NBER WORKING PAPER SERIES

MARKET EFFECTS OF ADVERSE REGULATORY EVENTS: EVIDENCE FROM DRUG RELABELING

Matthew J. Higgins
Xin Yan
Chirantan Chatterjee

Working Paper 24957
http://www.nber.org/papers/w24957

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2018, Revised May 2019

We thank Maria Arbatskya, Lee Branstetter, Dana Goldman, Dietmar Harhoff, David Howard, Darius Lakdawalla, Sara Markowitz, John Romley, Neeraj Sood and seminar participants at Emory University, University of Michigan, Max Planck Institute for Innovation and Competition and University of Southern California for helpful comments and suggestions. We thank IMS Health Incorporated, now IQVIA, for their generous support and access to their data. The statements, findings, conclusions, views, and opinions contained and expressed herein are not necessarily those of IQVIA, IMS Health Incorporated or any of its affiliated or subsidiary entities. The statements, findings, conclusions, views, and opinions contained and expressed in this article are based in part on data obtained under license from the following IMS Health Incorporated or affiliate information service(s): IMS Midas™. The usual disclaimers apply and any remaining errors are our own. Chatterjee acknowledges the ICICI Bank Chair in Strategic Management, IIM Ahmedabad and the Campbell and Edward Teller National Fellow Program, Hoover Institution, Stanford University for supporting this research. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Market Effects of Adverse Regulatory Events: Evidence from Drug Relabeling
Matthew J. Higgins, Xin Yan, and Chirantan Chatterjee
NBER Working Paper No. 24957
August 2018, Revised May 2019
JEL No. I18,L51,L65

ABSTRACT

We provide causal evidence that regulatory shocks associated with drug safety label changes lead to aggregate demand declines of 16.9 percent within two years of a relabeling event. After accounting for all plausible substitution patterns by physicians along with competitor actions, aggregate demand declines by 4.7 percent; this decline represents consumers that prematurely leave the market. Results are robust to variation across types of relabeling, market sizes, and levels of competition. Our findings complement recent work that shows negative upstream innovation impacts from these downstream regulatory shocks. Importantly, drugs receiving expedited FDA review are more likely to incur serious safety label changes. Thus, it appears we may be trading off quicker access to new drugs today for less innovation tomorrow. Implications for welfare and policy are discussed.

Matthew J. Higgins
Scheller College of Business
Georgia Institute of Technology
800 West Peachtree Street
Atlanta, GA 30308
and NBER
matt.higgins@scheller.gatech.edu

Chirantan Chatterjee
Indian Institute of Management Ahmedabad
Room 15F, Wing 15, Heritage Campus
Ahmedabad
India
chirantan@gmail.com

Xin Yan
Emory University
Department of Economics
1602 Fishburne Drive
Atlanta, GA 30307
xin.yan@emory.edu

1.0 Introduction

The drug development process is long and expensive with a low probability of receiving Food and Drug Administration (FDA) approval (Wong *et al*., 2018).[1] As part of the approval process drug candidates undergo clinical trials designed to test their safety and efficacy. These studies culminate with large-scale, randomized clinical trials to test a drug candidate's effectiveness before being submitted to the FDA for approval. In the post-approval period the FDA maintains a surveillance program that continues to monitor drugs. The FDA Adverse Events Reporting System (FAERS) database was designed to collect complaints and adverse events for approved drugs. Depending on the situation and severity of the data collected the FDA will act and move to change the safety label associated with a drug. The regulatory process underlying these safety label changes or relabeling is well defined, however, the overall impacts on aggregate demand is under investigated.

While prior studies have focused on various types of relabeling (*e.g*., Macher and Wade, 2016; Qureshi *et al.*, 2011; Dorsey *et al*., 2010), most have limited their analyses to a single or small number of therapeutic markets (*e.g*., Olfson *et al.,* 2008; Jacoby *et al.*, 2005). These studies are important because we learn about the intricacies and nuances of specific markets but we are unable to draw conclusions about the overall impact of relabeling across markets. Using a dataset of all drugs sold across all therapeutic markets in the U.S. and U.K. we use a difference-in-differences (diff-in-diffs) specification to provide causal evidence relating the impacts of FDA drug relabeling on aggregate consumer demand.[2] We find that, on average, aggregate demand declines by 16.9 percent within two years of a relabeling event. Our data allows us to capture intra- and inter-market substitution patterns as well as competitive responses. Critically, after accounting for these factors we still find that aggregate demand declined by 4.7 percent, an estimate that represents consumers that prematurely leave the market.[3]

[1] Wong *et al* (2018) place the probability of a drug candidate reaching FDA approval at 13.8 percent.
[2] For ease of exposition we use the term aggregate consumer demand interchangeably with demand. To be precise we are referring to aggregate consumer demand. Our data is at the standard unit level and not at the individual prescription level. Standard units are determined by IMS Health and are intended to equate pills, tablets and liquids.
[3] We engage in back of the envelope calculations to translate these changes to represent numbers of prescriptions and consumers. As a conservative lower bound, we can assume that the entire decline, after accounting for substitution patterns, represents consumers that are chronic patients. This allows us to transform the decline in aggregate demand into monthly prescriptions and as a result an estimate of chronic patients. Clearly, all conditions are not chronic so as the number of acute prescriptions increases in the sample so will the number of consumers that plausibly leave the market.

Next, we explore the variation across the severity of relabeling events. Not unexpectedly, we find an increasing aggregate demand response as relabeling severity increases, ranging from a decline of 15.6 percent for the least severe to a decline of 36.3 percent for the most severe. After accounting for all plausible substitution patterns we find declines in aggregate demand ranging from 4.0 percent for the least severe to 8.3 percent for the most severe relabel. Again, these estimates can be viewed as consumers that prematurely leave the market. This pattern, however, is not homogenously distributed across all markets. When we focus on the variation in relabeling activity across individual markets, interesting patterns begin to emerge. In "low-intensity markets" or those with low levels of relabeling activity we find declines in focal drug aggregate demand are completely absorbed by intra-market substitution. In contrast, in "high-intensity markets" or those with high levels of relabeling activity, after accounting for plausible substitution patterns, we find that consumers leave the market.

Our findings have implications for firms. Our results for focal firms suggest that their current efforts to counteract the impacts from these negative product shocks, on average, appear to be failing. Importantly, the magnitude of our results for relatively minor safety relabeling suggests that physicians may be proactively shifting consumers to other drugs. This implies that while detailing (*i.e.*, direct advertising to physicians) may be effective at influencing initial physician prescription behavior (*e.g.*, Datta and Dave, 2017), this influence appears to break down when confronted with negative safety information. Unfortunately, while we can detect the shift in behavior we can only conjecture on the underlying motivations driving physician behavior. Explanations range from physicians practicing "defensive medicine", being concerned that less serious safety concerns will eventually unmask more serious concerns ("where there is smoke, there is fire"), lack of adequate information, or being induced by competitor firm detailing efforts (Macher and Wade, 2016).[4]

Unfortunately, the effects on firms are not isolated to just the downstream aggregate demand impacts that we document but also extend upstream. In a recent paper, Krieger *et al.* (2018) explore how pharmaceutical firms react to negative shocks to existing products. Specifically, they look at how firms respond to FDA Public Health Advisories, which include the

[4] Current work by the authors involve a large-scale survey with a national association of physicians to understand what drives prescription changes in the face of negative safety-related information. Preliminary, qualitative evidence seems to suggest some combination of defensive medicine and marketing efforts by competitors - consist with those described in Macher and Wade (2016).

relabeling of drugs. They show that affected firms *increase* R&D expenditures but those expenditures are more likely to go towards the acquisition of new pipeline candidates versus internally developed candidates.[5] Importantly, they also show that competitors move resources away from affected therapeutic categories, reshuffling their own drug portfolios. Our findings and those of Krieger *et al.* (2018) are intimately linked; we provide evidence of the initial downstream aggregate demand impacts from negative regulatory shocks while they provide evidence of subsequent upstream innovation changes.[6]

These results suggest there are plausible welfare implications. If consumers that leave the market should be treated, then this shift to the non-treated population could be a detriment to welfare. Moreover, if consumers remain treated but are switched to drugs that are less effective, this will again be a detriment to welfare. On the other hand, it is widely believed that some drugs are overprescribed (Lembke *et al*., 2018; Sacarny *et al*., 2016; Forgacs, 2008; Price *et al*., 1986). If it is these consumers that exit the market then the impact on welfare may be dampened. Additionally, as Krieger *et al*. (2018) demonstrates there are also potential innovation impacts to consider. Whether the change in focal innovation and churn in competitor portfolios is a positive or negative for welfare is still unknown. However, what we can say definitively is that the downstream impacts that we document were significant enough to elicit an upstream response.

More broadly, our findings contribute empirical evidence to the economics of regulation literature dating back to Brown *et al*. (1964), Nelson (1970) and Stigler (1971). The breadth and depth of our data allow for a unique analysis of aggregate demand that captures all plausible substitution patterns. Regulation is rarely without cost, as is the case here. Importantly, this is a market that is under immense time constraints given the limited nature of patent protection. Regulators therefore face a tension between length of trials and getting new drugs to market. Into this mix the FDA has developed pathways for expedited development and review including priority review, breakthrough therapy, accelerated approval and fast track.[7] Recent evidence suggests, however, that there has been an increase in safety label changes for drugs that have

[5] The relationship between product or pipeline shocks and subsequent technology acquisition in the pharmaceutical industry was previously considered in Higgins and Rodriguez (2006), Danzon *et al.* (2007) and Chan *et al.* (2007). The importance of Krieger *et al.* (2018), however, is they provide causal evidence of this relationship.

[6] The linkage of downstream product shocks and upstream innovation has been explored in other contexts. For example, Ball *et al.* (2018) examines the upstream innovation impacts as a result of downstream medical device recalls.

[7] https://www.fda.gov/forpatients/approvals/fast/default.htm. These are coupled with other initiatives such as the 21st Century Cures Act.

moved through some form of expedited pathway (Mostaghim *et al*., 2017; Moore and Furberg, 2014; Carpenter *et al*., 2008). These label changes are not trivial; Mostaghim *et al.* (2017) report a doubling of the most severe types of relabel for expedited drugs relative to non-expedited drugs. With impacts from safety label changes rippling both downstream and upstream, it suggests that regulators may have tipped the balance too far towards getting new drugs to market. More fundamentally, our results combined with those of Krieger *et al*. (2018) suggest that we may be trading off quicker access to new drugs today for less innovation tomorrow.[8]

The remainder of the paper is organized as follows. In Section 2.0 we discuss the FDA drug relabeling process and in Section 3.0 we focus on adverse regulatory events. This is followed by our empirical strategy and data in Section 4.0. Results and robustness are reported in Sections 5.0 and 6.0, respectively, before we conclude in Section 7.0

2.0 FDA drug relabeling

The pharmaceutical industry in the U.S. is highly regulated and drug candidates undergo rigorous clinical testing prior to being submitted to the FDA for approval. During this rigorous process possible risks and side effects of a drug candidate are identified. This information becomes part of the FDA approved label and drug insert that accompanies a newly approved drug. Unfortunately, some side effects do not become known until after a drug has been approved. To help with the reporting and collection of these adverse events the FDA founded MedWatch in 1993. Healthcare professionals or consumers (patients) can voluntary report to Medwatch. In more recent times this adverse events data has been made available via FAERS.[9]

During the post-approval time period the FDA monitors adverse reporting along with results from post-approval studies and peer-reviewed literature. Negative safety-related information is scrutinized and the FDA can form an investigation team to determine if a safety label update is needed. If they believe a safety label change is warranted the manufacturer is notified and is required to report back to the FDA within a predetermined period. The agency works privately with a manufacturer to determine which type of safety label change will be made. At the end of the process the FDA will publish this information online while allowing

[8] We must include the caveat that while there may be less innovation tomorrow, we cannot say anything about the type or novelty of the lost innovation. In a Health Affairs blog post, Aurora *et al* (2016) conjectured about possible innovation implications from priority review vouchers.
https://www.healthaffairs.org/do/10.1377/hblog20160615.055372/abs/

[9] http://www.nber.org/data/fda-adverse-event-reporting-system-faers-data.html.

firms additional time to change actual printed material.[10] Prior to 2016 product safety data was available via MedWatch but has since shifted to the FDA Drug Safety Label Change database.

The main safety labeling changes that the FDA issues include: *adverse reaction*, *precaution*, *warning*, *contraindication*, and *box warning*.[11] These classifications serve to inform physicians and consumers of possible health concerns that have been clinically identified, anticipated to occur, or associated with unapproved uses. A box or "black box" warning is the most severe of type of label change and is intended to communicate potentially severe health risks resulting from taking a drug. For example, in 2004 the FDA issued box warnings for all anti-depressant drugs over the concern that these drugs could lead to suicide in patients younger than 18 years of age.[12] While these changes can be sensational, most changes are much less so. For example, Topomax® was the target of a "precaution and warning" label change due to an elevated risk of kidney stones.[13] As this example demonstrates, drugs can undergo several types of relabeling simultaneously. Additionally, a drug that has been relabeled can undergo additional safety label changes in the future, if warranted.

