelihood	-75881.751	-71369.659	-71867.903	-71708.892
Num. obs.	192,620	192,620	192,620	192,620

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Note: This table shows the point estimates and standard errors of four Logistic regression models. The columns contain estimates based on pooled data, fixed-effects, random intercepts, and random intercepts and random slopes.

sor of a drug has on the way in which science is conducted prior to approval. It is obviously in the interest of the sponsor of the drug to have the new drug application associated with as few ADRs as possible, and sponsors of new drug applications are therefore incentivized to limit the number of publications that could potentially jeopardize the success of their product in the approval process.[17] This explanation would also be consistent with prior research showing that there is supply-side[18] driven publication bias – firms holding back publications that could negatively impact the product that they represent (Ahmed et al. 2012; Chalmers et al. 2013; Perlis et al. 2005; Higgins et al. 2011).

In order to distinguish between these two explanations, Figure 4.1 shows a simple scatterplot of the bivariate relationship between the proportion of pre-approval ADR papers (all pre-approval ADR papers / all pre-approval papers) and the difference between the proportion of post-approval ADR papers and the proportion of pre-approval ADR papers.[19] The plot demonstrates that drug debates with relatively few pre-approval ADR publications are characterized by higher proportions of post-approval ADR publications.[20] The second interpretation differentiates between drugs (i.e. if pre-approval ADR publications are held low, the increase in post-approval ADR publications will be high because ADRs that could have been identified before are now identified only after

[17] One could argue that sponsors of drugs would not exert their power to steer the process at the risk of the drug being targeted by post-approval regulatory action. However, if one accounts for the fact that pretty much all drugs are targeted by regulatory action and that market withdrawals are rare events, the trade-off may favor exerting pre-approval influence on content of scientific publications.

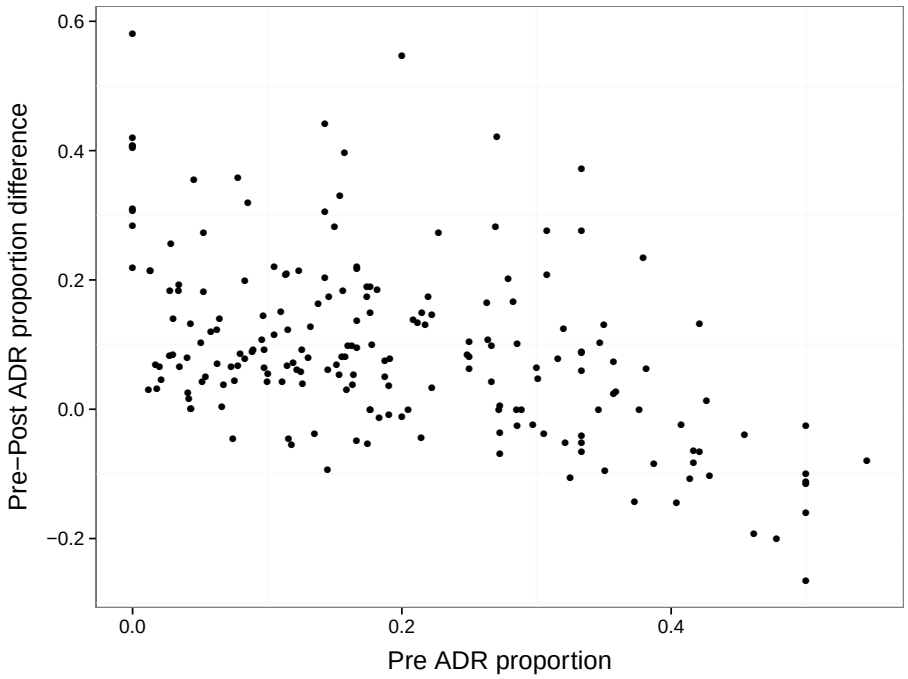
[18] Supply-side driven publication bias is defined here as publication bias that results from research that is not being conducted or research that is not being submitted. Conversely, demand-side driven publication bias results from research publishers exhibiting bias in the acceptance of research for publication.

[19] Making this comparison makes more sense than directly comparing the pre-approval and post-approval proportions, because some drugs are inherently more problematic than others and will therefore exhibit a strong and positive correlation by definition.

[20] Obviously, since proportions are bounded between 0 and 1, a very high pre-approval ADR proportion is by definition paired with a smaller pre-post difference. However, even if one were to focus only on the cases in which the pre-approval proportion was less than 0.19 – which is the average accross all observations, the negative correlation is still present and highly significant at 0.29. For this sample, the maximum difference between the pre and the post proportions is 0.58 which would still leave ample room for additional ADR papers.

approval) and is consistent with the graph. The first interpretation, however, seems less plausible since it cannot account for the heterogeneity in differences and the strong correlation between the pre-approval ADR publication proportion and the difference with the post-approval proportion. Although the graph does not completely resolve the causal question, it shows a pattern that is consistent with the idea that if the sponsor of a drug is able to reduce attention for ADRs prior to approval, attention for those ADRs will be picked up in later stages of the drug life-cycle.

Figure 4.1: Scatterplot of pre-approval ADR papers and change in ADR paper proportion



Note: This scatterplot shows the bivariate relationship between the proportion of pre-approval ADR papers (all pre-approval ADR papers / all pre-approval papers) and the difference between the proportion of post-approval ADR papers and the proportion of pre-approval ADR papers. The correlation coefficient is -0.49 and is significant at the 0.01 level.

The second set of findings are derived from the ERGMs that I have estimated for the citation networks of all drugs approved between 2000 and 2010.

Before going into the specifics of the variables of interest, Figure 4.2 describes the distribution of the coefficient estimates across all 200 models. The first variable, Pre nodefactor, captures the tendency of pre-approval papers to cite other papers. The reference group comprises all post-approval papers. Obviously the estimated coefficients are always negative: papers published prior to approval have much fewer options to cite other papers in the debate than papers published after approval. The average estimate for this variable indicates that pre-approval publications are citing 97% fewer other papers in the debate than do post-approval publications. The second variable, Pre-Pre citation is positive – as expected – and implies that edges between papers that were both published prior to approval are much more likely than edges between papers published after approval. While this finding is trivial, what's not trivial is its variation across models. I will discuss this variation shortly. Figure 4.2 also shows that the Post-Pre citation variable is mostly positive. This implies that, in general, publications that came out after approval of a prescription drug are more likely to cite papers that came out before approval than papers that came out after approval. This finding is in line with the idea that the making of facts in these networks is a cumulative process, where recent work builds on earlier work. The following three variables capture the mixing patterns of papers that discuss ADRs and papers that do not discuss ADRs. The reference group comprises ties between papers that do not address ADRs. Among these three variables, the tendency of ADR papers to cite each other is the strongest.

