Generic Aversion and Observational Learning in the Over-the-Counter Drug Market[1]

Mariana Carrera[2]
Case Western Reserve University

Sofia Villas-Boas[3]
University of California- Berkeley

October 2, 2019

ABSTRACT.

In a set of four-week labeling interventions at six locations of a national retailer, we tested three hypotheses for consumer aversion to generic over the counter (OTC) drugs: lack of information on the comparability of generic and brand drugs, inattention to the price difference between generic and brand drugs, and general uncertainty about generic quality or risks that can be reduced with information on peer purchase rates. We use a difference-in-differences approach to measure the average treatment effects of each type of information on the purchases of treated products at treated stores. We find that posted information on the purchases of other customers increases generic purchase shares significantly. Consumers without prior purchases of a generic product appear particularly responsive to this information.

[1] We thank Stefano DellaVigna, Ulrike Malmendier, Molly Van Dop, Mark Votruba, and seminar participants at U.C. Berkeley, Case Western, Kent State, Vanderbilt, and Duke University for helpful comments. We thank the retailer, whom we cannot name, for sharing data and allowing us to conduct a field experiment. We thank you the Giannini Foundation for support.
[2] Department of Economics, 11119 Bellflower Road, Cleveland, Ohio. Mariana.carrera@case.edu
[3] Agricultural and Resource Economics, Berkeley, California. sberto@berkeley.edu

1. Introduction

Consumers' choices are influenced by several non-standard factors including the salience of prices and other product attributes [Chetty, Looney, and Kroft (2009), Bordalo, Gennaioli and Schleifer (2013)], the difficulty of comparing attributes across alternatives [Allcott (2013), Hossain and Morgan (2006)], and potentially biased beliefs [Bollinger, Leslie and Sorensen (2010)]. Salience and beliefs about risks are acutely important when consumers choose experience goods or credence goods, which encompass most health care treatments. Mistaken beliefs can lead to either overutilization (e.g., antibiotics) or underutilization (e.g., chronic preventative drugs) of treatments [Baicker, Mullanaithan and Schwarstein (2015)]. The difficulty of making price comparisons in the health care sector may also drive the overuse of costly treatment options in place of lower-cost options [Carrera et al. (2018)].

Despite mounting evidence of such "behavioral hazards" affecting consumers' medical decisions, little is known about how to counteract these forces and lead consumers towards evidence-based and cost-effective choices. Given the growing emphasis in the U.S. healthcare system on giving patients a more active role in the choice of their treatments, it is crucial to better understand how consumers' preferences over health products are formed and how they respond to new information. Provision of information has been shown to positively affect the take-up of some preventative health products such as birth control [Delavande (2008)], but can also backfire, as in the case of vaccination promotion [Nyhan et al. 2014, 2015] and information about product safety [Ma, Wang and Khanna (2017)].

In this paper, we test how three different types of information, posted at the point of sale, affect consumers' choices between brand and generic over-the-counter (OTC) drugs. The OTC drug market is a ripe setting for studying how consumers form preferences over pharmaceutical products. In contrast to prescription drugs, which physicians select for their patients, consumers choose OTC drugs autonomously. To facilitate the comparison of different products, the FDA requires standardized "Drug Facts" labels to be posted on every package, and visible price tags make price comparisons much more straightforward than they are in the prescription drug market. Nevertheless, more than half of the sales of familiar household drugs are for branded versions which cost 40-60% more than their generic equivalents without any treatment or safety advantage. Bronnenberg et al. (2015) argue that the high market shares of branded OTC products reflect consumers' low awareness of the existence and comparability of a generic substitute. Consumers might also perceive a greater degree of uncertainty regarding the safety or quality of the generic product, leading ambiguity-averse individuals to prefer the brand [Muthukrishnan, Wathieu and

Xu (2009)]. In addition, cognitive effort is required to locate and compute the savings offered by the generic versions, which often have different product names at different retailers.

We conducted a set of temporary labeling interventions at six locations of one national retailer, to test three hypotheses for consumer aversion to generic OTC drugs: (1) lack of awareness of the equivalence of generic drugs to their brand counterparts, (2) inattention to the price difference between brand and generic drugs, and (3) significant uncertainty regarding product quality which might be reduced by information on other customers' purchases. Perceived quality encompasses efficacy as well as other attributes such as safety or taste.

For each hypothesis, we designed product-specific information labels to display relevant facts to consumers. Labels used to test the first hypothesis displayed information on the FDA approval, bioequivalence, and active ingredients of the generic matching the brand. Labels testing the second hypothesis highlighted the brand-generic price difference in percentage terms. To test the third hypothesis, we displayed generic purchase rates for each product, calculated using pre-intervention sales data specific to each product in each store. We also introduced exogenous variation in the posted share (within store and product) by alternating weekly the length of pre-period time used to calculate the posted generic purchase rates. For two of the tests, we designed labels with two different ways of framing the relevant information in different stores.