While the FDA's procedure and process for a drug safety relabel is well established, there is no guarantee that the updated information will be read by physicians or consumers. In a world of perfect-information we might assume that this new information will be read, however, the evidence seems to suggest otherwise. One form of communication that firms use to convey new safety information, the "Dear Doctor letter" (DDL) was found in 28% of cases to be deficient in their overall level of effectiveness (Mazor *et al*., 2005).[14] Other studies have shown that fewer than one in ten physicians routinely read drug labels.[15] Similarly, Hoy and Levenshus (2018) find that consumers routinely ignore safety related information.

Macher and Wade (2016) shed an important light on the underlying mechanism of how physicians may be learning about safety label changes. In the case of black-box warnings, they find that affected firms themselves may increase physician detailing (*i.e.*, direct-to-physician advertising) but they also find that competitor firms also increase detailing efforts. So while

[10] https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm250783.pdf
[11] Guidance for industry is provided by: https://www.fda.gov/downloads/drugs/guidances/ucm075096.pdf
[12] https://www.medpagetoday.com/psychiatry/depression/210
[13] https://www.accessdata.fda.gov/scripts/cder/safetylabelingchanges/index.cfm?event=searchdetail.page&DrugNameID=1063
[14] Attempts to improve the effectiveness of labels is on-going with draft guidance as recently as 9 July 2018 intended to assist applicants in writing drug labels: https://www.policymed.com/2018/07/fda-releases-draft-guidance-on-indications-and-usage-labeling-sections.html
[15] https://www.nytimes.com/2006/01/19/us/new-drug-label-rule-is-intended-to-reduce-medical-errors.html

affected firms may try to actively deal with the problem, it does appear that competitors take advantage of the opportunity and try to pull physicians to their products. While this study is limited to black-box warnings, there is no reason to believe that there wouldn't some type of similar response across the spectrum of relabeling activity.

There appears to be some hope in that new technology may be able to help with this information asymmetry. In a recent paper, Arrow *et al* (2017) show that physicians with access to a drug reference database changed their prescribing behavior. While this study focused on the shift from branded to generic drugs, they suggest that physicians may be responding to non-clinical information such as whether a drug is covered by a consumer's insurance plan or plan-specific drug pricing. The database in this study included FDA drug safety information and alerts but the information was not explicitly analyzed. However, given the fact that physicians were taking the time to interact with this type of technology does suggest that it may be an effective mechanism to deliver safety relabel information. A major concern with new technology is ensuring that it diffuses out to physicians, especially those practicing in non-academic or rural settings.

3.0 Adverse regulatory events

We draw on several strands of literature starting with the economics of regulation. Early work in this area theorizes on the impact of regulation on consumer and firm behavior (*e.g.*, Stigler, 1971; Peltzman, 1976; Migue, 1977). Brown *et al*. (1964) argued that regulation could be viewed as an information transmission process. As consumers receive new information they are able to update and change their behavior. Subsequent work built on this idea to show how information influences consumer perception of product quality (Zeithaml, 1988) and how behavior changes with positive information (Nelson, 1970). In contrast, Hartley (1994) showed how negative product information led to decreased sales. More broadly, Oberholzer-Gee and Mitsunari (2006) examined how non-related negative events, in their case the release of pollution information, decreased property values. In our context, the process of relabeling in the pharmaceutical industry can be viewed as an information transmission process that subsequently impacts behavior.

Our paper also draws on studies in health economics where scholars have explored the implementation of regulatory procedures on public health (*e.g.*, Gruenspecht and Lave, 2006). In the case of the pharmaceutical industry, Dranove (2011) stresses the importance of quality

certification for efficient and optimal regulation. For drugs, this certification comes in the form of the FDA approval process. This process can be divided into pre- and post-approval stages. The pre-approval stage includes clinical testing and provides the first line of defense to ensure safety and efficacy of products. This creates a tension, however, for regulators between length of trials and getting new drugs to market. For example, adverse events have been increasing (Moore *et al*, 2007) and have been associated with declines in pre-approval times (Olson, 2003). This makes post-approval safety monitoring critically important. In recent times FAERS has served as an important source of data for updating safety labeling information (Wysowski and Swartz, 2005).

Based on these prior studies, adverse safety related information should improve physician awareness about the potential safety of a drug and lead to changes in behavior. Presumably physicians (or consumers) may shift away from a drug given a safety concern. A number of studies focused on specific therapeutic markets support this association (*e.g.*, Dranove and Olsen, 1994; Smalley *et al*., 2000; Cheah *et al*., 2007; Olfson and Marcus, 2008; Tekin and Markowitz, 2008; Bunniran *et al*., 2009; Dorsey *et al*., 2010; Kales *et al.*, 2011; Dusetzina *et al*., 2012; Briesacher *et al*., 2013; Lu *et al*., 2014). Prior work also documents that this association could be differential; new drugs tend be impacted more than existing drugs (Wilkinson *et al*., 2004) and geographic variation could cause the usage of a warned drug to be different (Shah *et al*., 2010).

What remains unknown from this batch of prior work is what constitutes a rational, medically appropriate response? One might expect physicians to switch some consumers to other drugs as the severity of relabeling events increase but are all of these changes medically appropriate or is there some other underlying motivation driving the switch? Are physicians being influenced by detailing as suggested by Macher and Wade (2016), anticipating future problems (*i.e.*, "where there is smoke, there is fire"), practicing defensive medicine or responding to consumer concerns? Similarly, do consumers just seamlessly switch to different drugs or might they decide to stop treatment altogether and leave the market?

Prospect theory (Tversky and Kahneman, 1992; Kahneman and Tverksy, 1979) provides a behavioral explanation as to why physicians may switch consumers to other drugs. While a safety relabel is serious it is not necessarily relevant for all consumers, in all situations. However, physicians may overestimate the probability of a negative event occurring an incorrectly switch a consumer to another drug. Prospect theory can also help explain why consumers may ultimately

choose to leave the market. When confronted with new information about a drug from their physicians, consumers may also vastly overestimate the probability of a negative event. As a result, they may incorrectly attribute these same negative effects to substitute drugs that physicians prescribe. If consumers make this link they may incorrectly conclude that the benefit of a new drug does not outweigh the risk and exit the market.

There is experimental evidence that supports these negative responses by consumers. For example, Bunniran *et al.* (2009) study trust and blame due to the withdrawal of pharmaceutical products as a result of safety related concerns. They found that consumers taking the withdrawn drug or those taking another drug within the same class were highly likely to blame pharmaceutical companies and the FDA. After an event trust in both institutions remained fairly low. These declines provide one plausible explanation as to why consumers may formulate and attribute the negative effects described by Tversky and Kahneman (1992) and Kahneman and Tverksy (1979) to a focal drug or a substitute. It also suggests that there are behavioral considerations at play that physicians (and regulators) need to consider when consumers get switched to a new drug due to safety-related concerns.

Regulation can also have unintended consequence from spillovers. For example, toy recalls due to safety reasons tend to cause negative industry-wide spillover effects for similar types of toys (Freedman and Lederman, 2009). In our setting, such spillovers would manifest in the drugs within the same market or related market as the focal drug that is relabeled due to a safety concern. Recent work by Krieger *et al*. (2018) supports this notion and demonstrates that there are also innovation impacts to consider. They find a decline in the total number of drugs developed in a focal area, implying that these negative shocks may slow overall innovation in a given therapeutic category. Whether this decrease in focal innovation and churn in competitor portfolios is a positive or negative for welfare remains unknown. However, the downstream impacts we consider appear significant enough to have implications for upstream innovation.

These results are concerning given the increase in drugs receiving some type of expedited review by the FDA.[16] On the one hand, it has been shown that drugs cleared via expedited review appear to offer greater quality-adjusted life years (QALYs) than those approved via

[16] These issues were discussed in a recent JAMA Forum post: https://newsatjama.jama.com/2018/05/23/jama-forum-the-risks-and-benefits-of-expedited-drug-reviews/

conventional methods (Chambers *et al*., 2017). It appears that the expedited review process has helped the FDA prioritize drugs that offer greater health gains (0.182 versus 0.003 QALYs). On the other hand, these approvals have come at a cost. The evidence appears to suggest that the drugs receiving some type of expedited review are more likely to receive some type of serious safety label change (*e.g.*, Mostaghim *et al*., 2017; Moore and Furberg, 2014; Carpenter *et al*., 2008).

By its nature, regulation should be welfare enhancing but there is evidence exists that this may not always be the case. For example, the milk industry was regulated in the 1960s but that regulation was shown to be a detriment to welfare (Kessel, 1967). Bartel and Thomas (1987, 1985) found that the Occupational Safety and Health Administration did not have a significant impact on national injury rates. Similarly, Ter-Martirosyan and Kwoka (2010) found that incentive regulation caused quality degradation in the U.S. electricity industry. In healthcare, mandatory prescription regulation has not significantly improved health outcomes (Peltzman, 1987) and in hospitals regulation of quality standards has inflated costs and diminished patient welfare (Sloan and Steinwald, 1980). Given the role that the drug relabeling process plays in helping monitor drugs in the post-approval period, it makes understanding their impacts all the more critical.

4.0 Empirical strategy and data

4.1. Empirical strategy

We exploit FDA relabeling events to estimate a diff-in-diffs specification. As we discussed above, the relabeling process involves private interactions between the FDA and a focal firm and remains unknown to consumers and physicians prior to formal action. We use two groups of observations. The first group (treated) includes drugs sold in the U.S. Because FDA relabeling events only affects drugs sold in the U.S., our treated group is exposed to treatment in the post-relabel period but not in the pre-relabel period. The second group (control) is comprised of the same exact drugs as those in the treated group but sold in the U.K. As such, we estimate the following model:

$$(1) \qquad Y_{i,t} = \alpha + \beta_1 Relabel_{i,t} + \beta_2 US_i + \beta_3 \left(Relabel_{i,t}\, x\, US_i\right) + \gamma Controls_{i,t} + \mu_i + \delta_t + \varepsilon_{i,t}$$

where $Y_{i,t}$ is demand (*i.e.*, drug sales). $Relabel_{i,t}$ is a dummy variable for the post-treatment period represented by drug relabeling events and captures aggregate factors that would cause changes in $Y_{i,t}$ even in the absence of the treatment. US_i is a dummy variable and captures possible

differences between the treatment and control groups. We include a variety of controls, discussed below, as well as drug-level (μ_i) fixed effects to control for time-invariant heterogeneity between drugs and year fixed effects (δ_t) to control for common shocks impacting all drugs across time. This base specification is estimated at differing levels of aggregation so μ_i will also represent therapeutic market fixed effects. The coefficient of interest across all models is β_3 and it represents the impact induced by drug relabeling events on U.S. drugs relative to U.K. drugs.

Our identification strategy relies on the fact that the control group is not exposed to treatment in either period. Importantly, the FDA does not have regulatory jurisdiction over drugs sold in the U.K.[17] This can be visually shown in Figure 1 where the pre-trends do not appear to violate the parallel trend assumption. To test the parallel trend assumption more formally we take our pre-trend data and split it in half, defining the midpoint as an arbitrary treatment event and estimating our diff-in-diffs specification. If the parallel trend assumption is violated the coefficient $\beta_3$will be statistically significant. The results for this placebo test are reported in Appendix Table 1. The coefficient of interest is not statistically significant across any model or level of analysis. Combined, the visual evidence along with these placebo test results suggest that the parallel trends assumption is not violated.

4.2. Data

Our sample consists of all drugs sold in both the U.S. and U.K. during 2003 to 2009 as identified by IMS MIDAS™. Relabeling data for drugs sold in the U.S. was collected from the FDA MedWatch database and we restricted the data to those drugs that experienced a first-instance of a drug relabel.[18] Relabeling data for drugs sold in the U.K. was gathered from Datapharm's electronic Medicines Compendium that covers all drugs approved by the U.K. Medicines and Healthcare Products Regulatory Agency (MHRA).[19] In order to create a clean control group we further restricted our treated drugs to include only those that experienced a relabel in the U.S. but no relabel in the U.K. within eight quarters of the U.S. relabel. Table 1 provides the distribution of relabel activity in the U.S. and U.K. For those drugs that were

[17] While other countries in the E.U. could have been chosen for a control sample, the U.K. was chosen for reasons of common language and legal frameworks.

[18] It is possible to have multiple different types of relabeling activity at the same time. This is not a concern for our baseline models. However, when we examine the variation across types of relabeling activity we include those observations in each type of relabeling activity. We focus on four types of relabeling events: precaution, adverse reaction, warning and box warning. There was only one first-instance of a contraindication that met our sample criteria. It was excluded from the final sample; our results do not change with this exclusion.

[19] https://www.medicines.org.uk/emc/

subsequently relabeled in the U.K. the average time until relabel was 12.95 months after the relabel event in the U.S. This was shorter than 18.5 months documented by Pfistermeister *et al*. (2013) for a limited sample of psychiatric drugs.[20] Importantly, we could find no evidence that drug relabeling in the U.S., on average, systematically impacted contemporaneous physician prescription patterns in the U.K. (Figure 1). This further validates our U.K. sample as a clean control for causal estimates in our study.

Next, we gathered quarterly drug-level sales, detailing (*i.e.*, direct to physician promotions), and price data from IMS MIDAS™. Sales or quantity data is standardized by IMS into a 'standard unit' that equates pills, tablets and liquids. The data for both the U.S. and U.K. includes both hospital and retail channels. IMS MIDAS™ includes all branded and generic drugs and covers every therapeutic category. Detailing or direct-to-physician promotion data is available for all approved drugs. Financial variables from the U.K. have been converted by IMS to U.S. dollars and all financial variables have been converted to real 2009 dollars using a GDP deflator.[21] Descriptive statistics are presented in Table 2.