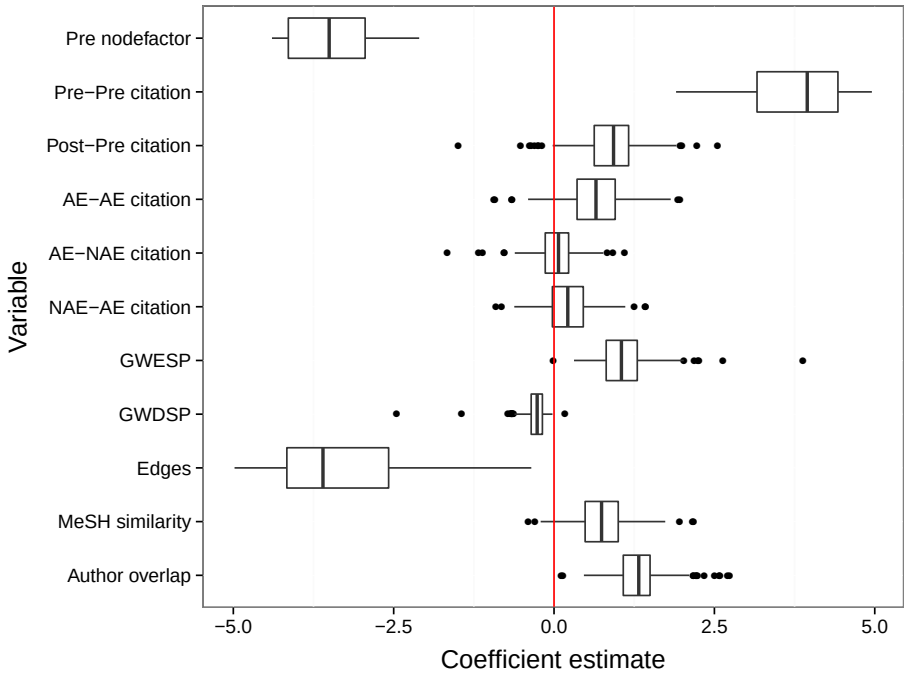
The models also contain endogenous network statistics. The GWDSP variables – which are mostly negative – and the GWESP variables – which are mostly positive – indicate that citations are likely to form ties that complete transitive triads. In other words, a citation from paper A to paper B (A $\rightarrow$ B) is more likely to occur if A was also citing paper C and paper C had also cited paper B (A $\rightarrow$ C and C $\rightarrow$ B). Please note that this effect is the net effect of those

structural properties and that the effect cannot be explained by similarities in content, or any of the other variables in the model. Finally, the two variables that capture similarity in terms of authors and MeSH terms both indicate that citations are more likely to be present between papers that share authors and between papers that have similar sets of MeSH terms associated with them. The majority of the models provides estimates that confirm the findings in earlier work on how citation networks form and this provides confidence that the modeling strategy allows me to make correct inferences in the processes that have guided the formation of the network.

However, to assess if the models indeed do a good job of describing the network formation process, I computed the GOF statistics for each of the 200 models. Although for a few individuals cases other model specifications provided better GOF statistics, the model specification shown here provides the best average fit among a series of model specifications that I have tested.

To assess the hypothesis developed earlier in the chapter about the relationship between the internal structure of the citation network in the pre-approval stage and the extent to which those papers become integrated in the post-approval debate, Figure 4.3 plots the paired coefficient estimates for Pre-Pre citation and Post-Pre citation. The left panel of Figure 4.3 plots the sorted coefficient estimates for Pre-Pre citation, while the right panel plots the paired coefficient estimates for Post-Pre citation. That is, figure 4.3 contains 200 lines (along the vertical axis of the graph) and each line contains the estimates of the Pre-Pre citation variable and Post-Pre citation variable. The graph shows that there is a strong tendency for debates with relatively low internal cohesion in the pre-approval stage to be characterized by relatively high levels of Post-Pre mixing. That is, the likelihood that papers in the post-approval debate cite papers in the pre-approval debate (versus other papers in the post-approval debate) is higher when the likelihood of papers in the pre-approval debate citing

Figure 4.2: Boxplot of coefficient estimates

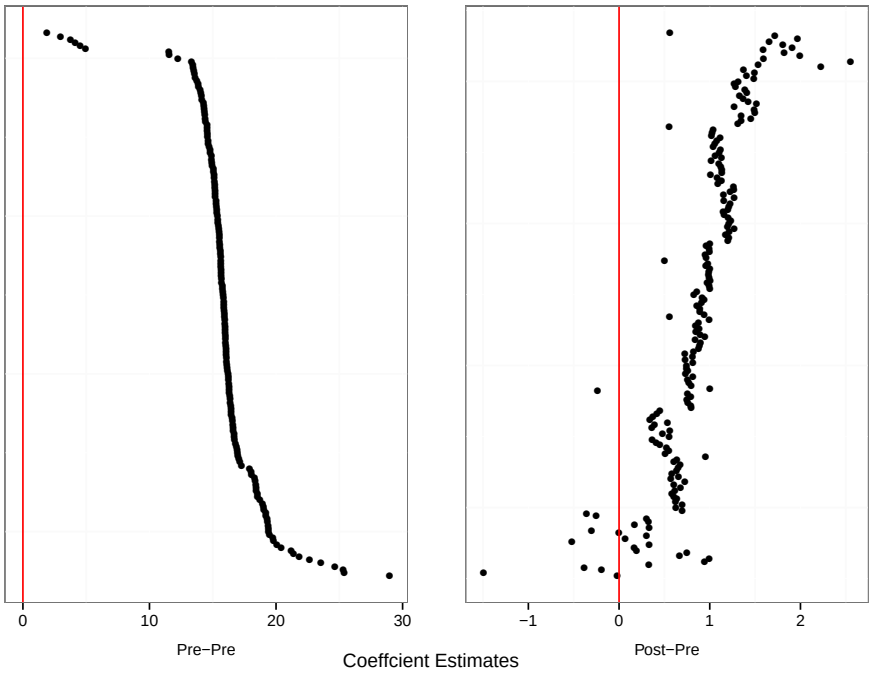


Note: This graph provides a summary of the coefficient estimates for all 200 ERGMs estimated for the drugs in the sample. The boxplots show what the range of the estimates is and the vertical red line shows the boundary between a positive and a negative estimate.

each other is lower. This implies that as there is less cohesion among papers in the pre-approval debate, these papers are more likely to be integrated in the post-approval debate. This finding suggests that a lack of pre-approval intra-debate cohesion is reflective of the wider range of possible hypotheses that are studied for those drugs and that research is therefore less redundant. Or to put it in opposite terms, if all pre-approval research only focuses on one or a set of closely related questions, post-approval research will only cite one or a few of those papers because the questions that they address are redundant.

I now move from the question about how citation networks form to the question of how the structure of a debate is associated with the rate and volume of regulatory action. Table 4.2 shows the coefficient estimates for three models

Figure 4.3: Coefficient estimates of Pre-Pre and Post-Pre nodemix terms



Note: These two plots show the relationship between the Pre-Pre and Post-Pre mixing coefficients extracted from the ERGMs. The left panel is sorted from small to large and the estimates on the left are matched – meaning that two points that line up horizontally belong to the same ERGM. Similar to Logistic regression, exponentiating the coefficient estimates provides one with the odds-ratios.

predicting volume of regulatory action and three models predicting duration to regulator action. Across all six models, the final set of findings show that drugs characterized by high levels of pre-approval network modularity are targeted by fewer regulatory actions. The coefficient in model 3 implies that an increase in modularity of 0.1 reduces the number of label changes by 7.4%. The Cox PH models also show that drug debates characterized by high levels of pre-approval network modularity are slower to accumulate regulatory action. This is consistent with the hypothesis 3 developed earlier in this chapter. High levels of modularity suggest that there is little redundancy across the publications and the pre-approval set of publications is therefore likely to have tackled a wider

range of questions.