We chose a fixed set of OTC categories to be "treated" with a label attached to the shelf price tag, and each of six treatment stores was randomly assigned to one of five label types (three different types of information, two of which had two framing variations). Using OTC sales data from previous years, we identified six similar stores to use as controls, and use a difference-in-differences approach to measure how consumers respond to the different labels. We estimate treatment effects on our main outcome of interest, the generic share of purchases, as well as the total quantity purchased of brand and generic versions combined. Using household-level sales data, we explore heterogeneity based on customers' past purchase patterns and test for post-treatment persistence of changes in purchasing behavior.

Our results are as follows. First, information on the comparability of generics has no effect on their purchase rate relative to the national brand. This null effect is surprising, but serves to alleviate the concern that our labels might increase generic purchase rates simply through a salience effect, i.e. by drawing customers' attention to the presence of a labeled generic product or reducing the time cost of identifying it. Second, we find mixed evidence for the hypothesis of inattention to the price difference between brand and generic drugs. The average generic purchase

share of treated products increased when the price difference between brand and generic versions was highlighted as a "savings" relative to the brand, but not when this difference was framed as an additional cost to choosing the brand. Third, we find strong evidence that consumers respond to information on the purchases of other consumers. This is consistent with general uncertainty about product desirability, perhaps encompassing attributes beyond stated equivalence of active ingredients. We find that the generic share of purchases increases by 8.5 percentage points, or 20% relative to the pre-treatment level. This effect is more than 60% as large as the increase associated with a price promotion on the generic store brand, and is particularly evident among brand-loyal customers. [4]

A large theoretical literature examines the ways in which uncertainty regarding product quality affects consumer demand (see Akerlof, 1970; Nelson, 1970; Wiggins and Lane, 1983; and Wolinsky, 1995). A closely related empirical literature analyzes how product quality information affects consumer behavior through branding (Montgomery and Wernerfelt, 1992), mandatory product labeling (Jin and Leslie, 2003), and advertising (Ackerberg, 2001; Ackerberg, 2003). A smaller but growing literature finds evidence of consumer inattention to non-salient components of costs, such as taxes [Chetty, Looney and Kroft (2009)] and eBay shipping costs [Hossain and Morgan (2006)]. We assume that all consumers observe the price of the brand-name product and hypothesize that they may be inattentive to, and may tend to underestimate, the savings offered by the generic product.

Our third test adds to a growing literature on observational learning.[5] There are three main channels through which an individual's demand for a good might be affected by information on other consumers' purchases: (1) goods that are socially visible may confer status, (2) there may be network benefits or "social utility" to using the same product as a peer, and (3) others' usage could

[4] A price promotion of the generic reflects, on average, an additional 9.3% savings relative to the brand price per unit. Thus, if we assume each percentage point additional savings to have the same effect on generic share, the effect of the labels is equivalent to the effect of reducing the generic's price by an additional 5.8 percent of the brand price.

[5] On the theoretical side, Becker (1991) developed a formal model in which the demand for a good depends positively on its aggregate quantity demanded, i.e. on peer demand. McFadden and Train (1996) formalize consumer learning about a new good's quality through a rational, forward-looking decision process in which they learn from their own experience and from the experiences of their peers. Moretti (2011) derives predictions from a model of social learning for the time path of movie sales after their release. He analyzes how sales diverge over time for movies that perform well or poorly relative to prior expectations and finds a reinforcing pattern: When a movie exceeds expectations in its opening week, consumers update their expectations via input from their peers, leading to greater future sales.

serve as signals of a good's quality.[6] While other experimental studies have found positive effects of peer usage disclosure, it is typically difficult to disentangle whether these effects are driven by updated quality priors as opposed to the social channels (status, network effects).[7] For example, Cai et al. (2009) found that restaurant demand for menu items increases significantly when they are identified as "Top 5" most popular items. But at a restaurant where diners eat "family-style," the orderer might seek to order items that others are likely to enjoy (instead of using the information to update his own preferences for the item). Similarly, in an artificial music market where participants select unknown songs to download, Salganik et al. (2006) found that their choices were influenced by information on each song's current download count, but this might be driven by a socially-motivated desire to know the songs that others play.

Since the consumption of pharmaceutical products is a largely private behavior, social status and network effects are arguably absent in our context. Thus, the third channel—learning about quality—is likely to be the only one through which the purchases of other customers would influence one's purchasing decision. Since most consumers do not buy over-the-counter drugs often, and seldom observe their peers' purchases, even simple statistics of generic purchase rates may contribute new and useful information, just as physicians may learn about prescription drug quality through the choices of other physicians (Saxell, 2014).[8] However, it is not theoretically clear whether disclosing market shares to customers will lead to an increase in generic share, given that they average 40-50% in the stores that we study. We use in-store customer surveys to complement our field experiment by soliciting priors on the purchase shares of other consumers and assessing self-reported likelihood of being influenced by peer market shares. Findings suggest that brand-buying customers are more likely to be swayed towards the generic by peer-purchase information than generic-buying customers are to be swayed towards the brand, most likely because they are less likely to have personal experience with both the brand and generic.