Note that drugs are approved for use within 4-digit anatomical therapeutic chemical (ATC) markets. The ATC classification is controlled by the World Health Organization and was designed to categorize drugs into different groups according to the organ or systems that they treat.[22] There are four different levels of classification ranging from the most aggregate (1-digit ATC) to most disaggregate (4-digit ATC). For example, the 1-digit ATC market N comprises drugs for the nervous system. Within ATC N there are seven 2-digit ATC markets that contain 19 3-digit ATC markets. Each of these 3-digit ATC markets, in turn, contains 4-digit ATC markets. An advantage of our data is that it is available at the 4-digit ATC market level and can be aggregated as needed.

4.2.1. Dependent variable

Our baseline dependent variable is focal drug sales (quantity) in standard units as determined by IMS. Sales are aggregated across varying dosages to the drug level since a relabeling event will impact the drug similarly across dosage types. We define *Sales* as the

[20] In Appendix Table 2 we extend the time frame for our baseline model from eight quarters to 12 and 16 quarters; our results remain robust to these longer time frames.
[21] It is critical to note that the price data within IMS MIDAS™ is a wholesale price. It does not include adjustments as a result of back-end rebate payments or any other discounts that may be offered to insurance or prescription benefit companies.
[22] For a more detailed discussion: https://www.whocc.no/atc_ddd_methodology/purpose_of_the_atc_ddd_system/

natural logarithm of quarterly focal drug sales plus one. In addition to the baseline focal drug level, we will consider two additional aggregate models. First, we consider sales of all drugs within a focal drug's 4-digit ATC market. These drugs can be reasonably viewed as close substitutes. For example, both anti-viral drugs Invirase® and Norvir® are contained in the 4-digit ATC market J5AE (protease inhibitors). Importantly, this aggregation allows us to capture intra-market substitution by physicians.

Second, we move up one more level of aggregation to the 3-digit ATC market. At this level of analysis we capture all drugs within multiple 4-digit ATC markets but contained within the same 3-digit ATC market.[23] For example, the two 4-digit ATC markets J5AE (protease inhibitors) and J5AF (nucleotide reverse transcriptase inhibitors) are contained within the 3-digit ATC market J5A (direct acting antivirals). As a second example, the two 4-digit ATC markts N3AF (carboxamide derivatives) and N3AG (fatty acid derivatives) are contained within the 3-digit ATC market N3A (anti-epileptics). This level of aggregation allows us to capture inter-market substitution by physicians.[24]

4.2.2. Independent variables and controls

As indicated above, our sample includes drugs that were sold both in the U.S. and U.K. We define *U.S.* as a dummy variable that equals one if the drug was sold in the U.S., zero otherwise. In order to implement our diff-in-diffs strategy, we define a dummy variable (*Relabel*) that equals one for all observations after a drug's first relabeling event, zero otherwise. *Relabel* encompasses four types of events: precaution, adverse reaction, warning and box warning.

Prior work has demonstrated the importance of detailing on physician prescription behavior (*e.g.*, Datta and Dave, 2017; Manchanda and Honka, 2005) and reducing price elasticity (Windmeijer *et al*, 2006; Rizzo, 1999). However, contemporaneous detailing is a function of current sales, which can create a reverse causal relationship. To resolve this issue we use lagged

[23] These markets can be explored at: https://www.whocc.no/atc_ddd_index/?code=J05A.

[24] As a robustness check, and as a method to verify we have captured all reasonable substitution patterns, on average, we aggregate markets up one more level to the 2-digit ATC market. At this level of aggregation we capture all 3-digit ATC markets contained within a 2-digit ATC market. Each of those 3-digit ATC markets will include 4-digit ATC markets. For example, let's consider the 2-digit ATC market J04 (antimycobaterials). It contains two 3-digit ATC markets, J04A (drugs for treatment of tuberculosis) and J04B (drugs for treatment of lepra). The 3-digit ATC market J04A contains six 4-digit ATC markets: J04AA (aminosalicylic acid and derivatives), J04AB (antibiotics), J04AC (hydrazides), J04AD (thiocarbamide derivatives), J04AK (other drugs for the treatment of tuberculosis), and J04AM (combinations of drugs for the treatment of tuberculosis). The 3-digit ATC market J04B contains one 4-digit ATC market, J04BA (drugs for the treatment of lepra). Like our 3-digit ATC market level of analysis this 2-digit ATC market level of analysis can also be viewed as capturing inter-market substitution.

promotion stock as studies have shown that promotions have a carry-over effect (*e.g.*, Zhao *et al*, 2011). Importantly, prior promotion expenditures should not be impacted by contemporaneous sales. As such we define *Lagged promotion stock* as the discounted sum of the prior three quarters detailing expenditures. We follow the literature (Leone, 1995) and use a 70 percent discount rate, however our baseline results are not sensitive to inclusion or variation in the discount rate.[25]

Focusing on black-box warnings in the Type-2 diabetes market Macher and Wade (2016) found that affected firms took strategic actions with respect to promotions to mitigate losses from the relabeling event. *Lagged promotion stock* in the focal drug-level models will capture these effects. They also found that competitors take advantage of these adverse events by increasing promotion activity in order to try to steal market share. *Lagged promotion stock* in the aggregated models at the 4-digit and 3-digit ATC market-level will control for these competitive dynamics. These latter two models will also capture and control for any affected firm promotion response.

Next, we control for several drug and market characteristics that may influence sales or demand. First, we define *Vintage* as a measure of elapsed time, in quarters, from introduction. Drugs that have been on the market longer have time to build up brand loyalties with consumers and physicians even though they may become 'outdated' as newer treatments come to market. Finally, we include count variables for the *Number of brands* and *Number of generics*. The former controls for the intra-market substitution possibilities. The latter controls for cross-molecular substitution or the insurance companies ability to attempt to influence physicians to switch patients to a generic of another branded drug within a focal drug therapeutic market (Branstetter *et al*, 2016 and 2014).[26]

4.2.3. Endogeneity of price

As indicated above, for those drugs that have multiple dosages sold by the same firm we aggregate the data together to the drug-level. We define *Price* by dividing drug-level revenues by the quantity of drugs sold. It is important to note that we are capturing wholesale price and this

[25] Following Leone (1995) we vary the discount rate between 50 and 70 percent.
[26] As an example, assume there are three branded drugs within a market, Brand A, Brand B and Brand C along with two generic drugs, Generic B and Generic C. Cross-molecular substitution refers to an insurance company trying to convince physicians to shift patients from Brand A to either Generic B or Generic C, often by providing economic incentives to consumers.

does not include any unmeasured discounting (rebates) by pharmaceutical companies, which is not currently commercially available. This price variable, however, will be highly correlated with ultimate consumer price and as such will be endogenous.[27] To address this concern we follow the literature (*e.g.*, Nevo, 2001) and use the mean and median price of other drugs in closely related markets as instruments for the focal drug's price. Specifically, we use the mean and median price of other drugs within the same 2-digit ATC market. For example, if our focal drug is a MAO-inhibitor (4-digit ATC market C02KC) we take the mean and median price of drugs in the broader 2-digit ATC market, C02 (anti-hypertensives). Drugs within the same 2-digit ATC should, on average, be correlated due to similar marginal costs but uncorrelated with the focal drug's unobserved product characteristics. The instruments pass the usual tests and are reported in the bottom panel of each table.

5.0 Empirical findings

5.1 Impact of drug relabeling on demand

In Table 3 we present empirical results from Equation 1. Model 1 presents estimates at the focal drug level, Model 2 presents estimates at the 4-digit ATC market level and Model 3 presents estimates at the 3-digit ATC market level. Model 1 can be viewed as testing the casual impact of drug relabeling on aggregate focal drug demand while Model 2 captures intra-market drug substitution. In other words, Model 2 helps us understand if physicians switch consumers to another drug in the same 4-digit ATC market. An example of such a substitution would be a switch from the anti-viral Invirase® to Norvir®. Finally, Model 3 captures inter-market drug substitution. In this case, physicians switch patients to another drug in a different 4-digit ATC market but within the same 3-digit ATC market. In the prior example, both Invirase® and Norvir® are in the 4-digit ATC market J5AE (protease inhibitors). In the current example, a physician would be switching a patient from either of those two drugs to Retrovir®, which is in the 4-digit ATC market J5AF (nucleotide reverse transcriptase inhibitors). All three drugs are treatments for HIV and both 4-digit ATC markets, J5AE and J5AF, are contained within the 3-digit ATC market J5A (direct acting antivirals).

The dependent variable across all three models is *Sales* and includes our full set of controls. In Model 1 we include drug and time fixed effects while in Models 2 and 3 we include

[27] A significant body of prior research on the pharmaceutical industry uses earlier versions of the IMS Health data that we employ here. Like us, these prior researchers do not directly observe retail sales or prices.

market and time fixed effects. *Price* is instrumented in all models and passes the usual test statistics, which are reported at the bottom of the table. Standard errors are clustered at the 2-digit ATC market level. The coefficient of interest is the interaction term (*Relabel * U.S.*); it is negative and statistically significant across all models. In Model 1 we find a 16.9 percent decline in aggregate focal drug sales caused by the first instance of a drug relabel.[28] When we aggregate within 4-digit ATC markets in Model 2 we find a 5.1 percent decline in aggregate sales. Importantly, this model accounts for demand of the focal drug that was absorbed by other drugs *within* that same 4-digit ATC market. In other words, physicians engaged in intra-market substitution and switched patients to another drug within the same therapeutic market. From the previous example, this would be a switch from Invirase® to Norvir® within the 4-digit ATC market J5AE.[29]

This is not the only substitution that can take place. It is possible that physicians can engage in inter-market substitution and switch consumers to another drug in a different 4-digit ATC market but still within the same 3-digit ATC market. Again, in the above example, this would be a switch from Invirase® (4-digit ATC market J5AE) to Retrovir® (4-digit ATC market J5AF) which are both in 3-digit ATC market J5A. In Model 3 we find a 4.7 percent decline in sales for drugs within a 3-digit ATC market that experienced a relabel. Critically, the result in Model 3 implies that after controlling for focal firm and competitor actions and capturing intra- and inter-market substitution patterns aggregate demand still declined by 4.7 percent. This decline represents consumers that fall out of the market.

It is important to recall the process that is involved with these types of substitutions. Only a physician can switch a consumer to another drug. While we can detect *ex post* that a substitution has occurred, we do not know what precipitated the move.[30] There are several possibilities. First, consumers could become informed of the relabel and push a physician to

[28] In unreported regressions we exclude *ln(Price)* and *ln(Lagged promotion stock)*, results remain consistent. In a second set of unreported regressions we include competitor promotions in Model 1 (Table 3). Again, results remain consistent.

[29] In Appendix Tables 2 and 3 we test alternative treatment periods. First, in Appendix Table 2 we consider time periods of three (Model 2) and four years (Model 3) before and after a drug relabeling. Our base model (Model 1, Table 3) is included as Model 1 for comparative purposes. Second, in Appendix Table 3 we widen the treatment window around the actual drug relabel. As a reminder, our baseline model excludes the quarter when a relabeling event occurred. In Model 1 and Model 2 we increase that exclusion to one and two quarters, respectively, before the quarter of relabel. This increase in exclusion will help if information leaks prior to announcement. All of the robustness results are consistent with our main findings in Table 3.

[30] This would require data on why physicians switched or changed a prescription.

switch them. Second, physicians could independently learn about the relabel and decide to proactively switch a consumer either for medically related reasons or for defensive medicine concerns. Third, physicians could learn about the relabel through detailing, either by the affected company or by a competitor and then decide to switch a consumer to another drug. These explanations are not mutually exclusive and there is recent evidence to support the role of detailing (Macher and Wade, 2016).[31]

Given that our data is at the standard unit level we do not know exactly how many consumers this represents because prescription patterns will differ across drugs and consumers. We can, however, calculate a conservative, lower bound if we assume that the loss was for chronic conditions that require daily uptake. Under this assumption, we can multiply the decline in aggregate demand from Model 3 by average sales over the two-year sample period prior to the relabeling event. This translates into an estimated decline of 7.97 million standard units or slightly over 265,000 30-day prescriptions. If all of these prescriptions were for chronic conditions then this translates into approximately 11,000 consumers that fall out of the market.[32] Again, this is likely to be a conservative, lower bound estimate because not every prescription is for a chronic condition requiring a daily dose. As the number of prescriptions for acute conditions increase so would the number of consumers that fall out of the market.

5.2 Heterogeneous impacts across relabeling intensity

Relabeling intensity varies across therapeutic markets (see Appendix Table 4). In Tables 4 and 5 we explore how these differential intensities impact aggregate demand. We divide our data into two sub-samples and define 'low-intensity markets' and 'high-intensity markets'.[33] In Table 4, low-intensity markets are defined as those 4-digit ATC markets where there was only one relabeling event over our sample period. In contrast, in Table 5, we define high-intensity markets as those 4-digit ATC markets where more than one relabeling event occurred over the sample period. In Table 4, Model 1 the coefficient on the interaction term (*Relabel * U.S.*) is negative and statistically significant at the one percent level. We find a decline of 10.8 percent in aggregate demand for focal drugs in these low-intensity markets. Interestingly, however, in

[31] While detailing may help explain why physicians may switch consumers we are only able to conjecture why consumers choose to prematurely leave the market. The study of this consumer behavior is critical but beyond the scope of this paper.