4.6 Discussion and Conclusion

The aim of this chapter was to describe the process by which scientific evidence about prescription drugs is produced and to examine its relation to the regulatory trajectory that characterizes the drug after approval. Since the production of evidence prior to approval is controlled by the sponsor of the drug, I have compared evidence production in the two stages and I have looked at how the evidence produced by sponsors of the drug become integrated in the scientific debate that unfolds after approval.

The chapter presents three main substantive findings. First, publications of studies conducted prior to approval that are commissioned by drug sponsors are less likely to discuss in detail an ADR that is potentially associated with treatment with the drug. Additional analyses show that drug applications characterized by low levels of ADR publications prior to approval are associated, on average, with a higher number of post-approval publications that are specifically about the ADRs associated with a drug. This finding is consistent with the idea that some firms are able to develop publication strategies that benefit the drug that they sponsor. The second finding is that cohesive citation structures among studies published prior to approval are less likely to become integrated in the post-approval scientific debate about the safety and efficacy of a prescription drug. One interpretation of this finding – that is in line with the first finding –holds that there is an incentive for sponsors of prescription drugs to limit the number of different questions asked about the safety of a prescription drug. Doing so will protect them from critique that could jeopardize approval. However, as a result, studies conducted after approval will have to explore that questions left open by the sponsor of the drug and those publica-

Table 4.2: Predicting volume of and duration to label changes

	NB1	NB2	NB3	CPH1	CPH2	CPH3
Intercept	1.758***	1.798***	2.016***			
	(.167)	(.167)	(.172)			
Priority review	−.117	−.108	−.124	−.143	−.115	−.131
	(.109)	(.109)	(.106)	(.171)	(.173)	(.173)
Sponsor's prior approved drugs	.018***	.018***	.015***	.029***	.027***	.024***
	(.004)	(.004)	(.004)	(.006)	(.006)	(.006)
Review time	.000	.000	.000	.000	.000	.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Number of pre-approval pubs.		.000	.000		.000	.000*
		(.000)	(.000)		(.000)	(.000)
Pre-approval network modularity			−1.336***			−1.488**
			(.355)			(.549)
Year dummies included	Yes	Yes	Yes	Yes	Yes	Yes
AIC	1145.381	1144.700	1133.409	1586.642	1584.176	1577.432
BIC	1191.557	1194.174	1186.182			
Log Likelihood	-558.690	-557.350	-550.705			
Num. obs.	200	200	200	200	200	200
R^2				.156	.175	.210
Num. events				176	176	176

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Note: This table shows the point estimates and standard errors of two sets of regression models. Model 1 to 3 shows the estimates for the Negative Binomial regression and Model 4 to 6 show the estimates for the Cox PH regressions.

tions will therefore have no broad scientific basis to build on. The final set of findings show that prescription drugs characterized by cohesive pre-approval citation structures are subject to higher levels and higher rates of label changes after approval. The interpretation of this finding relates closely to the findings one and two: fewer questions answered prior to approval, leads to more problematic regulatory biographies of drugs after approval.

Although the analyses presented in this chapter do not allow me to causally identify the mechanisms that produce some of the patterns that I observe, the analyses are consistent with the idea that the adoption of certain publication strategies by drug sponsors can interfere with the goal of the FDA to regulate drugs efficiently. Prior work has shown that industry involvement in science can introduce publication biases (Ahmed et al. 2012; Chalmers et al. 2013; Perlis et al. 2005; Higgins et al. 2011) and it can affect the type of science that is conducted (Evans 2010). What the study presented in this chapter adds to this work is that companies may influence how facts about prescription drugs are produced and it shows that those practices have real effects in terms of the regulatory actions that follow approval.

There are two shortcomings of the study presented in this chapter that I would like to address in future work. First, I have shown that the structure of science prior to approval allows me to predict at which rate the drug will be targeted by post-approval regulatory action. Moreover, I interpret this as evidence of strategic action on behalf of the firm sponsoring the drug. However, I have not been able to make a compelling argument that the drug itself rather than the firm sponsoring the drug is responsible for the structure of science production prior to approval. For example, an alternative explanation would be that new drugs with a mechanism of action that is similar to already approved drugs are less likely to be subjected to a wide variety of tests. Although the Priority Review variable controls for this explanation to some extent, the analyses could

benefit from measures such as those used in King and Bearman (2015). They typify medications according to their innovativeness based on descriptions in the *Medical Letter* which is a non-profit publication.

A second limitation of the study presented in this chapter is that, since the number of observations is low and data that describe the firm is scant, I am unable to explore in detail how firms vary in their publication strategies and how this affects the direction of scientific debates and the rates and volumes of regulatory action. The analyses presented in this chapter show that there is substantial variation across the 200 drug debates that I have studied. Some pre-approval citation networks are characterized by low levels of modularity while others comprise papers that do not cite each other. And while I have shown that this variation has an effect on how the scientific debate that follows unfolds and that this variation explains the volume and rate of regulatory action, this chapter does not answer the questions of which firms are able to steer the process and how the regulator deals with these different strategies.

One way to extend the analyses presented in this chapter is by increasing the sample to exploit the increased variation between firms in a larger sample. For example, the 200 drugs analyzed in this chapter have been produced by 119 unique firms and the majority of firms only appears once in the sample. A larger sample with multiple observations per firm would allow me to ask and answer a set of interesting questions: does variation in the way in which publications become embedded in the pre-approval debate cluster at the firm level? And if so, which types of firms adopt which types of strategies?

Another extension would be to move from the level of the paper to the level of the individual authors of the paper. If the finding from Azoulay and Fishman (2006) – that clinical research has moved away from universities – holds, one would expect that the individuals employed by other research contexts are less embedded in academia and that that would drive a wedge between the pre-

approval debate and the post-approval debate. In particular, if that is the case, the narrative introduced in this paper would be one that heavily builds on a disconnect between organizational cultures. If some firms are more likely than others to conduct their studies in non-university research contexts and if the science produced in those contexts differs strongly from university science, a disconnect between the pre- and post-approval debate is likely to emerge.

