Safety in Numbers:
Cognitive Structures in the Making of Drug Safety

Mathijs de Vaan
Department of Sociology, Columbia University, New York, NY, USA
md2999@columbia.edu

Note: This is an early draft – please do not cite or circulate without my permission.
Abstract

Prescription drugs reduce pain and save lives, but they also kill and injure hundreds of thousands of patients each year. The drugs responsible for killing and injuring most patients have typically been on the market for many years, before being pulled off. This paper develops and tests a theory that can account for the lengthy market presence of unsafe drugs. The theory builds on one of the primary data sources used to detect unsafe drugs – the set of complaints about drugs filed by patients and physicians. The main argument developed in this paper holds that in constructing their complaints, patients and physicians use different cognitive schemata for attributing an effect to a cause. These schemata are invoked when new information about the safety of drugs is released. Differences between the schemata used by patients and physicians and the fact that these schemata are temporarily invoked are sources of ambiguity for the regulator, responsible for monitoring drug safety. The paper argues 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.
Introduction

Prescription drugs improve health, reduce pain, and save lives. However, recent years have been characterized by a large volume of drug-related safety issues, leading to hundreds of thousands of Adverse Drug Reactions (ADRs) per year. These ADRs include life-threatening injuries, hospitalizations, and deaths. While the cause of some ADRs can be identified prior to approval of a drug on the market, many safety issues will only reveal themselves after the drug has already been approved. The Food and Drug Administration (FDA) is responsible for monitoring post-approval drug safety and their main goal is to take adequate action when a prescription drug is found to be unsafe (Wysowski and Swartz, 2005). Yet, it often takes a long time before drugs, that cause patients to die, are withdrawn from the market by the FDA. For example, the average drug that has been pulled off of the 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as a result of serious safety problems has been on the market for more than 10 years, while the median drug that was withdrawn is more than 6 years old.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 paper takes up that question and points to ambiguity in the data that the FDA uses to detect safety problems. The argument developed in the paper 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 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

---

1Examples of drugs that caused high numbers of ADRs include Vioxx and Avandia, two blockbuster drugs marketed by Merck & Co. and GlaxoSmithKline respectively. Vioxx is claimed to have caused 27,785 acute myocardial infarctions and sudden cardiac deaths. 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. Avandia is claimed to have caused 83,000 deaths. 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. (1.usa.gov/1B3W9Sm), presented at the “Endocrinologic and Metabolic Drugs Advisory Committee and the Drug Safety and Risk Management Advisory Committee Meeting” on July 30, 2007.

2Throughout the paper 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.
of complaints of patients and physicians is one of the richest resources for studying drug safety\textsuperscript{3}, it comes with several drawbacks. First, the data is known to be plagued by high levels of underreporting (Martin et al., 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.\textsuperscript{4} 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. In the 1950s, Ralph Smith, director of the Food and Drug Administration, claimed that “entirely safe drugs do not exist” and the claim still holds today. 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. 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.

In this article, I argue that by identifying 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 about how and when patients and physicians attribute an adverse experience to a prescription drug. For the analyses in this paper I identify 13 events – the recalls of 13 prescription drugs – and I observe how patients

\textsuperscript{3}The data have been used by hundreds of scientific studies and it is the primary resource used by the FDA to detect unsafe drugs (Wysowski and Swartz, 2005).

\textsuperscript{4}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.
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 paper 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, Kocak, and Hsu, 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 paper shows that audiences can be partitioned based on the cognitive schemata to which they are tied.

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, Otero, and Rovira, 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 paper 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\(^5\) or technical defects in cars\(^6\). Besides its substantive contribution, this paper contributes to 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 known about the audience that provides the evaluations. This paper addresses that issue.

In section 2, I describe the context of post-approval evidence production and I further 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 paper. In section 5, I provide a rationale for the analyses conducted in the paper. Section 6 to 8 present the data, the empirical strategy, and the results respectively. In Section 9, I summarize and discuss the findings.

**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 pa-

\(^5\)See for example bit.ly/1uJOZjV
\(^6\)See for example nyti.ms/1qCq43B
tients 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 earlier, 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. The FDA monitors complaint

---

7Appendix A contains a more detailed description of the methods used by the FDA to detect unsafe
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. This paper shows that this assumption does not always hold. The argument advanced in this paper 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.

**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 drugs from data in AERS.
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 paper 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 (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 orga-
nization 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 paper 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 paper 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 paper.

Figure 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. This paper suggests that an understanding of the social process

8Please note that in this second case, the number of ADRs that occurs remains unchanged.
by which effects are attributed to causes allows one to distinguish between these two explanations.