[32] Average quarterly sales (21.2 million) x 4.7% = 0.99 million standard units x 8 quarters = 7.97 million standard units. Next, 7.97 million divided by 30 = 265,707 30-day prescriptions. Finally, 265,707 divided by 24 months = 11,071 chronic patients.

[33] At the 4-digit ATC market-level there are 61 markets categorized as low-intensity and 76 as high-intensity.

Model 2 and Model 3 the interaction is not statistically significant. This suggests that intra-market substitution absorbed the decline in aggregate focal drug demand. In other words, in these markets physicians were successfully able to switch consumers to another drug *within* that same 4-digit ATC market. To the extent that consumer or physician concerns are warranted due to a relabeling event, this is the expected outcome.

In high-intensity markets, on the other hand, results are more complex. Across all models in Table 5 the interaction term is negative and statistically significant. In Model 1 aggregate focal drug demand declined by 18.9 percent decline while in Model 2 aggregate demand declined by 6.0 percent for drugs within a focal drug's 4-digit ATC market. As before Model 2 represents intra-market substitution or consumers being switched to other drugs *within* the same 4-digit ATC market. Shifting to the 3-digit ATC market that incorporates inter-market substitution patterns, Model 3, aggregate demand declined by 5.0 percent. Critically, this 5.0 percent decline in aggregate demand represents consumers that fall out of the market.

In Appendix Tables 5 and 6 we redefine low-intensity and high-intensity markets as those markets in the bottom and top quartile of relabeling activity.[34] Results remain robust with those reported in Tables 4 and 5. In low-intensity markets, Appendix Table 5, Model 1 aggregate focal demand declined by 10.3 percent. The interaction was not significant in Model 2 or Model 3 again suggesting that intra-market substitution absorbed the entire decline. For the high-intensity markets, Appendix Table 6, Model 1 aggregate focal drug demand declined by 20.1 percent. In Model 2, which incorporates intra-market substitution patterns, aggregate demand declined by 13.0 percent. Finally, in Model 3 that incorporates inter-market substitution, aggregate demand declined by 8.3 percent. Again, this 8.3 percent decline in aggregate demand represents consumers that fall out of the market. These markets with repeated negative shocks appear to reinforce consumers' behavioral responses thereby causing them to leave.

5.3. Heterogeneous impacts across levels of relabeling severity

As discussed in Section 2.0 the severity of drug relabeling spans from precaution (least serious) through box warnings (most serious). Table 6 explores whether the aggregate demand response we document varies across this continuum of severity. We split the data into three sub-samples representing precaution (Model 1), adverse reaction (Model 2) and warning/box warning (Model 3). The categorization continues to be based on the first time a drug is relabeled and

[34] At the 4-digit ATC market-level there are 35 markets in the bottom quartile and 36 markets in the top quartile.

allows us to isolate out the effects of any potential prior relabeling activity. Drugs that have multiple types of relabeling are counted individually in each category.[35] Across all models the interaction remains negative and statistically significant. As expected, we see an increasingly negative aggregate demand response as severity increases; aggregate demand declines by 15.6 percent, 20.3 percent and 36.3 percent in Models 1, 2 and 3, respectively.

The increasing decline in aggregate demand as severity increases should not be surprising; physicians appear to be switching consumers to other drugs as new potential risks reveal themselves. Notwithstanding this general decline the magnitude of results in Model 1 are unexpected. This appears to be a rather strong aggregate demand response given the limited severity of the relabeling event. Unfortunately, we don't know what caused physicians to react in such a significant way. That said, if the response is medically warranted or if physicians believe there may be future problems with a relabeled drug, then we should see intra-market substitution absorb this decline.[36]

We examine this in Table 7 where we split the sample and combine the two least severe relabeling events (*i.e.*, precaution and adverse reaction) together. Again, across the models we find a negative and statistically significant coefficient on our interaction of interest. At the focal drug level, Model 1, aggregate demand declined by 14.7 percent while at the 4-digit ATC market level, which incorporates intra-market substitution, aggregate demand declined by 5.1 percent. At the 3-digit ATC market level, Model 3, which accounts for inter-market substitution aggregate demand declined by 4.0 percent and represents the consumers that leave the market.

In Table 6 Model 3, aggregate demand declined by 36.3 percent for drugs that received either a warning or box warning. This response should not be surprising given the severity of the relabeling event. In Table 8, we combine warnings and box warnings and examine their intra- and inter-market substitution patterns. Across all three models in Table 8 our coefficient on the interaction term is negative and statistically significant. At the 4-digit ATC market level that incorporates intra-market substitution patterns (Model 2), aggregate demand declined by 10.0 percent. At the 3-digit ATC market level that accounts for inter-market substitution patterns (Model 3), aggregate demand declined by 8.3 percent and again represents consumers that leave

[35] For example, if a relabeling event included both a precaution and an adverse reaction it would be included both as a precaution and adverse reaction individually.

[36] The average probability that a drug that has received a precaution receives another relabel is 72.2%. As such, physicians may be pre-emptively switching patients to another drug. However, in this case we should see the entirety of aggregate demand decline of a focal drug absorbed by intra-market substitution.

the market. As the severity of the relabeling event increases (Table 7, Model 3 versus Table 8, Model 3) the percentage of consumers that leave the market increases as well.[37] Importantly, given the substitution patterns captured within Model 3, consistent with prospect theory, consumers appear to be viewing potential substitutes in the same negative manner as the focal drug.

Finally, we combine the intensity levels of relabeling activity from the prior section and examine how it impacts the heterogeneity of relabeling severity that we considered in this section. In Appendix Tables 9 and 10 we replicate Tables 7 and 8 for low-intensity markets. Results are consistent with our prior findings (Table 4 and Appendix Table 5). In Appendix Tables 9 and 10 we see declines in aggregate focal demand (Model 1) of 6.6 and 45.0 percent, respectively. Results in Models 2 and 3 are not statistically significant, suggesting that the entire decline in aggregate focal drug demand was absorbed by intra-market substitution.

In Appendix Tables 11 and 12 we replicate Tables 7 and 8 for high-intensity markets. Again, results are consistent with our prior findings for high-intensity markets (Table 5 and Appendix Table 6). For relabeling events that involved precaution or adverse warnings in high intensity markets, aggregate demand declined by 17.3 percent (Appendix Table 11, Model 1). At the 4-digit ATC market (Model 2) that incorporates intra-market substitution patterns, aggregate demand declined by 5.9 percent. Finally, at the 3-digit ATC market level (Model 3) that incorporates inter-market substitution patterns, aggregate demand declined by 4.8 percent. This again represents consumers that leave the market. The most significant declines are in high-intensity markets with warnings or box warnings (Appendix Table 12). Aggregate demand declined by 34.3 percent at the focal drug level (Model 1), 10.4 percent at the 4-digit ATC market level (Model 2), and 15.8 percent at the 3-digit ATC market level (Model 3). Unlike low-intensity markets where intra-market substitution absorbed the decline in aggregate focal drug demand, in high-intensity markets we see significant movement by consumers out of the market.

6.0 Robustness

6.1. Variation across market concentration and market size

[37] In Appendix Tables 7a, 7b, 8a and 8b we consider alternative time periods. First, in Appendix Tables 7a and 7b we consider three and four years before and after a relabeling event (as opposed to two years in our baseline model). Second, our baseline model excludes the quarter in which a relabeling event occurred. In Appendix Tables 8a and 8b we exclude one and two quarters prior to the relabeling event (along with the quarter of the event). In both tables and across all models our results remain robust to our baseline findings.

It may be possible that variation in market size or the level of competition within markets may differentially influence physician prescribing behavior or consumer behavior. For example, business or general news stories may enhance physician or consumer awareness about a drug. We examine these issues in Appendix Table 13. In Models 1 and 2 we separate markets into the bottom and top quartiles of sales while in Models 3 and 4 we create a HHI index and separate markets into the bottom and top quartiles, respectively. Across all models we find a negative and significant coefficient on our interaction term. Aggregate demand declined by 9.5 percent and 19.8 percent in the bottom and top sales quartiles (Models 1 and 2), respectively. However, when we consider the bottom and top quartiles of HHI, the difference becomes negligible. In Models 3 and 4, aggregate demand declined by 22.8 percent and 21.3 percent, respectively. Thus, we appear to see some variation in response across market sizes but not across levels of competition.

6.2 Heterogeneity across therapeutic markets

A benefit of the breadth of our data is that we capture all therapeutic markets; the impacts we find are average effects across these markets. Lost in our analysis, however, is the potential heterogeneity that may exist between markets. Thus, we examine two therapeutic markets that, according to our discussions with physicians and prior research, exhibit significantly different adherence rates and treatment periods. The first market we consider is ATC N (nervous system), which is comprised of seven 2-digit ATC therapeutic markets: anesthetics (N01), analgesics (N02), antiepileptics (N03), anti-Parkinson (N04), psycholeptics (N05), psychoanaleptics (N06) and other nervous system drugs (N07).

Within these 2-digit ATC markets we have additional 3-digit and 4-digit ATC markets. For example, within N06 resides anti-depressants (N06A) and anti-dementia (N06D) drugs. In general, ATC N exhibits lower levels of non-adherence and longer treatment periods than our second therapeutic market. One study places the non-adherence rates of antiepileptic drugs at 26 percent (Faught *et al*, 2008). In Appendix Table 14 we find a decline in aggregate focal drug demand of 21.4 percent (Model 1), however, the coefficient of interest is not significant in Model 2 or Model 3. These markets experience greater declines in aggregate demand, in percentage terms, than we saw for the overall sample, however, the entire decline is absorbed by intra-market substitution. That is, physicians successfully switch consumers to other drugs within the same 4-digit ATC market.

The second market that we consider is ATC J (anti-infectives), which is comprised of six 2-digit ATC markets: anti-bacterials (J01), anti-mycotics (J02), anti-mycobaterials (J04), anti-virals (J05), immune sera and immunoglobulins (J06), and vaccines (J07). The 2-digit ATC market J01 includes 10 different 3-digit ATC markets comprising various classes of anti-bacterials; for example, tetracyclines (J01A) and beta-lactam anti-bacterials/penicillins (J01C). In general, these ATC markets exhibit greater rates of non-adherence and shorter treatment periods than ATC N. Two recent studies (Fernandes *et al.*, 2014 and Tong *et al.*, 2018) place the non-adherence rates for antimicrobial therapies at greater than 57 percent. In Appendix Table 15 we find a focal decline in aggregate demand of 24.2 percent (Model 1). In these markets, however, we also see declines of 13.8 percent and 13.5 percent in the 4-digit (Model 2) and 3-digit (Model 3) ATC markets, respectively.

While we only explore two markets we see rather significant heterogeneity in physician substitution patterns and consumer response. These two markets were intentionally chosen because they differed in non-adherence rates and average treatment lengths. Unfortunately, we lack the data to say for certainty what specific attribute of these markets caused the physician and consumer responses that we observed. What we can say, however, is that there appears to be significant heterogeneity across markets and this has implications for focal firm and competitor responses as well as for regulators. Further work exploring the *why* behind these movements is clearly warranted.

7.0. Discussion and conclusions

Regulatory interventions rarely occur without consequences. While we are not the first to analyze the impacts of drug relabeling in the U.S, we are the first to do so in such a comprehensive and causal manner. Given the breadth of our data we are able to incorporate all plausible intra- and inter-market substitution patterns along with focal firm and competitor actions. This allows us to estimate not only the causal impact of a relabeling event on a focal drug but also quantify consumers that ultimately fall out of the market. In our baseline regressions (Table 3) we find a decline in aggregate focal drug demand of 16.9 percent at the first instance of relabeling. Critically, after accounting for intra- and inter-market substitution patterns and potential competitor reactions we find a decline in aggregate demand of 4.7 percent that represents consumers that leave the market (Table 3, Model 3).

This translates into approximately 11,000 consumers with chronic conditions. We believe this is the first evidence to suggest that drug relabeling is causing consumers to leave the market. Importantly, this estimate should be viewed as a lower bound for two reasons. First, our estimate is calculated based on chronic uptake (see Footnote 24) as opposed to a mix of chronic and acute prescriptions, which would be more realistic. Unfortunately, because we don't know the exact kind of consumer that leaves the market (or their motivation for doing so) we are left with our chronic patient calculation. Second, we know that information asymmetries exist with respect to the uptake of disseminated relabeling information. Simply, it is probable that we are not capturing *all* physician responses to relabeling events. It could be the case that some physicians never respond because they remain uninformed or they respond outside of our two-year window.[38]

Luckily, we do not need to conjecture whether the aggregate demand shifts we find are significant. In a recent paper that highly complements our work, Krieger *et al.* (2018) causally demonstrates the upstream impacts from FDA Public Health Advisories, which includes drug-relabeling events. They find that these negative shocks lead focal firms to increase external R&D activities in the impacted therapeutic categories while competitor firms pivoted away from the same market. If the aggregate demand shocks that we observe were not significant events then we shouldn't expect to see the upstream responses documented by Kreiger *et al.*

Ultimately, the downstream aggregate demand shifts that we document and the upstream innovation responses documented by Krieger *et al*. have implications for policy. Our concern here is the consumer shift from the treated to untreated population. On the one hand, if these are consumers that should be treated but are no longer treated this will be a detriment to welfare. Adding to this negative effect will be those consumers that may have been switched to another drug that turns out to be less effective. On the other hand, there is ample evidence to suggest that some types of drugs are overprescribed so if the consumers that leave the market should never have been treated in the first place then these negative welfare impacts will be dampened. These declines, however, would have to be matched against potential gains from consumers that didn't

[38] If one of the main ways that physicians learn about relabeling activity is from detailing then this scenario is plausible. In preparation for a follow-up, on-going survey preliminary interviews with primary care physicians stated that they "…had little time in their schedules to dedicate to detailing, often trying to listen to sales reps while dictating charts and orders to nurses between scheduled patient visits…" and "…we have several physicians in our practice and will often rotate who has to spend a few moments with sales reps."

suffer a negative medical consequence because they were switched as a result of a relabeling event.