The rest of the paper proceeds as follows. Section 2 describes the background on over the counter (OTC) drug regulations and a simple conceptual framework to frame our three testable

[6] A fourth channel is the behavioral "nudge" that can come from identifying a particular behavior as an injunctive or descriptive norm. This channel is less relevant here for two reasons. First, purchasing generic OTC products is not really considered to be pro-social or an action for the collective good. Second, Cialdini (1984) and the related literature typically find that high existing shares are necessary for these messages to be effective, which is not the case in our context.

[7] An exception is Bursztyn et al. (2013), which separately estimates the influences of "social utility" versus "social learning" (updating quality expectations) in the case of paired peers purchasing financial assets.

[8] We conducted an in-store survey of customers, finding that 24% (of 294 respondents) do not know whether more of their friends use brand or generic OTC drugs. By contrast, only 5 percent respond that they do not know, when asked whether more of their friends drink diet or regular soda.

hypotheses. Section 3 describes the retail empirical setting, the experimental design and intervention, and the data used. Section 4 presents the empirical strategy and discusses the results and robustness checks, while section 5 concludes.

2. Background

2.1 OTC Drugs in the United States

Generic versions of over-the-counter drugs contain the same active ingredients as their name brand counterparts and are highly regulated. For newer drugs, each manufacturer who wishes to produce a generic version of the drug must obtain their own FDA approval prior to selling it. In the prescription drug market, the FDA tests generics for bioequivalence to the brand, defined as a similar time pattern of active ingredient release and absorption into the blood stream. Of the drugs in our sample, many, but not all, were tested for bioequivalence because they were sold in the prescription market prior to the OTC market. For older drugs, including, for example, acetaminophen (Tylenol) and diphenhydramine (Benadryl), the FDA publishes a "Monograph" specifying regulations for production, packaging, and labeling, but does not actively examine and approve the formulations sold by each manufacturer.

The FDA and clinical studies fail to find differences in safety or efficacy between versions of a drug produced by the original brand patent holder and generic entrants (see Kesselheim et al. 2008).[9] Despite this, the perceptions of consumers and, to a lesser degree, physicians, are that generic drugs are less desirable than the original brand product (Shrank et al. 2011, Shrank et al. 2009). Interestingly, Bronnenberg et al. [2015] find that pharmacists are far more likely to purchase generic over-the-counter drugs than the general population, implying that the generic's drug quality, relative to the brand, is higher than perceived by the average consumer.

2.2 Conceptual Framework

Drugs treating symptomatic conditions may be experience goods, meaning individual *i* only learns her utility levels v_{ib} (for the brand version of a given product) and v_{igx} (for the generic version of that product) after having tried each of them. We assume that since brands precede generics on the market, all buyers of product *x* have used the brand version in the past and thus

[9] An exception is drugs that have a "narrow therapeutic index (NTI)" meaning that patient response can be sensitive to very small differences in the timing and speed of ingredient absorption. No over-the-counter drugs are considered NTI.

know their private valuation v_{ibx}.[10] Those who have not tried the generic version do not know their private valuation v_{ig}, but have some expectation over it, $E[v_{igx}]$. We do not assume this expectation is unbiased, since lack of information, prior experience in different markets, and differential advertising could create bias.[11] A risk-neutral shopper who has not yet tried the generic will continue to purchase the brand version if:

$$(1)\ v_{ibx} - E[v_{igx}] > -\alpha_i (p_{bx} - p_{gx})$$

where α_i represents sensitivity to price, p_{bx} and p_{gx} are the prices of the brand and generic, respectively[12].

Each of our interventions aims to shift either $E[v_{igx}]$ or α_i. By displaying information on the active ingredient comparability of brand and generic drugs, we test whether $E[v_{igx}]$ is based on inaccurate knowledge of this. By displaying the typical price difference in percentage terms, we test whether increasing the salience of $p_{bx} - p_{gx}$ increases the consumer's response to the price difference (α_i).[13] By displaying the share of other customers who buy the generic or brand version, we test whether $E[v_{igx}]$ can be influenced by observational learning.

Observational learning is the process by which consumers update their expectations of the quality of a given good after observing others' decision to purchase it (or not). In our context of choice between two competing versions of a product, observing the share of other customers who buy the generic serves as a signal of the generic's popularity relative to the more expensive brand version. To the extent that this signal exceeds (or falls short of) a customer's prior guess about the share of shoppers who buy the generic, the customer might increase (or decrease) her estimate of the generic's value.