Based on research conducted in social psychology, 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 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. By the first option, 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 for increased reporting is built on the idea that an indication that others have expressed that a drug has caused an ADR might 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 paper, 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. 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

---

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.
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 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, Barabas (2007) 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 paper 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 implementation 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 fur-
ther 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\textsuperscript{10} 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 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\textsuperscript{11} 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

\textsuperscript{10}The number of days between the date at which the ADR occurred and the date that it was reported.

\textsuperscript{11}"Closely related" here refers to a co-occurrence of adverse effect $x$ and a drug $y$ that is higher than chance.
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 paper, 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.

Roadmap

The aim of the analyses in this paper 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.

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 coef-
ficient. 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 paper 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 paper 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.

Data and methods

Study population

AERS

The main data used in the paper is the FDA Adverse Event Reporting System (AERS). This reporting system is a vital source of information for the FDA as “post-marketing safety surveillance is key” to the 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 “drugcite.com”, “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 those reports 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\textsuperscript{12}. 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.

Table 1 lists the descriptive statistics for AERS. The data contains more than 3 million cases and some cases are characterized by multiple 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 many reports as patients. Some health outcomes are extremely severe, but most of the reports are accounted for by less severe cases such as “Hospitalization” and “Other”. Finally,\textsuperscript{12}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 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.

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\textsuperscript{13}, a 1989 Center for Drug Evaluation and Research publication, NDA Pipeline, and the 1999 - 2011 Drug and Biologic Approval Reports\textsuperscript{14}. 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 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

\textsuperscript{13}\url{1.usa.gov/1pCpJaZ}
\textsuperscript{14}\url{1.usa.gov/1vr3Hr}
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).

**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.\textsuperscript{15} The complete classification of Insulin Lispro in Table 2 illustrates the structure of the system.

![Table 2 about here.](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 shows the salience of disease categories. Diseases are the most prominent categories along which understanding of conditions and treatments is shaped.

\textsuperscript{15}http://www.whocc.no/atc/structure_and_principles/
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 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 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.

[Figure 2 about here.]

Drug withdrawals

The focus is on drugs recalled or withdrawn by the FDA or the manufacturer because of safety reasons. The list used in this paper is constructed by going through a record of regulatory action taken by the FDA and identifying instances in which a drug is withdrawn. The record of regulatory action is accessible through Medwatch and can be found through the FDA website. I code the day at which the FDA communicated (in an FDA talk paper or a Public Health Advisory) about the withdrawal as the day of the communication. In

\textsuperscript{16} 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.
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 contains an overview of the drugs that were withdrawn between 1997 and 2013.

[Figure 3 about here.]

**Empirical strategy**

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

\[
T_i = \begin{cases} 
1, & \text{if } x \geq c. \\ 
0, & \text{if } x < c. 
\end{cases}
\] (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\(^{17}\)

**Descriptive Statistics**

Before turning to the analyses, I briefly review a few descriptive statistics for the data analyzed in this paper. AERS is a rich and complex dataset and I therefore think that it is useful to share some of its characterizing properties. Figure 4 shows the number of reports

\(^{17}\)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 with 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.
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 4 about here.]

Table 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 about here.]

Table 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 4 about here.]

Results

The regression estimates for the first set of regressions are shown in Table 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.
Figure 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 – 80.78 – and statistically significant at the 0.01 level. This implies that – if the assumptions of the model hold – the 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 6 shows the regression lines for four control variables. The first variable, which is found in the upper left panel of the graph shows the interval 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), 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 next variable which captures the gender ratio for the reports that are filed is also continuous at the cutoff. This implies that it is unlikely that prior to the cutoff women or men were increasingly and discontinuously exposed to the prescription drug, while differing in their likelihood of being of experiencing an ADR given that he or she was administered the drug. The other two variables essentially show the same pattern: no discontinuity at the cutoff. Here, I used death proportion as an example but any other health outcome shows essentially the same trend. Again, this variable shows that it is unlikely that, prior to the withdrawal of a drug, an unobserved factor caused the population prescribed with the specific drug to change. Had the population changed it would have been unlikely that
the health outcome was continuous in the cutoff. Finally, the proportion of consumers also is continuous at the cutoff. Had the underlying population of those to which the drug was prescribed changed discontinuously prior to the withdrawal one would have likely seen a discontinuity in the trend of this variable.

[Figure 6 about here.]

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 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.

[Table 6 about here.]

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 6. 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.

[Table 7 about here.]