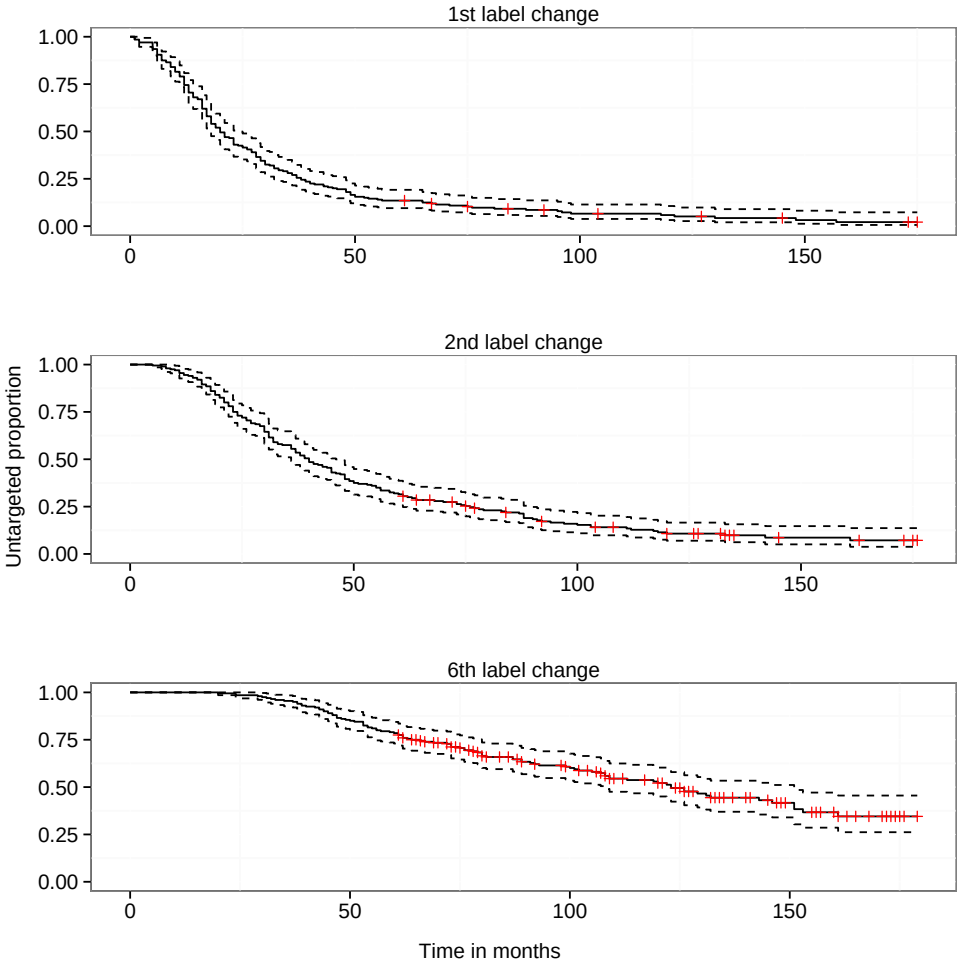
Appendix A Time to Regulatory Action: Kaplan-Meier plots

Virtually all drugs receive at least one label change throughout their life-cycle but there is substantial variation in the rate at which this happens. Figure A4.1 show the Kaplan-Meier plots for the duration until the 1^{st}, 2^{nd}, and 6^{th} label change. Kaplan-Meier plots are typically used to graph a survival function. In the context of drug safety and label changes, the function captures the population that, at time t, has not yet received n label changes. In Figure A4.1, n equals 1 in the upper panel, 2 in the middle panel, and 6 in the lower panel. Formally, the Kaplan-Meier survival estimate is computed as follows:

$$S_t = \frac{\#\ of\ non\text{-}targeted\ drugs\ -\ \#\ of\ targeted\ drugs}{Number\ of\ drugs\ in\ sample} \tag{4.5}$$

The upper panel of Figure A4.1 shows that after about 20 months, 50% of all drugs approved between 2000 and 2010 have received at least one label change and that this number increases to 75% after 50 months. The middle panel shows that it takes about 40 months for half of the sample to receive the second label change. The bottom panel shows that after 59 months, more than 50% of all drugs have received 6 or more label revisions. Moreover, about 30 of the sample does not receive 6 or more label changes during the observed time window.

Figure A4.1: Kaplan-Meier plots



Note: Two Kaplan-Meier plots for the time to first regulatory action (upper panel) and time to median number of regulatory actions (bottom panel). The y-axis shows the percentage of drugs that had not yet been targeted while the x-axis shows time in months. The dashed lines show the confidence interval around the estimate and the red symbols indicate right-censored drugs.

Chapter 5

Conclusion – Structure in the Production of Evidence

5.1 Synthesis and Discussion

The main contribution of this dissertation has been to show how social and cognitive structures that bind those who generate evidence about drug safety can be used to make sense of this evidence. Moreover, this dissertation has provided answers to the question of why it may take a long time before prescription drugs are identified as unsafe by the FDA.

In the first empirical study, I have shown that the communication of the recall of a prescription drug causes patients and physicians to file a large number of reports for the drug that is recalled. While this finding may seem trivial at first, one could argue that it is counterintuitive: the drug has already been taken of off the market and patients and physicians cannot use the anonymous report to strengthen their case against the sponsor of the drug (if they decide to pursue legal charges). Although my analyses do not allow me to identify the motives

of the patients and physicians filing the additional reports, one motive seems most plausible: patients and physicians file reports for the recalled drug after it is recalled to make their voices heard. When patients and physicians were treated with and prescribing the drug respectively, they were doing so under the impression that the drug was safe. The substance had been approved by the FDA – which implies that the FDA argues that the drug is safe and effective. However, the recall of the prescription drug shows that the expertise held by the regulator was not sufficient to protect the health of patients and the credibility of physicians. In such a context, the filing of a report can be interpreted as an instance in which the patient or physician expresses his or her concerns, anger, or discontent.

An alternative explanation is that people report because they *think* that reporting would strengthen their legal case. However, if this is the mechanism that would account for the observed effect, one would expect that the effect was larger for patients than it was for physicians. Although a recall is a rare event, physicians will experience a recall several times during their career and are expected to learn from their prior experiences. For patients, it is much less likely that they will experience a recall of a drug that they are treated with several times during their lives

A second finding shows that physicians respond to the recall of the prescription drug not just by reporting cases in which the recalled drug had been the suspect drug, but also cases in which drugs that treat the same health condition had been the suspect drug. What is more, this finding does not hold for patients: their response is limited to the drug that is recalled. Identifying these patterns and the way in which they are created is important for the regulator that needs to distinguish between increased reporting caused by an increase in the number of ADRs versus an increase in the likelihood of reporting. This latter type of increase could be termed a false positive response. Obviously, these

two data generating processes require different responses from the regulator and the inability to distinguish between them could cause serious problems – including withdrawing perfectly safe drugs and failing to withdraw unsafe drugs. Finally, preliminary evidence suggest that episodes of stimulated reporting cause problems for the FDA because the speed with which they detect true safety problems is reduced. This finding is consistent with the idea that the presence of false positive signals creates ambiguity for the FDA in the interpretation of drug safety signals.

Besides the impact that these findings have on the medical profession, the findings in Chapter 3 also have important sociological implications. The study shows that diffusion of behavior may be structured not in terms of social interactions but in terms of the cognitive maps that underlie the behavior of social actors. If individuals have been conditioned in a way that connects signal to behavior, individuals do not have to interact to create cascading effects.