In addition to our baseline results, we find increasing impacts across all levels of relabeling severity (Table 6). Consistent with prior literature (*e.g.*, Dorsey *et al*., 2010) we find the greatest impact for the most severe type of relabel. Less intuitive, however, is why we see such a significant demand response for the least severe relabel (*i.e.*, precaution). Conditional on receiving a precaution, there is a significant probability that a drug will be relabeled again in the future. So it is plausible that physicians are preemptively switching consumers to other drugs. After accounting for intra- and inter-market substitution (Table 7) we find a 4.0 percent decline in aggregate demand, which represents consumers falling out of the market. While we conjectured in the paper as to physician and consumers motivations, understanding their respective *why* is left for future work.

We exploit other variation in our data. For example, we break markets into "low-intensity" and "high-intensity" markets based on the level of relabeling activity within a particular 4-digit ATC market. In the case of low-intensity markets (Table 4 and Appendix Table 5) and low-intensity markets across types of relabeling (Appendix Tables 9 and 10), we find that the entire decline in aggregate focal demand was absorbed by intra-market substitution. That is patients were all successfully switched to other drugs within the same 4-digit ATC market. In contrast, in the case of high-intensity markets (Table 5 and Appendix Table 6) and high-intensity markets across types of relabeling (Appendix Tables 10 and 11) we find not only declines in aggregate focal drug demand but also find that consumers leave the market. This split is an important caveat to prior work, especially the work focused on box warnings (*e.g.*, Dorsey *et al*., 2010; Olfson *et al*., 2008; Jacoby *et al*, 2005) because it suggests the impacts are more nuanced.

A significant body of work has focused on elasticity and brand loyalty within the pharmaceutical industry. These issues are critical, for example, for pricing strategies and how firms respond to competitors and structure end of life strategies of branded products. Our findings suggest that firms should also be concerned with the magnitude of consumer (and physicians) response to adverse news from relabeling events. While some of these shifts may be medically warranted, others may be due to competitor behavior (Macher and Wade, 2016), physicians responding defensively, consumers acting irrationally or some combination of these. Given that we control for focal firm detailing activity our findings suggest that this may not be

enough to stem the decline in aggregate demand. All of this suggests that *how* physicians (and consumers) receive information may have important implications.

Finally, there are a number of FDA programs that offer expedited development and review for new drugs. These programs all attempt to bring new, novel drugs to market more quickly. Evidence exists that these programs have been successful (Chambers *et al.*, 2017). However, drugs approved through these expedited pathways are also more likely to suffer from serious safety label changes (Mostaghim *et al.*, 2017; Moore and Furberg, 2014; Carpenter *et al.*, 2008). As we have documented throughout this analysis, those changes have impacts on downstream aggregate demand as well as upstream innovation (Krieger *et al.*, 2018). These impacts add another layer of complication for regulators to consider in balancing safety with speed.

Clearly, post-marketing safety marketing is important and these expedited approval processes have been successful at bringing novel drugs to market more quickly (Chambers *et al.*, 2017). There are serious issues, however, that need to be resolved by regulators. First, while some aggregate demand shifts should be expected from negative product shocks, drugs are not like normal products - consumers cannot switch products themselves. Understanding *why* physicians switch consumers and *why* some consumers prematurely leave the market is paramount. Optimally, the only shift we should see is that which is medically warranted. Is this a problem with how medical information is being conveyed to physicians? Would systems, such as those studied by Arrow *et al.* (2017), allow physicians to make more informed decisions? Resolving these issues may ultimately diminish the aggregate demand shifts we observe thereby dampening the upstream innovation impacts documented by Kreiger *et al.* (2018). Otherwise, the benefits from expedited programs may not outweigh their longer-term costs; we may be trading quicker access to new, novel drugs today for less innovation tomorrow.

References

Arrow, K., Bilir, K., Sorenson, A., 2017. The impact of information technology on the diffusion of new pharmaceuticals. NBER working paper 23257.

Ball, G., Macher, J., Stern, A., 2018. Negative shocks and innovation: Evidence from medical device recalls. Harvard Business School working paper 19-028. https://hbswk.hbs.edu/item/negative-shocks-and-innovation-evidence-from-medical-device-recalls

Bartel, A. P., Thomas, L. G., 1985. Direct and indirect effects of regulation: A new look at OSHA's impact. Journal of Law and Economics, 28(1): 1-26.

Bartel, A. P., Thomas, L. G., 1987. Predation through regulation: The wage and profit effects of the occupational safety and health administration and the environmental protection agency. Journal of Law and Economics, 30(2): 239-264.

Branstetter, L., Chatterjee, C., Higgins, M.J., 2014. Starving (or fattening) the golden goose: Generic entry and incentives for early-stage pharmaceutical innovation. NBER working paper 20532.

Branstetter, L., Chatterjee, C., Higgins, M.J., 2016. Regulation and welfare: Evidence from Paragraph-IV generic entry in the pharmaceutical industry. NBER working paper 17188. RAND Journal of Economics 47(4), 857-890.

Briesacher, B. A., Soumerai, S. B., Zhang, F., Toh, S., Andrade, S. E., Wagner, J. L., Shoaibi, A., Gurwitz, J. H., 2013. A critical review of methods to evaluate the impact of FDA regulatory actions. Pharmacoepidemiology and Drug Safety, 22(9): 986-994.

Brown, G. W., Lu, J. Y., Wolfson, R. J., 1964. Dynamic modeling of inventories subject to obsolescence. Management Science, 11(1): 51-63.

Bunniran, S., McCaffrey, D. J., Bentley, J. P., Bouldin, A. S., 2009. Pharmaceutical product withdrawal: Attributions of blame and its impact on trust. Research in Social and Administrative Pharmacy, 5(3): 262-273.

Carpenter, D., Zucker, E., Avorn, J., 2008. Drug-review deadlines and safety problems. The New England Journal of Medicine, 358: 1354-1361.

Chambers, J., Thorat, T., Wilkinson, C., Neumann, P., 2017. Drugs cleared through the FDA's expedited review offer greater gains than drugs approved by conventional process. Health Affairs, 8: 1408-1415.

Cheah, E. T., Chan, W. L., Chieng, C. L. L., 2007. The corporate social responsibility of pharmaceutical product recalls: An empirical examination of US and UK markets. Journal of Business Ethics, 76(4): 427-449.

Datta, A., Dave, D., 2017. Effects of physician□directed pharmaceutical promotion on prescription behaviors: Longitudinal evidence. Health Economics, 26(4): 450-468.

Dorsey, E. R., Rabbani, A., Gallagher, S. A., Conti, R. M., Alexander, G. C., 2010. Impact of FDA black box advisory on antipsychotic medication use. Archives of Internal Medicine, 170(1): 96-103.

Dranove, D., 2011. Health care markets, regulators, and certifiers. *Handbook of Health*

Economics, 2: 639-690. Elsevier.

Dranove, D., Olsen, C., 1994. The economic side effects of dangerous drug announcements. The Journal of Law and Economics, 37(2): 323-348.

Dusetzina, S. B., Higashi, A. S., Dorsey, E. R., Conti, R., Huskamp, H. A., Zhu, S., Garfield, C. F., Alexander, G. C., 2012. Impact of FDA drug risk communications on health care utilization and health behaviors: A systematic review. Medical Care, 50(6): 466-478.

Faught, E., Duh, M., Weiner, J., Guerin, A., Cunnington, C., 2008. Nonadherence to antiepileptic drugs and increased mortality: Findings from the RANSOM study. Neurology, 71(20): 1572-1578.

Fernandes, M., Leite, A., Basto, N., Nobre, M., Vieira, N., Fernandes, R., Nogueira, P., Nicola, P., 2014. Non-adherence to antibiotic therapy in patients visiting community pharmacies. International Journal of Clinical Pharmacology, 36(1): 86-91.

Forgacs, I., Loganayagam, A., 2008. Overprescribing proton pump inhibitors. British Medical Journal 336: 2.

Freedman, S., Kearney, M., Lederman, M., 2012. Product recalls, imperfect information, and spillover effects: Lessons from the consumer response to the 2007 toy recalls. Review of Economics and Statistics, 94(2): 499-516.

Gruenspecht, H. K., Lave, L. B., 2006. The economics of health, safety, and environmental regulation. *Handbook of Industrial Organization*, 2(7): 1507-1550.

Hartley, R. F. 1994. *Management Mistakes & Successes*: John Wiley and Sons.

Jacoby, J. L., Fulton, J., Cesta, M., Heller, M., 2005. After the black box warning: Dramatic changes in ED use of droperidol. The American Journal of Emergency Medicine, 23(2): 196.

Hoy, M., Levenshus, A., 2018. A mixed methods approach to assessing actual risk readership on branded drug websites. Journal of Risk Research, 21(5): 521-528.

Kahneman, D., Tverksy, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica, 47: 263-291.

Kales, H. C., Zivin, K., Kim, H. M., Valenstein, M., Chiang, C., Ignacio, R. V., Ganoczy, D., Cunningham, F., Schneider, L. S., Blow, F. C., 2011. Trends in antipsychotic use in dementia 1999-2007. Archives of General Psychiatry, 68(2): 190-197.

Kessel, R. A., 1967. Economic effects of rederal Regulation of milk markets. Journal of Law and Economics, 10: 51-78.

Lembke, A., Papac, J., Humphreys, K., 2018. Our other prescription drug problem. New England

Journal of Medicine, 378(8): 693-695.

Leone, R. P., 1995. Generalizing what is known about temporal aggregation and advertising carryover. Marketing Science, 14(3): G141-G150.

Lu, C. Y., Zhang, F., Lakoma, M. D., Madden, J. M., Rusinak, D., Penfold, R. B., Simon, G., Ahmedani, B. K., Clarke, G., Hunkeler, E. M., Waitzfelder, B., 2014. Changes in antidepressant use by young people and suicidal behavior after FDA warnings and media coverage: Quasi-experimental study. British Medical Journal, 348: 3596.

Macher, J., Wade, J., 2016. The 'black box' of strategy: Competitive responses to and performance responses to adverse regulatory events. https://ashecon.confex.com/ashecon/2016/webprogram/Paper4635.html

Manchanda, P., Honka, E., 2005. The effects and role of direct-to-physician marketing in the pharmaceutical industry: An integrative review. Yale Journal of Health Policy, Law and Ethics, 5(2): 785-822.

Mostaghim, S., Gagne, J., Kesselheim, A., 2017. Safety related changes for new drugs after approval in the US through expedited regulatory pathways: Retrospective cohort study. British Medical Journal 358: 3837.

Mazor, K., Andrade, S., Auger, J., Fish, L., Gurwitz, J., 2005. Communicating safety information to physicians: An examination of dear doctor letters. Pharmacoepidemiology & Drug Safety 14(12): 869-875.

Migue, J. L., 1977. Controls versus subsidies in the economic theory of regulation. Journal of Law and Economics, 20(1): 213-221.

Moore, T. J., Cohen, M. R., Furberg, C. D., 2007. Serious adverse drug events reported to the Food and Drug Administration, 1998-2005. Archives of Internal Medicine, 167(16): 1752-1759.

Moore, T.J., Furberg, C.D., 2014. Development time, clinical testing, postmarket follow-up, and safety risks for the new drugs approved by the US food and drug administration: The class of 2008. JAMA Internal Medicine, 174(1): 90-95.

Nelson, P., 1970. Information and consumer behavior. Journal of Political Economy, 78(2): 311-329.

Nevo, A., 2001. Measuring market power in the ready-to-eat cereal industry. Econometrica, 69(2): 307-342.

Oberholzer-Gee, F., Mitsunari, M., 2006. Information regulation: Do the victims of externalities pay attention? Journal of Regulatory Economics, 30(2): 141-158.

Olfson, M., Marcus, S. C., 2008. A case-control study of antidepressants and attempted suicide

during early phase treatment of major depressive episodes. Journal of Clinical Psychiatry, 69(3): 425-432.

Olson, M. K., 2003. Pharmaceutical policy change and the safety of new drugs. Journal of Law and Economics, 45(S2): 615-642.

Peltzman, S., 1976. Toward a more general theory of economic regulation. Journal of Law and Economics, 19(2): 211-240.

Peltzman, S., 1987. The health effects of mandatory prescriptions. Journal of Law and Economics, 30(2): 207-238.

Price, D., Cooke, J., Singleton, S., Feely, M., 1986. Doctors' unawareness of the drugs their patients are taking: A major cause of overprescribing? British Medical Journal, 292(6513): 99-100.

Qureshi, Z., Seoane-Vazqquez, E., Rodriguez-Monguio, R., Stevenson, K., Szeinbach, S., 2011. Market withdrawl of new molecular entities approved in the United States from 1980 to 2009. Pharmacoepidemiology & Drug Safety, 20(7): 772-777.