Thus, the effect of posted generic shares could be different on customers who typically buy the brand versus those who typically buy the generic, for two reasons. First, these types of customers may have different priors about generic drugs' popularity, and second, those who have never before purchased the generic might have considerably weaker priors about its quality.

---

[10] In consumer surveys we found that 94% of customers reported having purchased the brand version of their preferred headache remedy at least once in the past, whereas only 65% had ever purchased any generic version of it (i.e. not restricted to the specific store-brand of the retailer we studied).

[11] For example, immigrants from countries where generic drugs are unregulated and often counterfeit may be unaware that U.S. generics are regulated by the FDA.

[12] Although we do not explicitly model risk aversion, it is easy to see that if v_{ibx} is known and v_{igx} is unknown, then as risk aversion increases, an individual becomes less likely to choose the generic and may continue to purchase the brand, even if the inequality in (1) goes the other direction.

[13] Highlighting the "savings" associated with buying the generic might also increase generic purchases by increasing the perceived deal-value of the product (described as transaction utility by Thaler (1985).

Consider the decision of a consumer who has previously used both the brand and generic versions of a drug. If they are pure experience goods with fixed private utilities, this consumer's past experiences have resolved all uncertainty regarding her private utility from each one, and the purchases of other people should be irrelevant to her next choice. On the other hand, this consumer may still perceive v_{igx} and v_{ibx} as uncertain; for example, there may be some risk that the manufacturer has produced a bad batch, or that long-term adverse effects of the generic drug have not been realized. To the extent that such uncertainty exists, the perceived utility from purchasing a drug might be sensitive to peer purchase information even after a consumer has tried it.

Suppose the posted generic purchase share is 50% and all customers overestimate the share of other customers who make the same choice as them. If we find that generic shares increase when popularity information is displayed, two possible explanations are that (1) generic-buying customers put less weight on the behavior of their peers as a signal of a drug's quality, consistent with the notion that these products are experience goods, and/or (2) brand-buying customers have priors farther away from the posted generic shares. In Section 5, we discuss customer surveys we conducted to explore these distinct possibilities, with results suggesting that generic-buying customers put less weight on the behavior of other customers as a signal of a drug's quality.

3 Setting and Intervention

3.1 Retail Setting

We conducted a four-week intervention in six northern California locations of a national supermarket chain. The retailer offers its own store-label (i.e. in-house brand) household and food items in addition to over-the-counter drugs. A large share of the locations have in-store pharmacies, where consumers can ask a pharmacist questions about any drugs. However, a pre-treatment survey we conducted suggests that only about 5% of OTC customers seek the input of pharmacists.[14] The shelf layout of OTC products is largely uniform across stores, with store-label versions often, but not always, placed adjacent to their brand counterparts. We were given

[14] We conducted an in-store pre-treatment survey of pharmacists in the store locations selected for treatment. Pharmacists reported that approximately 10 customers per week approach them to ask questions about OTC products. Our household-level sales data indicate that at these stores, the average number of loyalty card holders purchasing an OTC product each week is 190. Thus, approximately 5% of OTC customers seek the input of pharmacists. Nevertheless, we made sure that all treated stores had in-store pharmacies and that the typical response of pharmacists to customer inquiries was consistent in these stores.

permission to post labels beneath the price tags of generic versions of selected products (see Figure 1 below).[15]

Figure 1. Example of Label Placement

3.2 Experimental Intervention: Three labeling tests

The four-week intervention consisted of posting labels beneath the price tags of generic versions of products in treated categories in treated stores. We now present the content of the labels for three tests, two of which had two different versions, numbered as tests 1, 2a, 2b, 3a, and 3b, corresponding to our three hypotheses. Tests 2b and 3b differ from tests 2a and 2b only in how the same information was framed. With the exception of Test 3a, which was conducted in two different stores, each of these tests was conducted in one store.

Test 1. To test the hypothesis that consumers lack basic information on brand/generic drug comparability, we created labels that described the similarity between brand and generic products as specifically as possible. Each treated drug product received one of the three labels displayed in Figure 2. The strongest statement we used was: *"The FDA determined this product to be therapeutically equivalent and bioequivalent to* [corresponding brand product]," taken verbatim from the FDA approval letter, for drugs with such approval letters available on the FDA website. The second statement used was *"This product contains the same active ingredient as* [corresponding brand product] *and has been approved by the FDA,"* shown with the reference