Rossman (2014) makes a somewhat similar argument about how innovations diffuse. Building on models of exogenous diffusion and endogenous diffusion, Rossman (2014) asks whether the legitimacy of the innovation moderates the adoption of the innovation under the exogenous diffusion condition. The simple idea behind this hypothesis is that when individuals find that an innovation belongs to a category that is legitimate (i.e. the category contains older innovations that have been well-established and socially adopted), they do not need social influence to adopt; all they need is awareness about the availability of the innovation. However, when the innovation is not a member of a legitimate category, (Rossman 2014) argues that the innovation will only diffuse if there is social influence (i.e. endogenous contagion). The crux here is thus to understand how an innovation is embedded in a set of other innovations. If the social actors who are at risk of adopting have internalized shared understandings about the embeddedness of an innovation in a set of innovations, their

actions will by synchronized – even if there is no social influence.

The second empirical study shows that the way in which scientific publications build one one another varies between drugs. Drugs for which the pre-approval debate is narrow – meaning that the scientific publications are linked through a cohesive structure – are less likely to be of value for post-approval scientific work and are more likely to be targeted by large numbers of regulatory actions. Regulatory action is also more likely to occur *fast* for those drugs. In contrast, prescription drugs that are characterized by pre-approval scientific publications that are not tightly connected to the other pre-approval publications are more likely to be of value for later research and those drugs are not as likely to be targeted by regulatory action. This research has implications for the way in which we think about pre-approval regulation by the FDA. Which studies should be conducted? Which linkages – between a drug and a specific ADR – should be investigated? And what is the role of firms versus the FDA in terms of answering those questions?

The research also reveals an interesting link between what happens before approval and after approval. One might argue that a "spiral trajectory" (Shwed and Bearman 2010) is what the science conducted to understand the safety and efficacy of prescription drugs should look like. However, some drugs are characterized by spiral trajectories that emerge only after the drug is approved. In particular, the post-approval science in those cases essentially starts from scratch. Although the chapter is unable to establish a causal link between firm characteristics and the way in which the debate develops, evidence is consistent with the idea that experienced firms are able to set up evidence production in a way that favors positive evaluations of their product.

The second empirical study also shows how the structure in which claims are embedded is formed and how it impacts they way in which a regulator can build upon those claims. In particular, by using network analytic techniques,

the study reveals initial insights on how organizations can discipline social networks. Although it is difficult to identify a causal effect, the evidence is consistent with the idea that some organizations are able to exert influence on the ways in which connections are made. Once the organization loses control over the way in which networks are formed, the initial mode of network formation can have a lasting impact on the way in which future connections are made and the ways in which regulators can extract the meaning from those networks.

5.2 Conclusion

This dissertation has provided insights into a context that has never been studied by sociologists: the production of evidence about drug safety. Studying this context is important for two main reasons: First, many patients are dying from treatment with unsafe drugs and earlier detection of unsafe drugs may reduce that number. Second, the detailed accounts of how the behaviors of social actors translate into bodies of knowledge advances the sociology of knowledge production.

Building on sociological theories about how knowledge claims are constructed and how scientific associations form, the broad argument laid out in the dissertation holds that to extract meaning from massive sets of data, accounting for the social and cognitive structure in which the data production process is embedded is crucial. That is, complex systems such as the one in which the evidence about the safety and efficacy of prescription drugs is generated are characterized by heterogeneous actors who's behavior varies over time. Theorizing about those complexities allows one to capture the meaning of that aggregated behavior.

The implications of this argument extend beyond the context of drug safety. For example, the recently popularized idea that "big data" will allow science

and industry to accurately predict future behavior of social actors is meaningless if one does not have a clear understanding of how the data used to generate the predictions is produced. Chapter 3 of this dissertation has shown that the production of data varies between groups and over time. For the data scientist trying to use big data to predict future patterns of human behavior without using social theories, that variation will represent noise leading to less accurate predictions.

The research presented in this dissertation also has implications for research on diffusion studies. The analyses presented in Chapter 3 show that behavior diffuses and becomes clustered in time, not because social actors interact with one another, but because the response of social actors to media events spreads to contexts that go beyond the media event. In particular, the central object in the media event is embedded in a larger set of objects and social actors include those related objects to inform their behavior. This suggests that recent research on social networks and complex systems should also pay attention to interdependencies that represent commonly held beliefs.

Finally, this dissertation has implications for how science is conducted. The analyses in Chapter 4 show that a scientific debate may become characterized as disconnected if industry involvement in science steers science into certain directions. Although the pharmaceutical industry is obviously one of the most well-known industries to become involved in science, many other disciplines accommodate non-university science. Understanding how these different affiliations affect how science evolves is important if the science is consumed by policy makers trying to extract best practices from a body of scientific literature.

Bibliography

Abbott, A. (1988). *The system of professions: An essay on the division of expert labor.* University of Chicago Press.

Abraham, J. and R. Ballinger (2012). Power, expertise and the limits of representative democracy: genetics as scientific progress or political legitimation in carcinogenic risk assessment of pharmaceuticals? *Journal of Community Genetics 3*(2), 91–103.

Adams, S. H. (1907). *The great American fraud*. New York: P.F. Collier & Son.

Adut, A. (2005). A theory of scandal: Victorians, homosexuality, and the fall of oscar wilde. *American Journal of Sociology 111*(1), 213–248.

Ahmed, I., A. J. Sutton, and R. D. Riley (2012). Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ 344*, d7762.

Annas, G. J. and S. Elias (1999). Thalidomide and the titanic: reconstructing the technology tragedies of the twentieth century. *American Journal of Public Health 89*(1), 98–101.

Azoulay, P., W. Ding, and T. Stuart (2009). The impact of academic patenting on the rate, quality and direction of (public) research output*. *The Journal of Industrial Economics 57*(4), 637–676.

Azoulay, P. and A. Fishman (2006). Doctors, and drug development: The rise of for-profit experimental medicine.

Baer, A. R., S. Devine, C. D. Beardmore, and R. Catalano (2011). Clinical investigator responsibilities. *Journal of Oncology Practice 7*(2), 124–128.

Barabasi, A.-L., N. Gulbahce, and J. Loscalzo (2011). Network medicine: a network-based approach to human disease. *Nature Reviews Genetics 12*(1), 56–68.

Barley, S. R. (2007). Corporations, democracy, and the public good. *Journal of Management Inquiry 16*(3), 201–215.

Berkman, N. D., S. L. Sheridan, K. E. Donahue, D. J. Halpern, and K. Crotty (2011). Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine 155*(2), 97–107.

Beunza, D. and D. Stark (2004). Tools of the trade: the socio-technology of arbitrage in a wall street trading room. *Industrial and Corporate Change 13*(2), 369–400.

Bowker, G. C. and S. L. Star (1999). *Sorting Things Out: Classification and Its Consequences*. MIT Press.

Carpenter, D. (2010). *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA* (1 ed.). Princeton University Press.

Carpenter, D., J. Chattopadhyay, S. Moffitt, and C. Nall (2012). The complications of controlling agency time discretion: Fda review deadlines and postmarket drug safety. *American Journal of Political Science 56*(1), 98–114.