Figure 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 0.001, and is statistically significant at the 0.05 level. For the absolute number of reports, the effect is 1.10 and statistically significant at the 0.01 level.\(^\text{18}\) Although the estimates for

\(^{18}\text{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...}

29
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 100%. So, besides its statistical significance, the effect of a communication of a withdrawal seems to be economically salient too. These results conform hypothesis 2.

[Figure 7 about here.]

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

[Figure 8 about here.]

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. 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.

[Table 8 about here.]

**Discussion and conclusion**

The results presented in this paper 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

100 placebo events and checked how the effects lined up. They suggest that the effect shown in Table 7 is exceptional and is unlikely to stem from random variation only. More details on the estimations can be found in Appendix B.
withdrawn and for drugs that treat the same disease as the drug that was originally with- 
drawn. The increase in the number of reports filed for drugs that treat the same disease 
is accounted for by additional reports from physicians, not patients. Moreover, the analy-
ses presented in the paper 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 paper 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 dispro-
portionality 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 paper is that the increase of 100% 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 addi-
tional reports, thereby removing ambiguity in the data.

More broadly, the findings presented in this paper are especially salient given the promi-
nence of Evidence Based Medicine (EBM) in the medical discipline. EBM has been a dom-
inant 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 pro-
cesses should be accounted for when identifying best practices.

This paper 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 paper 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 paper 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 legitimize the contribution of an evaluation. If patients and physi-
cians 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 paper 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 paper 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 answer 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 paper 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.

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
This paper 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.

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.
References


Eichler, H.-G., B. Bloechl-Daum, D. Brasseur, A. Breckenridge, H. Leufkens, J. Raine,


Rees, G. (2011). morphology is a witness which doesn't lie: Diagnosis by similarity relation and analogical inference in clinical forensic medicine. Social Science & Medicine 73(6), 866–872.


Figures

Figure 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.
Figure 2: The Network of Drugs that Treat the same Disease
Figure 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.
Figure 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.
Figure 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.
Figure 6: RD controls graphs

Note: Each panel represents one control variable. The upper left panel plots the daily reporting interval (the number of days between the event date and the reporting date), the upper right panel plots the daily proportion of females, the lower right panel shows the daily proportion of reports that indicate death as the outcome of the adverse event, and the lower left panel plots the daily proportion of consumers in the total number of daily reports. The calculations are based on data for the withdrawn drugs only. The regression lines are estimated using three polynomials. Additional tests - using linear regression to as many as 10 polynomials - show that varying the number of polynomials does not affect the interpretation of the analysis.
Figure 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 8: RD controls graphs

Note: Each panel represents one control variable. The upper left panel plots the daily reporting interval (the number of days between the event date and the reporting date), the upper right panel plots the daily proportion of females, the lower right panel shows the daily proportion of reports that indicate death as the outcome of the adverse event, and the lower left panel plots the daily proportion of consumers in the total number of daily reports. The calculations are based on data for the withdrawn drugs only. The regression lines are estimated using three polynomials. Additional tests - using linear regression to as many as 10 polynomials - show that varying the number of polynomials does not affect the interpretation of the analysis.
Tables

Table 1: Descriptive Statistics for AERS

<table>
<thead>
<tr>
<th>Database Size</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique Cases</td>
<td>3,257,695</td>
</tr>
<tr>
<td>Unique Reports</td>
<td>4,184,706</td>
</tr>
</tbody>
</table>

Demographic Data

<table>
<thead>
<tr>
<th>Gender</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>2,414,006 (57.7%)</td>
</tr>
<tr>
<td>Male</td>
<td>1,474,466 (35.2%)</td>
</tr>
<tr>
<td>Mean Age (SE)</td>
<td>52.29 (0.012)</td>
</tr>
</tbody>
</table>

Reporter Type

<table>
<thead>
<tr>
<th>Type</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer</td>
<td>1,296,218 (31%)</td>
</tr>
<tr>
<td>Physician</td>
<td>758,972 (18.1%)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>188,199 (4.5%)</td>
</tr>
<tr>
<td>Other Health Professional</td>
<td>453,935 (10.8%)</td>
</tr>
</tbody>
</table>

Health Outcome

<table>
<thead>
<tr>
<th>Outcome</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>438,511 (8.9%)</td>
</tr>
<tr>
<td>Life-Threatening</td>
<td>170,868 (3.5%)</td>
</tr>
<tr>
<td>Disability</td>
<td>164,125 (3.3%)</td>
</tr>
<tr>
<td>Hospitalization</td>
<td>1,225,324 (24.9%)</td>
</tr>
<tr>
<td>Other</td>
<td>1,664,297 (33.8%)</td>
</tr>
</tbody>
</table>