[15] Labels were reposted each week after price tags were updated by the retailer. We thank Yann Pannasie, Raymond Gong, Karen Yao, Caitlin Crooks, Feyisola Shadiya, Brian Mitchell, Roni Hilel, Kyle Kennelly, Jonathan Arenas, Kathy Hua, Fanglin Sun for research assistance in gathering data and implementing the label experiment. We also thank Ishita Arora, Samantha Derrick, Kathy Hua, and Ye Zhong for helping us gather auxiliary data and perform in-store surveys.

number and date of FDA approval. This label appeared on products for which we found notices of FDA approval, but either no electronically available letter, or a letter that did not include any statement about bioequivalence. The third statement, which was posted for older-generation drugs whose manufacturers need not seek explicit approval from the FDA prior to marketing a generic, was *"This product contains the same active ingredient as* [corresponding brand product]*."*[16]

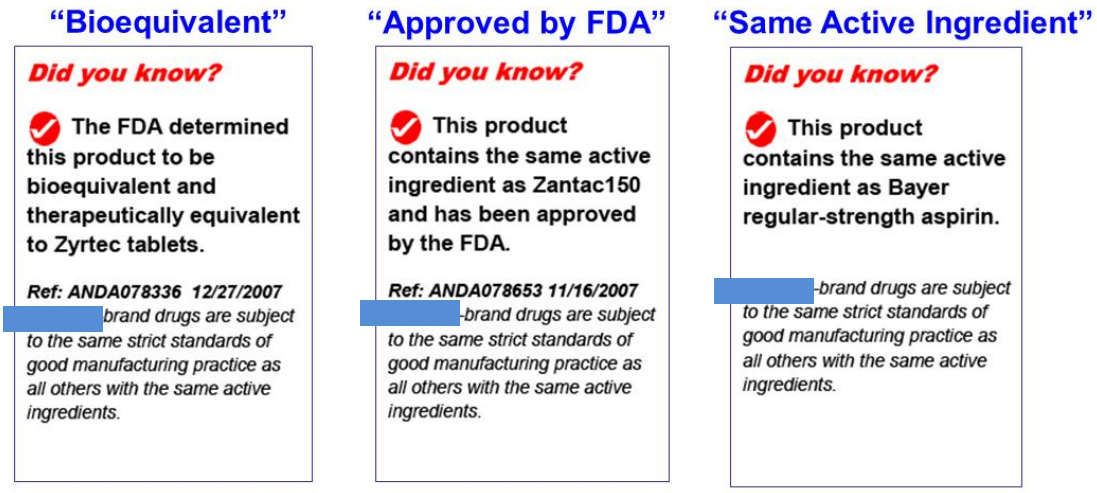


Figure 2. Labels highlighting comparability/quality

Test 2a. To test for inattention to price differences, we posted labels stating "*Customers who choose this product save X%"* with a footnote specifying that the savings was relative to the specified brand product per dose. X ranged from 14% to 68% in the products labeled and an example of one such label is below.

[16] As described in Section 2, for older-generation drugs such as acetaminophen or aspirin, rules regarding the production of the drug are reported in an FDA monograph, and new manufacturers are not required to apply for approval to market their own versions of such drugs. Thus, we could not use a statement as strong as the others for these products.

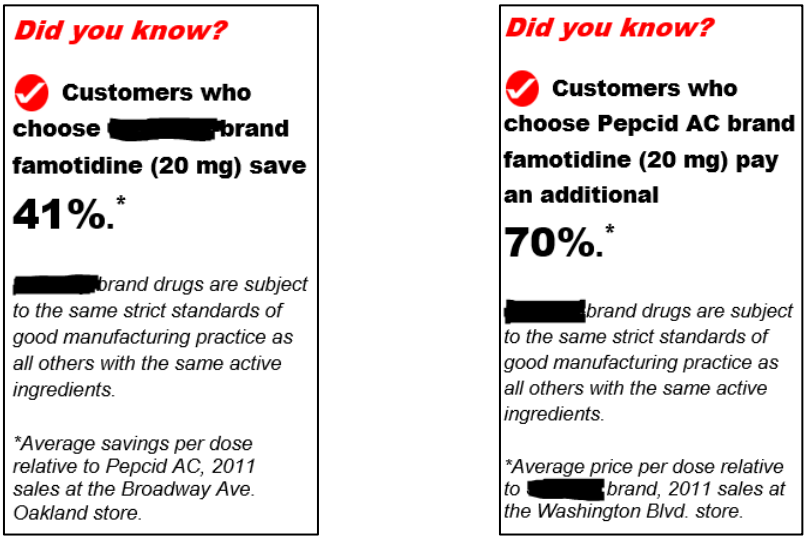


Figure 3. Labels highlighting the price difference.

[On left: Test 2a label example; On right: Test 2b label example]

Test 2b. In another store, we highlighted the price differences in a different way, by stating "*Customers who choose* [corresponding brand product] *pay Y% more than the generic alternative*." In this type of label, the price difference is framed as a loss rather than a gain. Also, for the same brand and generic prices, Y will be a larger number than X, because the generic price is a smaller denominator. For these reasons, we hypothesized that Test 2b would have a stronger effect than Test 2a. Note, however, that the label was placed below the generic product, as we were not permitted to place labels below branded products.