Rizzo, J., 1999. Advertising and competition in the ethical pharmaceutical industry: The case of antihypertensive drugs. The Journal of Law and Economics, 42(1): 89-116.

Sacarny, A., Yokum, D., Finkelstein, A., Agrawal, S., 2016. Medicare letters to curb overprescribing of controlled substances had no detectable effect on providers. Health Affairs, 35(3): 471-479.

Shah, N. D., Montori, V. M., Krumholz, H. M., Tu, K., Alexander, G. C., Jackevicius, C. A., 2010. Responding to an FDA warning—geographic variation in the use of rosiglitazone. New England Journal of Medicine, 363(22): 2081-2084.

Sloan, F. A., Steinwald, B., 1980. Effects of regulation on hospital costs and input use. Journal of Law and Economics, 23(1): 81-109.

Smalley, W., Shatin, D., Wysowski, D. K., Gurwitz, J., Andrade, S. E., Goodman, M., Chan, K. A., Platt, R., Schech, S. D., Ray, W. A., 2000. Contraindicated use of cisapride: Impact of food and drug administration regulatory action. The Journal of the American Medical Association, 284(23): 3036-3039.

Stigler, G. J., 1971. Theory of economic regulation. Bell Journal of Economics, 2(1): 3-21.

Tekin, E., Markowitz, S., 2008. The relationship between suicidal behavior and productive activities of young adults. Southern Economic Journal: 300-331.

Ter-Martirosyan, A., Kwoka, J., 2010. Incentive regulation, service quality, and standards in U.S. electricity distribution. Journal of Regulatory Economics, 38(3): 258-273.

Tong, S., Pan, J., Lu, S., Tang, J., 2018. Patient compliance with antimicrobial drugs: A Chinese study. American Journal of Infection Control 46, 25-29.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty, 5(4): 297-323.

Wilkinson, J. J., Force, R. W., Cady, P. S., 2004. Impact of safety warnings on drug utilization: Marketplace life span of cisapride and troglitazone. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy, 24(8): 978-986.

Windmeijer, F., de Laat, E., Douven, R., Mot, E., 2005. Pharmaceutical promotion and GP prescription behavior. Health Economics, 15(1): 5-18.

Wong, C., Siah, K., Lo, A., 2018. Estimation of clinical trial success rates and related parameters. *Forthcoming* Biostatistics: https://doi.org/10.1093/biostatistics/kxx069

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

Zeithaml, V. A., 1988. Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. Journal of Marketing, 52(3): 2-22.

Zhao, X., Li, Y., Flynn, B. B., 2013. The financial impact of product recall announcements in China. International Journal of Production Economics, 142(1): 115-123.

Figure 1. Focal drug demand in U.S. and U.K. surrounding relabel events. The figure shows the sales quantity of focal drugs in the U.S. (treated) and U.K. (control) before and after relabeling. The relabeling event is set at *t*=0 where time horizon is in quarters and labeled on the x-axis. Sales are shown over eight quarters before and after the quarter of relabeling. Drug sales (thousands) are in standardized units determined by IMS Health and natural logarithms are taken (y-axis).

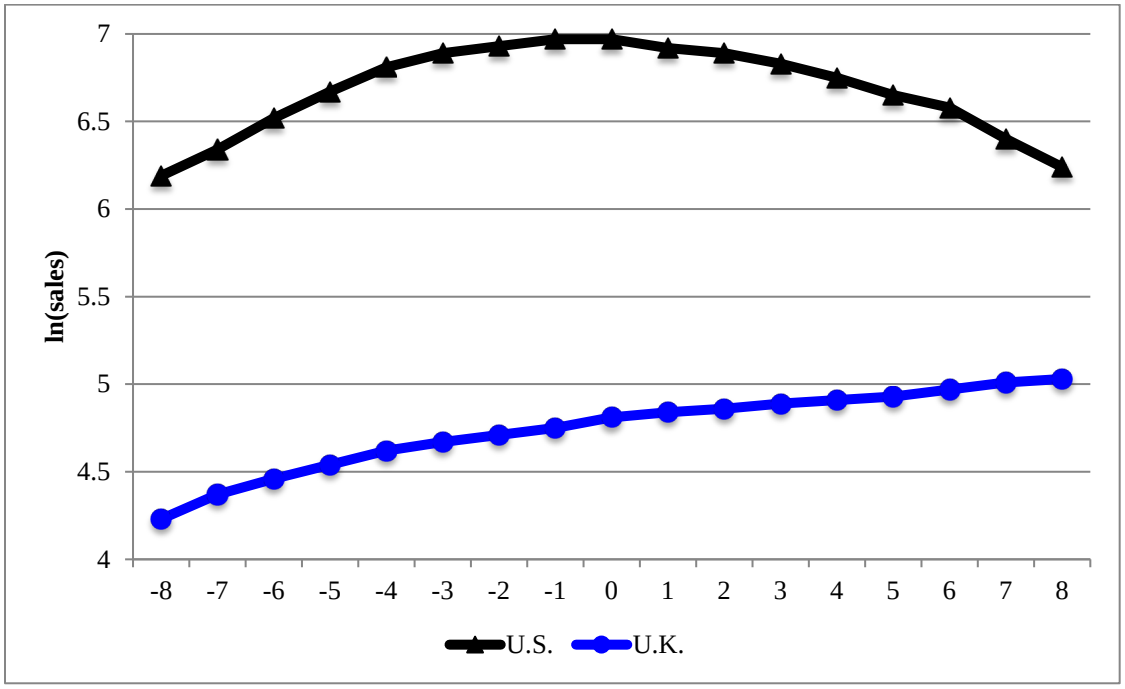

Table 1. Distribution of relabel activity between the U.S. and U.K. Our sample consists of drugs sold both in the U.S. and U.K. In order to create a clean control window we excluded drugs if they were relabeled in the U.K. within eight quarters of a U.S. relabel. This table shows the variation in relabeling types across the U.S. and U.K. for our sample. Within the imposed restrictions the average elapsed time between relabeling in the U.S. and U.K. is 12.95 quarters.

Relabeling Type	U.S.	U.K.	Average time (Quarters)
Precaution	226	166	13.48
Adverse Reaction	176	134	11.83
Warning	161	115	12.06
Box Warning	53	35	9.40
Label Changes	251	180	12.95

Table 2. Descriptive statistics. Sales (quantity) are measured in millions of standardized units. IMS Health has converted financial variables for U.K. drugs to U.S. dollars. All financial variables have been converted to real 2009 U.S. dollars using a GDP deflator.

Variable	N	Mean	Median	Std. Dev.	Min	Max
U.S.	6,519	0.54	1.00	0.50	0.00	1.00
Sales (standard units)	6,519	13.87	0.88	45.81	0.00	577.85
Promotion	6,519	1.73	0.02	5.40	0.00	63.18
Lagged promotion stock	6,519	6.84	0.68	15.46	0.00	135.17
Price	6,519	91.78	2.36	357.43	0.01	5352.50
Relabel	6,519	0.27	0.00	0.45	0.00	1.00
Precaution	6,519	0.22	0.00	0.42	0.00	1.00
Adverse reaction	6,519	0.16	0.00	0.37	0.00	1.00
Warning	6,519	0.11	0.00	0.31	0.00	1.00
Box warning	6,519	0.03	0.00	0.17	0.00	1.00
Vintage	6,519	23.53	24.00	11.53	1.00	56.00
Number of brands	6,519	7.98	6.00	6.31	0.00	32.00
Number of generics	6,519	13.70	5.00	23.70	0.00	149.00

Table 3. Effects of relabeling on demand. Dependent variable is the natural logarithm of sales, ln(*Sales*). The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln(Sales)	Model 1 Focal drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.108***	0.058**	0.023
	(0.032)	(0.024)	(0.020)
U.S.	0.712***	1.796***	1.375***
	(0.034)	(0.040)	(0.036)
Relabel * U.S.	-0.185***	-0.052*	-0.048**
	(0.025)	(0.028)	(0.023)
ln(Price)	-0.610***	-1.158***	-0.544***
	(0.053)	(0.049)	(0.049)
ln(Lagged promotion stock)	0.742***	0.186***	0.151***
	(0.015)	(0.009)	(0.007)
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	6,519	5,946	4,946
Adjusted R^2	0.531	0.765	0.820
First stage F-statistic	37.12	64.79	26.56
Hansen J-statistic	2.12	0.15	2.621
Hansen J p-value	0.145	0.698	0.105
Marginal effects:			
Relabel * U.S.	**-0.169**	**-0.051**	**-0.047**

Table 4. Effects of relabeling on demand: Low-intensity markets. Dependent variable is the natural logarithm of sales, ln(*Sales*). Low-intensity markets are defined as those 4-digit ATC markets where there was only one relabeling event over our sample period. The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln(Sales)	Model 1 Focal drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.096*	-0.008	-0.071
	-0.052	-0.094	-0.089
U.S.	0.350***	1.640***	0.842***
	-0.056	-0.094	-0.085
Relabel * U.S.	-0.114***	-0.012	-0.015
	-0.04	-0.07	-0.064
ln(Price)	-0.522***	-1.073***	-0.445***
	-0.055	-0.046	-0.072
ln(Lagged promotion stock)	0.691***	0.153***	0.166***
	-0.034	-0.021	-0.019
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	1,576	1,576	749
Adjusted R^2	0.655	0.561	0.638
First stage F-statistic	79.67	150.21	16.34
Hansen J-statistic	1.235	0.477	0.092
Hansen J p-value	0.267	0.490	0.761
Marginal effects:			
Relabel * U.S.	**-0.108**		

Table 5. Effects of relabeling on demand: High-intensity markets. Dependent variable is the natural logarithm of sales, ln(*Sales*). High-intensity markets are defined as those 4-digit ATC markets where there was more than one relabeling event over our sample period. The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln(Sales)	Model 1 Focal Drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.118***	0.072***	0.026
	-0.037	-0.022	-0.02
U.S.	0.904***	1.874***	1.539***
	-0.038	-0.034	-0.043
Relabel * U.S.	-0.210***	-0.062**	-0.051**
	-0.03	-0.028	-0.023
ln(Price)	-0.846***	-1.137***	-0.539***
	-0.099	-0.081	-0.076
ln(Lagged promotion stock)	0.706***	0.180***	0.129***
	-0.019	-0.008	-0.006
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	4,943	4,370	4,197
Adjusted R^2	0.526	0.833	0.852
First stage F-statistic	17.61	48.22	33.22
Hansen J-statistic	1.501	0.218	0.778
Hansen J p-value	0.221	0.640	0.378
Marginal effects:			
Relabel * U.S.	**-0.189**	**-0.060**	**-0.050**

Table 6. **Heterogeneous impacts across levels of relabeling severity.** Dependent variable is the natural logarithm of sales, ln(*Sales*). Data is split into three sub-samples representing precaution (Model 1), adverse reaction (Model 2) and warning/box warning (Model 3). The categorization is based on the first time a drug is relabeled and allows us to isolate out the effects of any potential prior relabeling activity. The unit of analysis across all models is the focal drug level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

DV = ln(Sales)	Model 1 Precaution	Model 2 Adverse reaction	Model 3 Warning/Box
Relabel	0.111***	0.143***	0.176***
	(0.030)	(0.040)	(0.050)
U.S.	0.838***	0.749***	0.607***
	(0.039)	(0.041)	(0.051)
Relabel * U.S.	-0.170***	-0.227***	-0.451***
	(0.027)	(0.036)	(0.059)
ln(Price)	-0.759***	-0.528***	-0.709***
	(0.069)	(0.039)	(0.070)
ln(Lagged promotion stock)	0.725***	0.790***	0.756***
	(0.017)	(0.022)	(0.026)
Controls	Y	Y	Y
Drug fixed effect	Y	Y	Y
Market fixed effect	N	N	N
Time fixed effect	Y	Y	Y
N	5,183	3,166	2,236
Adjusted R^2	0.517	0.579	0.430
First stage F-statistic	29.39	65.76	37.76
Hansen J-statistic	0.81	5.821	1.451
Hansen J p-value	0.368	0.055	0.228
Marginal effects: **Relabel * U.S.**	**-0.156**	**-0.203**	**-0.363**

Table 7. Effects of precaution/adverse reaction relabeling on demand. Dependent variable is the natural logarithm of sales, ln(*Sales*). Sample includes the combination of precaution and adverse reaction. The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln(Sales)	Model 1 Focal Drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.068**	0.067***	0.021
	(0.032)	(0.024)	(0.021)
U.S.	0.659***	1.779***	1.440***
	(0.039)	(0.042)	(0.033)
Relabel * U.S.	-0.159***	-0.052*	-0.041*
	(0.035)	(0.028)	(0.023)
ln(Price)	-0.569***	-1.175***	-0.700***
	(0.053)	(0.045)	(0.027)
ln(Lagged promotion stock)	0.808***	0.186***	0.151***
	(0.018)	(0.009)	(0.007)
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	6,310	5,722	4,955
Adjusted R^2	0.407	0.768	0.812
First stage F-statistic	49.62	70.28	42.90
Hansen J-statistic	0.065	0.090	2.254
Hansen J p-value	0.799	0.765	0.133
Marginal effects:			
Relabel * U.S.	**-0.147**	**-0.051**	**-0.040**