Missing Data

<table>
<thead>
<tr>
<th>Missing Data</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmatched Drug Names (PS)</td>
<td>109,249 (2.6%)</td>
</tr>
<tr>
<td>Unmatched ATC Codes (PS)</td>
<td>120,257 (2.9%)</td>
</tr>
<tr>
<td>Gender Missing</td>
<td>296,234 (7.1%)</td>
</tr>
<tr>
<td>Age Missing</td>
<td>1,436,473 (34.3%)</td>
</tr>
<tr>
<td>Reporter Type Missing</td>
<td>1,308,949 (31.3%)</td>
</tr>
<tr>
<td>Health Outcome Missing</td>
<td>1,255,407 (25.5%)</td>
</tr>
</tbody>
</table>

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. Since each report represents a unique event, 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. The total number of drug fields is given by aggregating all unique standardized drug fields over all unique reports. The count of unmatched drug names or unmatched ATCs includes drug fields that are no drugs or drug fields that are ambiguous. Examples of such cases include “allergy medication” and “whole egg”.
<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1</strong></td>
<td>A - Alimentary tract and metabolism - anatomical main group</td>
</tr>
<tr>
<td><strong>Level 2</strong></td>
<td>A10 - Drugs used in diabetes - therapeutic subgroup</td>
</tr>
<tr>
<td><strong>Level 3</strong></td>
<td>A10A - Insulins and analogues - pharmacological subgroup</td>
</tr>
<tr>
<td><strong>Level 4</strong></td>
<td>A10AB - Insulins and analogues for injection, fast-acting - chemical subgroup</td>
</tr>
<tr>
<td><strong>Level 5</strong></td>
<td>A10AB04 - Insulin Lispro - chemical substance</td>
</tr>
</tbody>
</table>
Table 3: Event - report interval per day

<table>
<thead>
<tr>
<th>Day</th>
<th>Min</th>
<th>Max</th>
<th>Median</th>
<th>Mean</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>0</td>
<td>1533</td>
<td>64</td>
<td>207.08</td>
<td>0.42</td>
</tr>
<tr>
<td>Tuesday</td>
<td>0</td>
<td>1533</td>
<td>63</td>
<td>199.51</td>
<td>0.42</td>
</tr>
<tr>
<td>Wednesday</td>
<td>0</td>
<td>1533</td>
<td>60</td>
<td>195.33</td>
<td>0.43</td>
</tr>
<tr>
<td>Thursday</td>
<td>0</td>
<td>1533</td>
<td>59</td>
<td>198.79</td>
<td>0.45</td>
</tr>
<tr>
<td>Friday</td>
<td>0</td>
<td>1533</td>
<td>58</td>
<td>196.71</td>
<td>0.46</td>
</tr>
<tr>
<td>Saturday</td>
<td>0</td>
<td>1533</td>
<td>30</td>
<td>141.75</td>
<td>1.35</td>
</tr>
<tr>
<td>Sunday</td>
<td>0</td>
<td>1533</td>
<td>24</td>
<td>132.95</td>
<td>1.55</td>
</tr>
</tbody>
</table>
Table 4: Reports per day

<table>
<thead>
<tr>
<th>Day</th>
<th>Mean</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>1081.83</td>
<td>20.66</td>
</tr>
<tr>
<td>Tuesday</td>
<td>1077.57</td>
<td>17.74</td>
</tr>
<tr>
<td>Wednesday</td>
<td>999.79</td>
<td>15.85</td>
</tr>
<tr>
<td>Thursday</td>
<td>952.1</td>
<td>16.41</td>
</tr>
<tr>
<td>Friday</td>
<td>882.47</td>
<td>15.59</td>
</tr>
<tr>
<td>Saturday</td>
<td>68.05</td>
<td>1.84</td>
</tr>
<tr>
<td>Sunday</td>
<td>51.91</td>
<td>1.53</td>
</tr>
</tbody>
</table>

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.
Table 5: Point estimate for treatment effect on focal drug using Poisson regression