Test 3a. To test for observational learning, we posted labels stating *"X% of customers in this store choose this product instead of* [corresponding brand product]." The values of this share were calculated for each product and each store, using either the previous year's sales data (Jan-Dec 2011) or the first three months of the current year (Jan-March 2012). To obtain quasi-exogenous variation in the value of the share displayed, holding constant the product and the store, we alternated which method of calculation was used in each store's labels, each week. An example of one such label is below.

Figure 4. Labels with Generic purchase rate of Peers

[On left: Test 3a label example; On right: Test 3b label example]

Test 3b. An alternate way to frame the information displayed in Test 3a is to report the share of customers who buy the brand product, e.g. *"Y% of customers in this store choose* [corresponding brand product]" *instead of this product."* If the mere act of bringing attention to the purchase of a specific product leads consumers to buy it, or if the statement is read as an implicit endorsement of a particular product, then Test 3b could have a different effect than Test 3a. If, instead, both labels only affect purchases insofar as they shift customers' beliefs about what others buy, then Tests 3a and 3b should have the same effect.

3.3 Data and Summary Statistics

We use two types of data obtained from the retailer. First, we use a panel data set of store-level weekly sales for OTC drugs at the product version (UPC) level for all 60 stores in the geographic divisions that include our treated stores. Our main pre-period, for which we have store-level data available for all these stores, is the 6 weeks prior to treatment: weeks 14 to 19 of 2012 (the experiment was conducted during weeks 20-23). We also have data covering all stores for the same set of weeks in 2010 and 2011, which we use to conduct placebo tests, and post-intervention data from weeks 24 to 38 of 2012.

Second, we obtained a transaction-level dataset with household identifiers, based on the use of customer loyalty cards. This dataset includes the six treated stores and six stores that were pre-selected as a similar set of comparison stores, and spans the period from 2011 week 23 (one

year preceding our last treatment week) to 2012 week 38 (15 weeks after our treatment ended). This data allows us to investigate how the customers shopping during the treatment period changed their behavior relative to their prior purchase patterns, whether different types of shoppers responded differently, and whether any changes persisted beyond the intervention.

OTC Product Classes

We focus on the twelve largest OTC drug categories that offer generic (store-label) versions for the majority of the products. These twelve categories can be broadly grouped as pain relief, allergy relief, digestive/stomach relief, and relief for acute ailments such as colds and cough. In order to ensure that closely substitutable products were either all treated or all untreated, we defined treatment at the drug class level (for example, second-generation antihistamines) but also left at least one drug class within each of these four groups as untreated, for a more balanced representation.

For the six treatment stores, we were given pre-treatment data from January to May, 2012, which we used to compute mean percent savings and mean generic share at the store-product level, to be displayed in the labels for Tests 2 and 3. We report summary statistics for the twelve drug categories in Appendix Table A1.

Store-level data

The data contain the gross revenue, net revenue (net of promotions) and the total quantity sold of a particular UPC in a given week in a given store, with which we calculate each item's gross price and net price.[17] Prices are similar within price divisions and all stores in the treatment and control belong to one of two price divisions. Prices change at the same time each week for all stores in the same price region. For the rest of the paper, we use "price" to refer to the net or promotional price, given that this is the price faced by the majority of the customers.[18]

We collapse our data to the level of the active ingredient and dosage combination rather than the UPC. For example, if 500mg acetaminophen is sold in quantities of 12, 30, and 100, and in gelcaps as well as tablets, we will combine all of these UPCs into one observation, but another

[17] The revenue is obtained as two columns (net and gross) in the raw data that are equal to each other if the product was not on promotion during a certain week in a certain store. Those two revenue columns will differ if there are promotions: the net column will feature a smaller dollar value than the gross column. If we divide both revenue variables by the quantity sold, we obtain the gross shelf price and the average promotional price.

[18] A loyalty card for this retailer is required to obtain the promotional price. Since our household level data is obtained through loyalty card purchases, we can compute the share of purchases that are made using a loyalty card, and thus, at the promotional price. Across the six treated stores, this share varies from 79% to 86% with a mean of 83%.

strength acetaminophen would be grouped separately. We use the term "product" to refer to a set of brand and generic formulations with the same active ingredient at the same dosage level. Thus, we refer to product "generic share" as the share of quantity purchased, within this set of equivalent products, for the store's private label (generic) versions. We compute an average price per unit (i.e. per pill) for the brand and generic versions of each product by dividing the sum of net price paid by the total unit quantity sold in each store-week, and also the price of the generic version as a share of the unit price of the brand version. Lastly, we create indicator variables for "brand on sale" and "generic on sale" which equal 1 if any of the UPC's for the brand or generic versions of the product are offered at a promotional price.