Table 8. Effects of warning/box warning relabeling on demand. Dependent variable is the natural logarithm of sales, ln(*Sales*). Sample includes the combination of warning and box warning. The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

DV = ln(Sales)	Model 1 Focal Drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.176***	0.020	0.005
	(0.050)	(0.034)	(0.029)
U.S.	0.607***	1.821***	1.545***
	(0.051)	(0.048)	(0.042)
Relabel * U.S.	-0.451***	-0.105**	-0.087**
	(0.059)	(0.044)	(0.036)
ln(Price)	-0.709***	-0.997***	-0.533***
	(0.070)	(0.059)	(0.028)
ln(Lagged promotion stock)	0.756***	0.174***	0.117***
	(0.026)	(0.012)	(0.012)
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	2,236	2,189	1,991
Adjusted R^2	0.430	0.834	0.812
First stage F-statistic	37.76	96.64	753.48
Hansen J-statistic	1.451	0.824	1.653
Hansen J p-value	0.228	0.364	0.199
Marginal effects:			
Relabel * U.S.	**-0.363**	**-0.100**	**-0.083**

Appendix Table 1. Placebo test of parallel trend assumption. This placebo test is on the pre-trend data where it is divided in two with the mid-point being assigned as the arbitrary event date. The dependent variable is the natural logarithm of sales, ln(*Sales*). The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

DV = ln(Sales)	Model 1 Focal drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	-0.026	0.025	0.003
	(0.041)	(0.030)	(0.026)
U.S.	0.600***	1.693***	1.433***
	(0.059)	(0.054)	(0.043)
Relabel * U.S.	-0.069	-0.008	-0.006
	(0.050)	(0.037)	(0.032)
ln(Price)	-0.483***	-0.966***	-0.672***
	(0.066)	(0.059)	(0.048)
ln(Lagged promotion stock)	0.878***	0.202***	0.153***
	(0.024)	(0.013)	(0.009)
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	3,141	2,865	2,383
Adjusted R^2	0.481	0.790	0.823
First stage F-statistic	25.71	49.62	36.46
Hansen J-statistic	2.22	0.15	2.728
Hansen J p-value	0.157	0.658	0.135

Appendix Table 2. Baseline results across alternative time periods. Dependent variable is the natural logarithm of sales, ln(*Sales*). Model 1 replicates Model 1, Table 3. The time period of analysis is extended to three years (12 quarters) in Model 2 and four years (16 quarters) in Model 3. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln(Sales)	Model 1 2 years	Model 2 3 years	Model 3 4 years
Relabel	0.108***	0.113***	0.132***
	(0.032)	(0.027)	(0.025)
U.S.	0.712***	0.696***	0.711***
	(0.034)	(0.027)	(0.026)
Relabel * U.S.	-0.185***	-0.235***	-0.279***
	(0.025)	(0.022)	(0.021)
ln(Price)	-0.610***	-0.499***	-0.506***
	(0.053)	(0.033)	(0.036)
ln(Lagged promotion stock)	0.742***	0.760***	0.772***
	(0.015)	(0.013)	(0.012)
Controls	Y	Y	Y
Drug fixed effect	Y	Y	Y
Market fixed effect	N	N	N
Time fixed effect	Y	Y	Y
N	6,519	9,229	11,842
Adjusted R^2	0.497	0.498	0.484
First stage F-statistic	37.12	36.23	39.12
Hansen J-statistic	2.120	4.310	3.982
Hansen J p-value	0.145	0.635	0.679
Marginal effects:			
Relabel * U.S.	**-0.168**	**-0.209**	**-0.243**

Appendix Table 3. Baseline results across alternative treatment periods. Dependent variable is the natural logarithm of sales, ln(*Sales*). In our baseline model specification the quarter of relabel is excluded from analysis. Model 1 drops the quarter of relabel (t = 0) and the quarter prior (t = -1) Model 2 drops the quarter of relabel (t = 0) and the two quarters prior (t = -1, -2). Dropping prior quarters controls for any possible leakage of information. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln(Sales)	Model 1 T = 0, -1	Model 2 T = 0, -1, -2
Relabel	0.118***	0.124***
	(0.036)	(0.041)
U.S.	0.671***	0.673***
	(0.032)	(0.034)
Relabel * U.S.	-0.203***	-0.210***
	(0.027)	(0.028)
ln(Price)	-0.472***	-0.470***
	(0.040)	(0.042)
ln(Lagged promotion stock)	0.739***	0.745***
	(0.016)	(0.016)
Controls	Y	Y
Drug fixed effect	Y	Y
Market fixed effect	N	N
Time fixed effect	Y	Y
N	6,074	5,661
Adjusted R^2	0.501	0.498
First stage F-statistic	48.86	44.81
Hansen J statistic	3.791	3.457
Hansen J p-value	0.150	0.178
Marginal effects:		
Relabel * U.S.	**-0.183**	**-0.189**

Appendix Table 4. Variation in relabeling activity. Count of first instance relabeling activity across type and over time. If a drug has multiple relabel types they are each counted below.

ATC	Relabel Type	2003	2004	2005	2006	2007	2008	2009
A	Precaution	6	13	17	19	24	26	26
	Adverse Reaction	4	7	12	17	21	26	29
	Warning	1	3	4	6	7	12	13
	Boxed Warning	0	0	0	0	3	5	5
	Total Relabel	6	14	19	22	28	31	33
B	Precaution	4	5	7	7	7	8	8
	Adverse Reaction	4	7	9	9	9	9	9
	Warning	2	3	6	6	7	7	7
	Boxed Warning	0	0	0	0	1	1	3
	Total Relabel	4	7	9	9	9	10	11
C	Precaution	10	15	20	26	30	36	39
	Adverse Reaction	11	15	19	22	22	24	26
	Warning	9	11	15	17	20	23	24
	Boxed Warning	0	0	1	2	3	4	4
	Total Relabel	16	23	27	31	34	40	42
D	Precaution	1	5	7	7	7	7	7
	Adverse Reaction	0	3	4	4	5	5	5
	Warning	0	0	1	3	3	3	3
	Boxed Warning	0	0	0	2	2	3	3
	Total Relabel	1	5	7	8	8	8	8
G	Precaution	5	7	14	15	16	17	17
	Adverse Reaction	2	4	10	11	11	12	13
	Warning	0	0	1	1	5	9	11
	Boxed Warning	0	0	0	0	0	0	0
	Total Relabel	5	7	14	15	17	18	18
H	Precaution	1	5	5	6	6	7	7
	Adverse Reaction	1	4	5	6	6	7	7
	Warning	0	2	2	3	3	4	4
	Boxed Warning	0	0	0	0	0	0	0
	Total Relabel	1	5	5	6	6	7	7
J	Precaution	18	44	50	56	59	64	68
	Adverse Reaction	12	30	40	45	46	53	57
	Warning	6	16	22	30	41	53	56
	Boxed Warning	2	8	9	10	13	15	16
	Total Relabel	22	49	55	59	62	69	72

Appendix Table 4. Variation in relabeling activity (continued).

ATC	Relabel Type	2003	2004	2005	2006	2007	2008	2009
L	Precaution	14	25	31	38	44	46	52
	Adverse Reaction	11	19	30	35	41	46	51
	Warning	7	10	17	29	38	41	47
	Boxed Warning	4	5	6	16	21	21	25
	Total Relabel	17	30	39	43	52	54	57
M	Precaution	8	17	20	23	24	24	24
	Adverse Reaction	5	11	15	15	17	18	19
	Warning	1	6	9	14	16	18	18
	Boxed Warning	0	1	3	10	11	11	12
	Total Relabel	10	19	21	23	24	24	24
N	Precaution	12	24	32	40	45	47	51
	Adverse Reaction	7	14	18	27	35	38	39
	Warning	7	11	22	30	38	40	44
	Boxed Warning	0	0	9	14	16	19	19
	Total Relabel	14	28	36	44	52	55	56
P	Precaution	1	1	2	2	2	2	2
	Adverse Reaction	1	1	1	1	1	1	1
	Warning	0	0	0	0	0	0	0
	Boxed Warning	0	0	0	0	0	0	0
	Total Relabel	1	1	2	2	2	2	2
R	Precaution	3	5	6	6	9	11	11
	Adverse Reaction	3	3	4	4	5	7	7
	Warning	2	3	3	3	4	5	6
	Boxed Warning	1	1	1	1	2	2	2
	Total Relabel	6	8	8	8	10	12	12
S	Precaution	6	8	9	11	11	11	11
	Adverse Reaction	0	1	3	4	4	4	4
	Warning	0	0	0	0	0	0	0
	Boxed Warning	0	0	0	0	0	0	0
	Total Relabel	6	8	10	12	12	12	12
T	Precaution	2	2	3	3	5	5	5
	Adverse Reaction	0	1	1	1	1	1	1
	Warning	0	0	0	1	2	2	3
	Boxed Warning	0	0	0	0	1	1	1
	Total Relabel	2	3	4	4	5	5	6

Appendix Table 4. Variation in relabeling activity (continued).

ATC	Relabel Type	2003	2004	2005	2006	2007	2008	2009
V	Precaution	1	2	2	3	3	4	5
	Adverse Reaction	1	2	2	2	3	3	3
	Warning	1	1	1	1	3	3	3
	Boxed Warning	0	0	0	0	0	0	0
	Total Relabel	1	2	2	3	4	4	5

ATC Therapeutic Code Definition:

A: Alimentary tract and metabolism
B: Blood and blood forming organs
C: Cardiovascular system
D: Dermatological
G: Genitourinary system and sex hormones
H: Systemic hormonal preparations, excluding sex hormones
J: Anti-infectives
L: Anti-neoplastic and immunomodulating agents
M: Musculoskeletal system
N: Nervous system
P: Anti-parasitic products
R: Respiratory system
S: Sensory organs
T: Diagnostic agents
V: Various

Appendix Table 5. Effects of relabeling on demand: Low-intensity markets (bottom quartile). Dependent variable is the natural logarithm of sales, ln(*Sales*). Low-intensity markets are defined as those 4-digit ATC markets in the bottom quartile of relabeling activity over our sample period. The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln(Sales)	Model 1 Focal Drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.077	0.140**	0.007
	(0.054)	(0.059)	(0.051)
U.S.	0.386***	1.013***	1.092***
	(0.051)	(0.117)	(0.080)
Relabel * U.S.	-0.109***	-0.104	-0.018
	(0.041)	(0.067)	(0.054)
ln(Price)	-0.564***	-1.068***	-0.740***
	(0.129)	(0.097)	(0.047)
ln(Lagged promotion stock)	0.971***	0.321***	0.070***
	(0.024)	(0.028)	(0.016)
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	1,264	1,202	1,170
Adjusted R^2	0.725	0.726	0.599
First stage F-statistic	12.81	38.02	81.74
Hansen J-statistic	0.558	0.017	0.590
Hansen J p-value	0.455	0.897	0.442
Marginal effects:			
Relabel * U.S.	**-0.103**		

Appendix Table 6. Effects of relabeling on demand: High-intensity markets (top quartile).
Dependent variable is the natural logarithm of sales, ln(*Sales*). High-intensity markets are defined as those 4-digit ATC markets in the top quartile of relabeling activity over our sample period. The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln(Sales)	Model 1 Focal Drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.157***	0.172***	0.088***
	(0.051)	(0.063)	(0.028)
U.S.	0.812***	1.809***	1.574***
	(0.048)	(0.128)	(0.055)
Relabel * U.S.	-0.224***	-0.139**	-0.087***
	(0.043)	(0.064)	(0.033)
ln(Price)	-0.628***	-1.209***	-0.814***
	(0.051)	(0.056)	(0.028)
ln(Lagged promotion stock)	0.695***	0.259***	0.165***
	(0.036)	(0.026)	(0.012)
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	1,903	1,669	1,533
Adjusted R^2	0.508	0.692	0.882
First stage F-statistic	100.28	147.89	1054.4
Hansen J-statistic	0.147	1.459	0.066
Hansen J p-value	0.701	0.227	0.798
Marginal effects:			
Relabel * U.S.	**-0.201**	**-0.130**	**-0.083**

Appendix Table 7a. Heterogeneous relabeling severity across alternative time periods. Data split across types of relabeling activity and across alternative periods of analysis. Models 1-3 are extended to three years or 12 quarters before and after a relabeling event.

	Model 1	Model 2	Model 3
DV = ln(Sales)	Precaution	3 years Adverse Reaction	Warning/Box
Relabel	0.075***	0.108***	0.036
	-0.026	-0.03	-0.034
U.S.	0.731***	0.770***	0.566***
	-0.035	-0.032	-0.034
Relabel * U.S.	-0.207***	-0.237***	-0.285***
	-0.023	-0.027	-0.041
ln(Price)	-0.700***	-0.473***	-0.372***
	-0.063	-0.039	-0.045
ln(Lagged promotion stock)	0.773***	0.720***	0.698***
	-0.013	-0.014	-0.016
Controls	Y	Y	Y
Drug fixed effect	Y	Y	Y
Market fixed effect	N	N	N
Time fixed effect	Y	Y	Y
N	8,711	6,428	5,943
Adjusted R^2	0.487	0.558	0.433
First stage F-statistic	40.01	42.48	35.42
Hansen J-statistic	0.978	1.781	2.76
Hansen J p-value	0.323	0.619	0.43
Marginal effects:			
Relabel * U.S.	**-0.186**	**-0.211**	**-0.247**

Appendix Table 7b. Heterogeneous relabeling severity across alternative time periods. Data split across types of relabeling activity and across alternative periods of analysis. Models 4-6 are extended to four years or 16 quarters before and after a relabeling event.