Carpenter, D., S. I. Moffitt, C. D. Moore, R. T. Rynbrandt, M. M. Ting, I. Yohai, and E. J. Zucker (2010). Early entrant protection in approval regulation: Theory and evidence from fda drug review. *Journal of Law, Economics, and Organization 26*(3), 515–545.

Carpenter, D., E. J. Zucker, and J. Avorn (2008). Drug-review deadlines and safety problems. *New England Journal of Medicine 358*(13), 1354–1361.

Cavers, D. F. (1939). The food, drug, and cosmetic act of 1938: Its legislative history and its substantive provisions. *Law and Contemporary Problems 6*(1), 2–42.

Cerulo, K. A. (2010). Mining the intersections of cognitive sociology and neuroscience. *Poetics 38*(2), 115–132.

Chalmers, I., P. Glasziou, and F. Godlee (2013). All trials must be registered and the results published. *BMJ 346*, f105.

Collins, H. (1992). *Changing Order: Replication and Induction in Scientific Practice*. University of Chicago Press.

Collins, H. M. (1998). The meaning of data: Open and closed evidential cultures in the search for gravitational waves. *American Journal of Sociology 104*(2), 293–338.

Collins, H. M. and R. Evans (2002). The third wave of science studies studies of expertise and experience. *Social Studies of Science 32*(2), 235–296.

Cranmer, S. J. and B. A. Desmarais (2011). Inferential network analysis with exponential random graph models. *Political Analysis 19*(1), 66–86.

David, G., S. Markowitz, and S. Richards-Shubik (2010). The effects of pharmaceutical marketing and promotion on adverse drug events and regulation. *American Economic Journal: Economic Policy 2*(4), 1–25.

Davidoff, F., C. DeAngelis, J. Drazen, and et al. (2001). Sponsorship, authorship, and accountability. *Archives of Otolaryngology-Head & Neck Surgery 127*(10), 1178–1180.

Davis, L. W. (2008). The effect of driving restrictions on air quality in mexico city. *Journal of Political Economy 116*(1), 38–81.

DiMaggio, P. (1997). Culture and cognition. *Annual Review of Sociology 23*, 263–287.

Durkheim, E. (1963). *Primitive Classification*. University of Chicago Press.

Eichler, H.-G., B. Bloechl-Daum, D. Brasseur, A. Breckenridge, H. Leufkens, J. Raine, T. Salmonson, C. K. Schneider, and G. Rasi (2013). The risks of risk aversion in drug regulation. *Nature Reviews Drug Discovery 12*(12), 907–916.

Epstein, S. (2007). *Inclusion: the politics of difference in medical research*. Chicago: University of Chicago Press.

Etzkowitz, H. (1983). Entrepreneurial scientists and entrepreneurial universities in american academic science. *Minerva 21*(2-3), 198–233.

Etzkowitz, H. (1998). The norms of entrepreneurial science: cognitive effects of the new university-industry linkages. *Research Policy 27*(8), 823–833.

Evans, J. A. (2010). Industry induces academic science to know less about more. *American Journal of Sociology 116*(2), 389–452.

Eyal, G. (2013). For a sociology of expertise: The social origins of the autism epidemic. *American Journal of Sociology 118*(4), 863–907.

Falagas, M. E., E. I. Pitsouni, G. A. Malietzis, and G. Pappas (2008). Comparison of pubmed, scopus, web of science, and google scholar: strengths and weaknesses. *The FASEB Journal 22*(2), 338–342.

Furman, J. L. and M. MacGarvie (2012). When the pill peddlers met the scientists: the antecedents and implications of early collaborations between us pharmaceutical firms and universities. *Essays in Economic & Business History 26*(1).

Geiling, E. M. K. and P. R. Cannon (1938). Pathologic effects of elixir of sulfanilamide (diethylene glycol) poisening clinical and experimental correlation: Final report. *Journal of the American Medical Association 111*(10), 919–926.

Gieryn, T. F. (1999). *Cultural Boundaries of Science: Credibility on the Line*. University of Chicago Press.

Goffman, E. (1986). *Stigma: Notes on the Management of Spoiled Identity* (Reissue edition ed.). New York: Touchstone.

Gould, R. V. (2002). The origins of status hierarchies: A formal theory and empirical test. *American Journal of Sociology 107*(5), 1143–1178.

Green, D. P., T. Y. Leong, H. L. Kern, A. S. Gerber, and C. W. Larimer (2009). Testing the accuracy of regression discontinuity analysis using experimental benchmarks. *Political Analysis 17*(4), 400–417.

Hannan, M. T. (2010). Partiality of memberships in categories and audiences. *Annual Review of Sociology 36*(1), 159–181.

Harbour, R. and J. Miller (2001). A new system for grading recommendations in evidence based guidelines. *BMJ : British Medical Journal 323*(7308), 334–336.

Higgins, J. P. T., D. G. Altman, P. C. GÃÿtzsche, P. JÃijni, D. Moher, A. D. Oxman, J. SavoviÄĞ, K. F. Schulz, L. Weeks, and J. A. C. Sterne (2011). The cochrane collaborationâĂŹs tool for assessing risk of bias in randomised trials. *BMJ 343*, d5928.

Howard, R. L., A. J. Avery, S. Slavenburg, S. Royal, G. Pipe, P. Lucassen, and M. Pirmohamed (2007). Which drugs cause preventable admissions to hospital? a systematic review. *British Journal of Clinical Pharmacology 63*(2), 136–147.

Hunter, D. R., M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of statistical software 24*(3), nihpa54860.

Imbens, G. and K. Kalyanaraman (2011). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*.

Inman, W. H. (1985). Under-reporting of adverse drug reactions. *British Medical Journal (Clinical research ed.) 290*(6478), 1355.

Jackson, C. O. (1970). *Food and Drug Legislation in the New Deal*. Princeton University Press.

Jenkins, C. D. (1966). Group differences in perception: A study of community beliefs and feelings about tuberculosis. *American Journal of Sociology 71*(4), 417–429.

Jonsson, S., H. R. Greve, and T. Fujiwara-Greve (2009). Undeserved loss: The spread of legitimacy loss to innocent organizations in response to reported corporate deviance. *Administrative Science Quarterly 54*(2), 195–228.

Kaitin, K. I., M. Manocchia, M. Seibring, and L. Lasagna (1994). The new drug approvals of 1990, 1991, and 1992: Trends in drug development. *The Journal of Clinical Pharmacology 34*(2), 120–127.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist 28*(2), 107–128.

Kessler, D., A. Hass, K. Feiden, M. Lumpkin, and R. Temple (1996). Approval of new drugs in the united states: Comparison with the united kingdom, germany, and japan. *JAMA 276*(22), 1826–1831.

Kim, J. W. (2012). Arbiter of science: Institutionalization and status effects in fda drug review 1990-2004. *Strategic Organization 10*(2), 128–157.

King, M. and P. Bearman (2009). Diagnostic change and the increased prevalence of autism. *International Journal of Epidemiology 38*(5), 1224–1234.