Table 1 reports descriptive statistics across the stores assigned to Test 1, Test 2, Test 3, and the control group, during the pre-treatment period. As we would expect, there is no significant difference in prices across these stores. Sales quantities, however, tend to be lower in the stores assigned to Tests 2 and 3 than in the Test 1 store and the control stores. Our main outcome variable, the generic share of purchases, appears significantly smaller in the Test 2 stores than in the control stores, for treated products. The average share of products with at least one version of its name-brand packages being sold on promotion in a given week (*salebrand*=1), is more than half in all groups of stores. The average share with at least one of its store-brand packages being sold on promotion (*salegeneric*=1), is significantly smaller in the Test 2 and Test 3 stores than in the Test 1 and control stores. This is likely explained by two reasons, which are both related to the fact that Test 2 and Test 3 stores sold smaller quantities overall. First, we can only observe a UPC's promotional price when at least one customer purchases that UPC. Second, larger stores tend to sell a greater range of UPCs (packages varying in size or form) of any given product, so it is more likely that one is on sale in a given week.

Household-level data

Longitudinal purchases at the household-level data are available through purchases made via loyalty cards. At this retailer, discounts posted throughout the store are only available via loyalty card, so their use is near ubiquitous (e.g. 89% of OTC drug purchases).

Household-level data confirms that there is strong habit persistence in the choice of brand versus generic formulations. Of the 27,085 household-drug combinations that we observe with two

or more purchases of a treated drug, 51% only ever buy the brand.[19] Of the remaining 49%, 80% buy the generic in the first purchase that we observe, and this group continues choosing the generic at a rate of 83% in future purchases.

4. Empirical Specifications and Results

4.1 Effects of Labeling Interventions: Store-level analysis

In the store-level analysis, we measure the effects of the labeling interventions on OTC drug purchases at the store-week-product level. The two outcomes of interest are the generic share of each product sold, computed as $gs=(q_{gen}/(q_{gen}+q_{brand})$, and the total quantity sold $Q=(q_{gen}+q_{brand})$.[20] Using a differences-in-differences approach, we examine the effect of our treatments by comparing the change in the sales of treated OTC products from the six-week pre-treatment period to the four-week treatment period, in the treatment stores versus the control stores. We also implement a triple difference-in-differences by comparing the difference-in-difference of treated products to that of untreated products.

To illustrate the approach, we first report the pure difference-in-differences in generic share for all of our labeling interventions, pooled together as the "treated" set of stores, in Table 2.[21] This table shows the generic shares, averaged across weeks, products and stores, for the treatment period and the pre-treatment period, in treatment stores and control stores, among both treated and untreated products. The number of observations in each cell represents the number of product-store-week observations in the treated or untreated group of products, stores, and period.

The top panel, corresponding to treated products, shows that in the pre-treatment period, mean generic shares were 48.8% in the control stores and 45.5% in the treatment stores, an insignificant difference. From the pre-treatment period to the treatment period, average generic shares of treated products increased by 4.6 percentage points in the treatment stores and decreased by 1.2 percentage points in the control stores. The increase of average generic share in the treatment stores was marginally significant, as was the difference-in-differences ($DD_{t.p.}$) estimate

[19] This share is not very responsive to increasing the minimum number of purchases. With a minimum of 5 purchases of the same product, 46% have only bought the brand, and with a minimum of 10 purchases, 42% have only bought the brand.
[20] Note that we focus on the number of packages purchased as *quantity*, rather than the number of daily doses.
[21] Comparable results for Tests 1 and 2 are less informative and are shown in the appendix.

of 5.8 percentage point increase in treated products within treated stores pooled together, relative to control stores.

The bottom panel shows the parallel comparisons for untreated products. Among these products, the change in generic share from the pretreatment period to the treatment period was -0.014 in treatment stores and 0.001 in control stores, leading to a negative but insignificant difference-in-differences ($DD_{u.p.}$) estimate of -1.5 percentage points.