	Model 4	Model 5	Model 6
		4 years	
DV = ln(Sales)	Precaution	Adverse Reaction	Warning/Box
Relabel	0.083***	0.119***	0.063**
	-0.024	-0.028	-0.031
U.S.	0.708***	0.740***	0.741***
	-0.027	-0.03	-0.044
Relabel * U.S.	-0.238***	-0.270***	-0.213***
	-0.021	-0.026	-0.039
ln(Price)	-0.611***	-0.470***	-0.698***
	-0.043	-0.039	-0.072
ln(Lagged promotion stock)	0.795***	0.762***	0.703***
	-0.011	-0.013	-0.015
Controls	Y	Y	Y
Drug fixed effect	Y	Y	Y
Market fixed effect	N	N	N
Time fixed effect	Y	Y	Y
N	10,877	8,059	7,587
Adjusted R^2	0.505	0.549	0.418
First stage F-statistic	44.13	39.64	35.92
Hansen J-statistic	4.699	0.843	1.261
Hansen J p-value	0.454	0.974	0.261
Marginal effects:			
Relabel * U.S.	**-0.211**	**-0.236**	**-0.192**

Appendix Table 8a. Heterogeneous relabeling severity across alternative treatment periods. Data split across types of relabeling activity and across alternative exclusion windows. Models 1-3 exclude the quarter of a relabeling event and prior quarter from the analysis (T = 0, -1).

	Model 1	Model 2	Model 3
		T = 0, -1	
DV = ln(Sales)	Precaution	Adverse Reaction	Warning/Box
Relabel	0.124***	0.155***	0.179***
	-0.038	-0.05	-0.054
U.S.	0.847***	0.745***	0.468***
	-0.042	-0.044	-0.047
Relabel * U.S.	-0.191***	-0.255***	-0.456***
	-0.029	-0.039	-0.061
ln(Price)	-0.758***	-0.515***	-0.408***
	-0.073	-0.04	-0.044
ln(Lagged promotion stock)	0.725***	0.804***	0.738***
	-0.018	-0.024	-0.026
Controls	Y	Y	Y
Drug fixed effect	Y	Y	Y
Market fixed effect	N	N	N
Time fixed effect	Y	Y	Y
N	4,552	2,786	1,975
Adjusted R^2	0.491	0.552	0.374
First stage F-statistic	48.17	55.94	36.8
Hansen J-statistic	0.934	2.399	2.316
Hansen J p-value	0.334	0.301	0.314
Marginal effects:			
Relabel * U.S.	**-0.173**	**-0.225**	**-0.366**

Appendix Table 8b. Heterogeneous relabeling severity across alternative treatment periods. Data split across types of relabeling activity and across alternative exclusion windows. Models 4-6 exclude the quarter of a relabeling event and two prior quarters (T = 0, -1, -2).

	Model 4	Model 5 T = 0, -1, -2	Model 6
DV = ln(Sales)	Precaution	Adverse Reaction	Warning/Box
Relabel	0.134***	0.166***	0.184***
	-0.043	-0.056	-0.056
U.S.	0.848***	0.758***	0.471***
	-0.044	-0.047	-0.048
Relabel * U.S.	-0.203***	-0.273***	-0.455***
	-0.031	-0.041	-0.062
ln(Price)	-0.749***	-0.510***	-0.398***
	-0.075	-0.041	-0.046
ln(Lagged promotion stock)	0.734***	0.807***	0.730***
	-0.018	-0.025	-0.027
Controls	Y	Y	Y
Drug fixed effect	Y	Y	Y
Market fixed effect	N	N	N
Time fixed effect	Y	Y	Y
N	4,240	2,599	1,839
Adjusted R^2	0.49	0.548	0.365
First stage F-statistic	21.06	51.52	33.89
Hansen J-statistic	0.57	1.641	3.03
Hansen J p-value	0.45	0.44	0.22
Marginal effects:			
Relabel * U.S.	**-0.165**	**-0.238**	**-0.365**

Appendix Table 9. Effects of precaution/adverse selection relabeling: Low-intensity markets. Dependent variable is the natural logarithm of sales, ln(*Sales*). Sample includes the combination of precaution and adverse reaction along with low-intensity markets. The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln(Sales)	Model 1 Focal Drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.075*	0.024	-0.063
	(0.039)	(0.088)	(0.086)
U.S.	0.333***	1.559***	0.829***
	(0.052)	(0.101)	(0.088)
Relabel * U.S.	-0.068*	-0.020	0.011
	(0.035)	(0.072)	(0.065)
ln(Price)	-0.287***	-1.044***	-0.584***
	(0.053)	(0.045)	(0.036)
ln(Lagged promotion stock)	0.804***	0.165***	0.178***
	(0.034)	(0.021)	(0.020)
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	1,564	1,564	737
Adjusted R^2	0.622	0.558	0.638
First stage F-statistic	76.21	160.56	204.09
Hansen J-statistic	0.248	0.977	2.462
Hansen J p-value	0.618	0.323	0.117
Marginal effects:			
Relabel * U.S.	**-0.066**		

Appendix Table 10. Effects of warning/box warning relabeling: Low-intensity markets. Dependent variable is the natural logarithm of sales, ln(*Sales*). Sample includes the combination of warning and box warning along with low-intensity markets. The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln (Sales)	Model 1 Focal Drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.228*	-0.251**	-0.007
	(0.117)	(0.098)	(0.118)
U.S.	0.174**	1.842***	1.051***
	(0.072)	(0.078)	(0.095)
Relabel * U.S.	-0.597***	0.163	0.021
	(0.148)	(0.113)	(0.120)
ln(Price)	-0.310***	-0.812***	-0.449***
	(0.077)	(0.048)	(0.039)
ln(Lagged promotion stock)	0.596***	0.183***	0.181***
	(0.049)	(0.022)	(0.029)
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	654	762	430
Adjusted R^2	0.746	0.707	0.638
First stage F-statistic	18.28	410.61	580.23
Hansen J-statistic	0.708	1.67	1.276
Hansen J p-value	0.400	0.196	0.259
Marginal effects:			
Relabel * U.S.	**-0.450**		

Appendix Table 11. Effects of precaution/adverse selection relabeling: High-intensity markets. Dependent variable is the natural logarithm of sales, ln(*Sales*). Sample includes the combination of precaution and adverse reaction along with high-intensity markets. The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln(Sales)	Model 1 Focal Drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.035	0.071***	0.019
	(0.041)	(0.023)	(0.021)
U.S.	0.773***	1.889***	1.630***
	(0.047)	(0.034)	(0.032)
Relabel * U.S.	-0.190***	-0.061**	-0.049**
	(0.045)	(0.027)	(0.024)
ln(Price)	-0.658***	-1.176***	-0.781***
	(0.071)	(0.072)	(0.033)
ln(Lagged promotion stock)	0.803***	0.174***	0.133***
	(0.022)	(0.007)	(0.007)
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	4,746	4,158	4,218
Adjusted R^2	0.360	0.844	0.846
First stage F-statistic	31.25	54.03	405.85
Hansen J-statistic	0.011	0.218	1.095
Hansen J p-value	0.915	0.641	0.295
Marginal effects:			
Relabel * U.S.	**-0.173**	**-0.059**	**-0.048**

Appendix Table 12. Effects of warning/box warning relabeling: High-intensity markets. Dependent variable is the natural logarithm of sales, ln(*Sales*). Sample includes the combination of warning and box warning along with high-intensity markets. The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln (Sales)	Model 1 Focal Drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.212**	0.093***	0.041
	(0.087)	(0.034)	(0.031)
U.S.	0.670***	1.778***	1.675***
	(0.070)	(0.052)	(0.041)
Relabel * U.S.	-0.420***	-0.110***	-0.172***
	(0.091)	(0.040)	(0.038)
ln(Price)	-0.891***	-0.581***	-0.594***
	(0.098)	(0.068)	(0.039)
ln(Lagged promotion stock)	0.791***	0.132***	0.114***
	(0.034)	(0.014)	(0.011)
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	1,582	1,409	1,561
Adjusted R^2	0.425	0.902	0.862
First stage F-statistic	41.69	60.91	652.71
Hansen J-statistic	0.674	1.030	0.101
Hansen J p-value	0.412	0.310	0.751
Marginal effects:			
Relabel * U.S.	**-0.343**	**-0.104**	**-0.158**

Appendix Table 13. Effects of relabeling on demand across market size and concentration.
Dependent variable is the natural logarithm of sales, ln(*Sales*). Models 1-2 split the sample across the bottom and top quartile of sales within a 4-digit ATC market. Models 3-4 split the sample across the bottom and top quartile of market concentration or HHI within a 4-digit ATC market. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01

	Model 1	Model 2	Model 3	Model 4
	Bottom Quartile	Top Quartile	Bottom Quartile	Top Quartile
DV = ln(Sales)	Sales		HHI	
Relabel	0.055**	0.058	0.151*	0.175***
	(0.027)	(0.070)	(0.085)	(0.054)
U.S.	0.204***	1.116***	-0.029	0.890***
	(0.024)	(0.061)	(0.059)	(0.049)
Relabel * U.S.	-0.100***	-0.221***	-0.259***	-0.240***
	(0.025)	(0.055)	(0.063)	(0.043)
ln(Price)	-0.221***	-1.280***	0.173	-0.681***
	(0.036)	(0.125)	(0.119)	(0.046)
ln(Lagged promotion stock)	0.459***	0.721***	1.057***	0.685***
	(0.031)	(0.026)	(0.047)	(0.036)
Controls	Y	Y	Y	Y
Drug fixed effect	Y	Y	Y	Y
Market fixed effect	N	N	N	N
Time fixed effect	Y	Y	Y	Y
N	1,369	1,972	1,300	1,930
Adjusted R^2	0.241	0.504	0.332	0.431
First stage F-statistic	21.26	54.41	26.57	113.48
Hansen J-statistic	1.927	4.276	0.054	0.002
Hansen J p-value	0.165	0.233	0.817	0.961
Marginal effects:				
Relabel * U.S.	**-0.095**	**-0.198**	**-0.228**	**-0.213**

Appendix Table 14. Effects of relabeling on market ATC N. Dependent variable is the natural logarithm of sales, ln(*Sales*). The sample includes only those drugs within the 1-digit ATC market N or nervous system. The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln (Sales)	Model 1 Focal Drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.100	0.017	-0.010
	(0.092)	(0.040)	(0.042)
U.S.	0.289***	1.117***	1.026***
	(0.109)	(0.067)	(0.057)
Relabel * U.S.	-0.241***	-0.022	0.039
	(0.074)	(0.053)	(0.054)
ln(Price)	-0.236	1.744***	1.372***
	(0.311)	(0.148)	(0.155)
ln(Lagged promotion stock)	0.961***	0.170***	0.188***
	(0.032)	(0.014)	(0.010)
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	1,148	1,027	949
Adjusted R^2	0.602	0.852	0.859
First stage F-statistic	92.95	243.58	156.74
Hansen J-statistic	0.342	0.286	0.839
Hansen J p-value	0.559	0.593	0.360
Marginal effects:			
Relabel * U.S.	**-0.214**		

Appendix Table 1. Effects of relabeling on market ATC J. Dependent variable is the natural logarithm of sales, ln(*Sales*). The sample includes only those drugs within the 1-digit ATC market J or anti-infectives. The unit of analysis in Model 1 is the focal drug level, Model 2 is the 4-digit ATC market (ATC4) level and Model 3 is the 3-digit ATC market (ATC3) level. *Price* is instrumented in all models with relevant tests reported in the table. Controls include *Vintage*, *Number of brands*, and *Number of generics*. The models are log-linear, as such the marginal effects are calculated using the equation $\exp^{(\beta-1)}$ where β is the respective coefficient on our variable of interest, (*Relabel*U.S.*). Marginal effects for our variable of interest are reported in the lower panel. Standard errors are clustered at the 2-digit ATC market level. Constants are included in all specifications but omitted from the table. * p < 0.10, ** p < 0.05, *** p < 0.01.

DV = ln(Sales)	Model 1 Focal Drug	Model 2 ATC4 market	Model 3 ATC3 market
Relabel	0.174	0.292***	0.187***
	(0.128)	(0.076)	(0.040)
U.S.	1.549***	1.557***	1.888***
	(0.128)	(0.167)	(0.153)
Relabel * U.S.	-0.277***	-0.149**	-0.145***
	(0.103)	(0.072)	(0.044)
ln(Price)	-3.225***	-0.159	-1.216***
	(0.513)	(0.126)	(0.260)
ln(Lagged promotion stock)	0.541***	0.241***	0.137***
	(0.064)	(0.031)	(0.027)
Controls	Y	Y	Y
Drug fixed effect	Y	N	N
Market fixed effect	N	Y	Y
Time fixed effect	Y	Y	Y
N	1,126	1,108	1,022
Adjusted R^2	-0.107	0.781	0.882
First stage F-statistic	15.86	240.35	20.23
Hansen J-statistic	3.808	1.779	0.457
Hansen J p-value	0.149	0.182	0.499
Marginal effects:			
Relabel * U.S.	**-0.242**	**-0.138**	**-0.135**