King, M. and P. Bearman (2015). Gifts and influence: Marketing regulation and the diffusion of new medications.

Kocak, Ã., M. T. Hannan, and G. Hsu (2014). Emergence of market orders: Audience interaction and vanguard influence. *Organization Studies 35*(5), 765–790.

Kogut, B. and J. M. Macpherson (2011). The mobility of economists and the diffusion of policy ideas: The influence of economics on national policies. *Research Policy 40*(10), 1307–1320.

Krahenbuhl-Melcher, A., R. Schlienger, M. Lampert, M. Haschke, J. Drewe, and S. Krahenbuhl (2012). Drug-related problems in hospitals. *Drug Safety 30*(5), 379–407.

Kuehn, B. M. (2010). Frances kelsey honored for fda legacy: award notes her work on thalidomide, clinical trials. *JAMA 304*(19), 2109–2110, 2112.

Kuhn, T. S. (2012). *The Structure of Scientific Revolutions: 50th Anniversary Edition*. University of Chicago Press.

Landrine, H. and E. A. Klonoff (1992). Culture and health-related schemas: A review and proposal for interdisciplinary integration. *Health Psychology 11*(4), 267–276.

Langley, A. (1999). Strategies for theorizing from process data. *The Academy of Management Review 24*(4), 691–710.

Lasagna, L. (1989). Congress, the fda, and new drug development: Before and after 1962. *Perspectives in Biology and Medicine 32*(3), 322–343.

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*. Harvard University Press.

Lavertu, S. and D. L. Weimer (2011). Federal advisory committees, policy expertise, and the approval of drugs and medical devices at the fda. *Journal of Public Administration Research and Theory 21*(2), 211–237.

Law, M. T. (2006). How do regulators regulate? enforcement of the pure food and drugs act, 1907-38. *Journal of Law, Economics, and Organization 22*(2), 459–489.

Lazarou, J. and B. Pomeranz (1998). Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. *JAMA 279*(15), 1200–1205.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*(2), 281–355.

Legewie, J. (2013). Terrorist events and attitudes toward immigrants: A natural experiment. *American Journal of Sociology 118*(5), 1199–1245.

Lieberson, S. (2000). *A Matter of Taste: How Names, Fashions, and Culture Change*. Yale University Press.

Liu, K.-Y., M. King, and P. S. Bearman (2010). Social influence and the autism epidemic. *American Journal of Sociology 115*(5), 1387–1434.

Luke, D. A. and K. A. Stamatakis (2012). Systems science methods in public health: Dynamics, networks, and agents. *Annual Review of Public Health 33*(1), 357–376.

MacKenzie, D. (2011). The credit crisis as a problem in the sociology of knowledge. *American Journal of Sociology 116*(6), 1778–1841.

Martin, R. M., K. V. Kapoor, L. V. Wilton, and R. D. Mann (1998). Underreporting of suspected adverse drug reactions to newly marketed (âĂIJblack triangleâĂİ) drugs in general practice: observational study. *BMJ 317*(7151), 119–120.

McArthur, L. A. (1972). The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology 22*(2), 171–193.

McClellan, M. (2007). Drug safety reform at the fda - pendulum swing or systematic improvement? *New England Journal of Medicine 356*(17), 1700–1702.

McFarland, D. A., J. Moody, D. Diehl, J. A. Smith, and R. J. Thomas (2014). Network ecology and adolescent social structure. *American Sociological Review 79*(6), 1088–1121.

Merton, R. K. (1968). The matthew effect in science the reward and communication systems of science are considered. *Science 159*(3810), 56–63.

Moffitt, S. L. (2010). Promoting agency reputation through public advice: Advisory committee use in the fda. *The Journal of Politics 72*(03), 880–893.

Morris, M., M. S. Handcock, and D. R. Hunter (2008). Specification of exponential-family random graph models: Terms and computational aspects. *Journal of statistical software 24*(4), 1548–7660.

Murphy, S. and R. Roberts (2006). âĂIJblack boxâĂİ 101: How the food and drug administration evaluates, communicates, and manages drug benefit/risk. *Journal of Allergy and Clinical Immunology 117*(1), 34–39.

Negro, G., Ã. Kocak, and G. Hsu (2010). Research on categories in the sociology of organizations. *Research in the Sociology of Organizations 31*, 3–35.

Newman, M. E. J. and M. Girvan (2004). Finding and evaluating community structure in networks. *Physical Review E 69*(2), 026113.

Olson, M. K. (1997). Firm characteristics and the speed of fda approval. *Journal of Economics & Management Strategy 6*(1), 377–401.

Olson, M. K. (2013). Eliminating the u.s. drug lag: Implications for drug safety. *Journal of Risk and Uncertainty 47*(1), 1–30.

Oreskes, N. and E. M. Conway (2010). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Publishing USA.

Owen-Smith, J. (2003). From separate systems to a hybrid order: accumulative advantage across public and private science at research one universities. *Research Policy 32*(6), 1081–1104.

Papachristos, A. V., D. M. Hureau, and A. A. Braga (2013). The corner and the crew: The influence of geography and social networks on gang violence. *American Sociological Review 78*(3), 417–447.

Parascandola, J. (1992). *The development of American pharmacology: John J. Abel and the shaping of a discipline*. John Hopkins University Press.

Peirce, C. S. and J. Jastrow (1884). On small differences in sensation. *Memoirs of the National Academy of Sciences 3*, 75–83.

Peltzman, S. (1973). An evaluation of consumer protection legislation: The 1962 drug amendments. *Journal of Political Economy 81*(5), 1049–1091.

Perlis, R. H., C. S. Perlis, Y. Wu, C. Hwang, M. Joseph, and A. A. Nierenberg (2005). Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. *American Journal of Psychiatry 162*(10), 1957–1960.

Pinch, T. and W. Bijker (1986). Science, relativism and the new sociology of technology: Reply to russell. *Social Studies of Science 16*(2), 347–360.

Pinch, T. J. and W. E. Bijker (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science 14*(3), 399–441.

Pirmohamed, M., S. James, S. Meakin, C. Green, A. K. Scott, T. J. Walley, K. Farrar, B. K. Park, and A. M. Breckenridge (2004). Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ 329*(7456), 15–19.

Podolny, J. M. (1993). A status-based model of market competition. *American Journal of Sociology 98*(4), 829–872.

Podolny, J. M. (1994). Market uncertainty and the social character of economic exchange. *Administrative Science Quarterly 39*(3), 458–483.

Podolny, J. M. (2001). Networks as the pipes and prisms of the market. *American Journal of Sociology 107*(1), 33–60.

Poluzzi, E., E. Raschi, C. Piccinni, and F. De (2012). Data mining techniques in pharmacovigilance: Analysis of the publicly accessible fda adverse event reporting system (aers). In A. Karahoca (Ed.), *Data Mining Applications in Engineering and Medicine*. InTech.

Pontikes, E., G. Negro, and H. Rao (2010). Stained red a study of stigma by association to blacklisted artists during the âĂIJred scareâĂİ in hollywood, 1945 to 1960. *American Sociological Review 75*(3), 456–478.