<table>
<thead>
<tr>
<th>Interval width</th>
<th>Point estimate</th>
<th>Lower CB</th>
<th>Upper CB</th>
</tr>
</thead>
<tbody>
<tr>
<td>± 7 days interval</td>
<td>2.54</td>
<td>2.43</td>
<td>2.64</td>
</tr>
<tr>
<td>± 14 days interval</td>
<td>2.55</td>
<td>2.47</td>
<td>2.64</td>
</tr>
<tr>
<td>± 21 days interval</td>
<td>2.57</td>
<td>2.50</td>
<td>2.64</td>
</tr>
<tr>
<td>± 28 days interval</td>
<td>2.28</td>
<td>2.22</td>
<td>2.34</td>
</tr>
<tr>
<td>± 35 days interval</td>
<td>2.19</td>
<td>2.13</td>
<td>2.24</td>
</tr>
<tr>
<td>± 42 days interval</td>
<td>2.16</td>
<td>2.11</td>
<td>2.21</td>
</tr>
</tbody>
</table>

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.
Table 6: RD Estimates

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>SE</th>
<th>Z-score</th>
<th>P-value</th>
<th>CI Lower</th>
<th>CI Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthcare Provider - ratio</td>
<td>0.03</td>
<td>0.01</td>
<td>2.17</td>
<td>0.03</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Healthcare Provider - absolute</td>
<td>3.47</td>
<td>1.74</td>
<td>1.99</td>
<td>0.049</td>
<td>0.14</td>
<td>7.07</td>
</tr>
<tr>
<td>Patient - ratio</td>
<td>0.07</td>
<td>0.03</td>
<td>2.42</td>
<td>0.02</td>
<td>0.01</td>
<td>0.12</td>
</tr>
<tr>
<td>Patient - absolute</td>
<td>7.24</td>
<td>3.26</td>
<td>2.22</td>
<td>0.03</td>
<td>0.85</td>
<td>13.62</td>
</tr>
</tbody>
</table>

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.
Table 7: Point estimate for treatment effect on related drugs using Poisson regression

<table>
<thead>
<tr>
<th>Interval width</th>
<th>Point estimate</th>
<th>Lower CB</th>
<th>Upper CB</th>
</tr>
</thead>
<tbody>
<tr>
<td>± 7 days interval</td>
<td>0.42</td>
<td>0.36</td>
<td>0.48</td>
</tr>
<tr>
<td>± 14 days interval</td>
<td>0.33</td>
<td>0.28</td>
<td>0.37</td>
</tr>
<tr>
<td>± 21 days interval</td>
<td>0.22</td>
<td>0.18</td>
<td>0.26</td>
</tr>
<tr>
<td>± 28 days interval</td>
<td>0.07</td>
<td>0.03</td>
<td>0.10</td>
</tr>
<tr>
<td>± 35 days interval</td>
<td>0.13</td>
<td>0.10</td>
<td>0.16</td>
</tr>
<tr>
<td>± 42 days interval</td>
<td>0.15</td>
<td>0.12</td>
<td>0.18</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>SE</th>
<th>Z-score</th>
<th>P-value</th>
<th>CI Lower</th>
<th>CI Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthcare Provider - ratio</td>
<td>0.01</td>
<td>0.00</td>
<td>2.77</td>
<td>0.01</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>Healthcare Provider - absolute</td>
<td>0.97</td>
<td>0.36</td>
<td>2.70</td>
<td>0.01</td>
<td>0.26</td>
<td>1.67</td>
</tr>
<tr>
<td>Patient - ratio</td>
<td>0.00</td>
<td>0.00</td>
<td>1.46</td>
<td>0.14</td>
<td>-0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Patient - absolute</td>
<td>0.80</td>
<td>0.59</td>
<td>1.35</td>
<td>0.18</td>
<td>-0.36</td>
<td>1.97</td>
</tr>
</tbody>
</table>

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.
Appendices

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>
<thead>
<tr>
<th>Adverse Event (_i)</th>
<th>Drug (_i)</th>
<th>Drug (_i^C)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>( B )</td>
<td>( A + B )</td>
<td></td>
</tr>
<tr>
<td>( C )</td>
<td>( D )</td>
<td>( C + D )</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>( A + C )</td>
<td>( B + D )</td>
<td>( A + B + C + D )</td>
</tr>
</tbody>
</table>

\[ A = \text{number of reports where suspect drug is Drug}_i \text{ and adverse event is Adverse Event}_i \]
\[ B = \text{number of reports where suspect drug is Drug}_i \text{ and adverse event is Adverse Event}^C_i \]
\[ C = \text{number of reports where suspect drug is Drug}^C_i \text{ and adverse event is Adverse Event}_i \]
\[ D = \text{number of reports where suspect drug is Drug}^C_i \text{ and adverse event is Adverse Event}^C_i \]

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 (roughly divided into frequentist versus 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. Essentially, 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 then 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.

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 simulate - 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 paper. 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.