Proctor, R. and L. L. Schiebinger (2008). *Agnotology: The Making and Unmaking of Ignorance*. Stanford University Press.

Psaty, B. M. and S. P. Burke (2006). Protecting the health of the public - institute of medicine recommendations on drug safety. *New England Journal of Medicine 355*(17), 1753–1755.

Ray, W. A., T. M. MacDonald, D. H. Solomon, D. J. Graham, and J. Avorn (2003). Cox-2 selective non-steroidal anti-inflammatory drugs and cardiovascular disease. *Pharmacoepidemiology and Drug Safety 12*(1), 67–70.

Rees, G. (2011). âĂIJmorphology is a witness which doesnâĂŹt lieâĂİ: Diagnosis by similarity relation and analogical inference in clinical forensic medicine. *Social Science & Medicine 73*(6), 866–872.

Robb, M. A., J. A. Racoosin, R. E. Sherman, T. P. Gross, R. Ball, M. E. Reichman, K. Midthun, and J. Woodcock (2012). The us food and drug administration's sentinel initiative: Expanding the horizons of medical product safety. *Pharmacoepidemiology and Drug Safety 21*, 9–11.

Robins, G., P. Pattison, Y. Kalish, and D. Lusher (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks 29*(2), 173–191.

Rodriguez-Monguio, D. R., M. J. Otero, and J. Rovira (2003). Assessing the economic impact of adverse drug effects. *PharmacoEconomics 21*(9), 623–650.

Roehm, M. L. and A. M. Tybout (2006). When will a brand scandal spill over, and how should competitors respond? *Journal of Marketing Research 43*(3), 366–373.

Rossman, G. (2014). The diffusion of the legitimate and the diffusion of legitimacy. *Sociological Science*, 49–69.

Sackett, D. L. (2005). Evidence-based medicine. In *Encyclopedia of Biostatistics*. John Wiley & Sons, Ltd.

Salganik, M. J., P. S. Dodds, and D. J. Watts (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science 311*(5762), 854–856.

Sarkar, S. K. and P. J. de Jong (2006). Market response to fda announcements. *The Quarterly Review of Economics and Finance 46*(4), 586–597.

Schweitzer, S. O., M. E. Schweitzer, and M.-J. S.-L. Guellec (1996). Is there a u.s. drug lag? the timing of new pharmaceutical approvals in the g-7 countries and switzerland. *Medical Care Research and Review 53*(2), 162–178.

Shane, S. A. (2000). *A General Theory of Entrepreneurship: The Individual-opportunity Nexus*. Edward Elgar Publishing.

Shwed, U. (2015). Robust science: Passive smoking and scientific collaboration with the tobacco industry in the 1970s. *Sociological Science 2*, 158–185.

Shwed, U. and P. S. Bearman (2010). The temporal structure of scientific consensus formation. *American Sociological Review 75*(6), 817–840.

Smelser, N. J. (2011). *Theory of Collective Behavior*. Quid Pro Books.

Srivastava, S. B. and M. R. Banaji (2011). Culture, cognition, and collaborative networks in organizations. *American Sociological Review 76*(2), 207–233.

Stephens, T. and R. Brynner (2009). *Dark Remedy: The Impact Of Thalidomide And Its Revival As A Vital Medicine*. Basic Books.

Stirling, D. and S. Junod (2002). Arnold j. lehman. *Toxicological Sciences 70*(2), 159–160.

Strang, D. and M. Macy (2001). In search of excellence: Fads, success stories, and adaptive emulation. *American Journal of Sociology 107*(1), 147–182.

Suchman, E. A. (1964). Sociomedical variations among ethnic groups. *American Journal of Sociology 70*(3), 319–331.

Suvarna, V. (2012). Investigator initiated trials (iits). *Perspectives in Clinical Research 3*(4), 119–121.

Tatonetti, N. P., G. H. Fernald, and R. B. Altman (2012). A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. *Journal of the American Medical Informatics Association 19*(1), 79–85.

Tavory, I. and S. Timmermans (2013). A pragmatist approach to causality in ethnography. *American Journal of Sociology 119*(3), 682–714.

Tegeder, I., M. Levy, U. Muth-Selbach, R. Oelkers, F. Neumann, H. Dormann, T. Azaz-Livshits, M. Criegee-Rieck, H. T. Schneider, E.-G. Hahn, K. Brune, and G. Geisslinger (1999). Retrospective analysis of the frequency and recognition of adverse drug reactions by means of automatically recorded laboratory signals. *British Journal of Clinical Pharmacology 47*(5), 557–564.

Temple, R. and M. Himmel (2002). Safety of newly approved drugs: Implications for prescribing. *JAMA 287*(17), 2273–2275.

Timmermans, S. and M. Berg (2003). *The Gold Standard: The Challenge Of Evidence-Based Medicine* (1 ed.). Temple University Press.

Timmermans, S. and M. Buchbinder (2012). Expanded newborn screening: articulating the ontology of diseases with bridging work in the clinic. *Sociology of Health & Illness 34*(2), 208–220.

Timmermans, S. and S. Epstein (2010). A world of standards but not a standard world: Toward a sociology of standards and standardization*. *Annual Review of Sociology 36*(1), 69–89.

Timmermans, S. and A. Mauck (2005). The promises and pitfalls of evidence-based medicine. *Health Affairs 24*(1), 18–28.

Vallas, S. P. and D. L. Kleinman (2008). Contradiction, convergence and the knowledge economy: the confluence of academic and commercial biotechnology. *Socio-Economic Review 6*(2), 283–311.

Van Winkle, W., R. Herwick, H. Calvery, and A. Smith (1944). Laboratory and clinical appraisal of new drugs. *Journal of the American Medical Association 126*(15), 958–961.

Washburn, J. (2006). *University, Inc.: The Corporate Corruption of Higher Education*. New York :; New York :: Basic Books.

Wax, P. M. (1995). Elixirs, diluents, and the passage of the 1938 federal food, drug and cosmetic act. *Annals of Internal Medicine 122*(6), 456–461.

Wears, R. L. and A. W. Wu (2002). Dealing with failure: The aftermath of errors and adverse events. *Annals of Emergency Medicine 39*(3), 344–346.

Wimmer, A. and K. Lewis (2010). Beyond and below racial homophily: Erg models of a friendship network documented on facebook. *American Journal of Sociology 116*(2), 583–642.

Wuchty, S., B. F. Jones, and B. Uzzi (2007). The increasing dominance of teams in production of knowledge. *Science 316*(5827), 1036–1039.

Wysowski, D. K. and L. Swartz (2005). Adverse drug event surveillance and drug withdrawals in the united states, 1969-2002: The importance of reporting suspected reactions. *Archives of Internal Medicine 165*(12), 1363–1369.

Young, J. H. (2015). *The Medical Messiahs: A Social History of Medical Quackery in 20th Century America*. Princeton University Press.

Zuckerman, E. W. (1999). The categorical imperative: Securities analysts and the illegitimacy discount. *American Journal of Sociology 104*(5), 1398–1438.

Zuckerman, H. and R. K. Merton (1971). Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva 9*(1), 66–100.