Lastly, the table shows the triple-differences estimate, following Chetty, Looney and Kroft (2009), which is the difference between $DD_{t.p.}$ and $DD_{u.p.}$. For the three interventions pooled together, the estimate is 0.073 and marginally significant. This measure is not our primary focus, however, for two reasons: First, each of the three labeling tests could have a different impact, and second, it is plausible that by highlighting different aspects of store-label generic drugs, the treatments could have spillover effects on untreated products.[22]

In the tables that follow, we separately report the second difference estimates for treated products and untreated products, for generic share and total quantity sold, for each of our three labeling interventions. We add store-level and product-level fixed effects and estimate coefficients for three treatment dummies: T_1 T_2, and T_3, which are interactions of the treatment time dummy t_t and indicators for store s being one of the stores treated with Test 1, Test 2, or Test 3 labels, respectively. The equation estimated in the odd numbered columns of Table 3 is:

$$(2) \quad Y_{jst} = \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + t_t + \delta_j + \delta_s + \varepsilon_{ist}$$

where Y_{jst} denotes the quantity (Q_{ist}) or generic share (gs_{ist}) of purchases of product *i* in store *s* in time *t*, δ_s denotes store fixed effects to control for store-specific constant factors, δ_j denotes product fixed effects, and t_t is a "treatment time" dummy that is equal to one during the treatment month and equal to zero during the pre-treatment month. The coefficients on T_1 T_2, and T_3 can be interpreted as average treatment-specific changes between the pre-treatment month and the treatment month, after controlling for changes over this same time period in the control stores.

In a second specification, shown in even number columns, we add controls for whether any generic versions, brand versions, or both generic and brand versions of product *i* were on sale in store s during week *t*. Also, for the specifications where generic share is the outcome, we weight each product's observations by the total quantity sold during the period 2011wk1-2012wk12.

[22] The spillover effects could be either positive, if positive information about labeled products leads customers to infer similar positive information about unlabeled products, or negative, if customers infer that the labeled products were chosen based on having more positive attributes than the unlabeled products.

Table 3 reports the results of these regressions. The results in Columns 1 and 2 show that only Test 3 had a positive and significant effects on generic share of treated products, representing a sizeable 6.8-7.2 percentage point increase in generic share. The estimated effect of Test 2 on generic share is smaller and statistically insignificant, and the estimated effect of Test 1 is close to zero. By contrast, we find no significant differences between treatment and control stores in the pre-post changes in either outcome for the products which did not receive labels.[23] For untreated products, Columns 3 and 4 show that the estimated effects of all three tests are very close to zero. Columns 5-8 show the same specifications estimated for the outcome of total item quantity for treated and untreated products. Results show that Test 2 had a statistically significant effect, raising the average quantity purchased by slightly more than 1, relative to a baseline average quantity of 10.3. The magnitudes of the effects of Test 1 and Test 3 on quantity are similar but statistically insignificant on treated products, and the estimates in Columns 7-8 suggest no significant difference-in-differences for the quantity of untreated products.

Note that in Table 3, we constrain the effects of Test 2 and Test 3 to be equal across the two ways of framing the information that were used in different treated stores. In Table 4, we disaggregate these results by framing of the message content. The disaggregation of Test 2 shows that one of the methods of framing the price difference ("save x%" relative to the brand) has much larger point estimates on both generic purchase share and quantity purchased, although these effects are only marginally significant. The other framing method ("pay an additional y%" relative to the generic), by contrast, had small and insignificant point estimates for both outcomes. We interpret this as suggestive evidence that the "Save" framing might be more effective, due to its use of the familiar and attractive promotional wording of "save", whereas the other variation, with which we sought to provoke loss aversion, presented the information in a less familiar way. Also, we only had permission to post signs beneath the generic products. It is possible that putting "Pay an additional _%" signs beneath the generic products caused confusion regarding which products involved paying an additional amount.

Varying the framing of Test 3 did not change the impact of posted peer purchases on generic share. The point estimates for the two framing variations are very similar. This helps elucidate the mechanisms behind observational learning in this context. If the results had suggested that the statement "55% of customers buy the generic, " for example, had a stronger effect on generic share than the statement "45% of customers buy the brand," then we might infer that part

[23] In theory, we could have found spillover effects from the labels, to other generic products in the same store.

of the impact was driven either by pointing out the existence of the brand or generic product, or through an implicit perceived encouragement to buy the brand or generic. The fact that these statements have equivalent effects on generic share, however, supports the interpretation that these labels provide an informative signal that consumers use to update their expectations of the quality of the generic product: in a binary choice situation, 55% of customers buy the generic product conveys the same information as 45% of customers buy the brand product.

Since we also randomly varied whether the generic purchase shares displayed to consumers were calculated using 2011 sales data or data from the first quarter of 2012, we have some exogenous variation in the share displayed. Figure 5 graphically shows the correlation between the difference in posted share and the week-by-week differences in treatment effects for the Test 3 treatment stores and products. Each point in the scatterplot corresponds to one product (active ingredient) in one of the three stores that were treated with Test 3 labels. Note that we are underpowered in this analysis, due to the fact that the items with greatest differences in store-level purchase shares of 2011 and early 2012 tend to be items that are purchased less frequently, driving greater variance in their generic purchase shares. Nevertheless, the best fit linear prediction has a positive slope, suggesting that every 10 percentage point increase in the posted generic share increases the predicted generic purchase rate by an additional 6 percentage points.

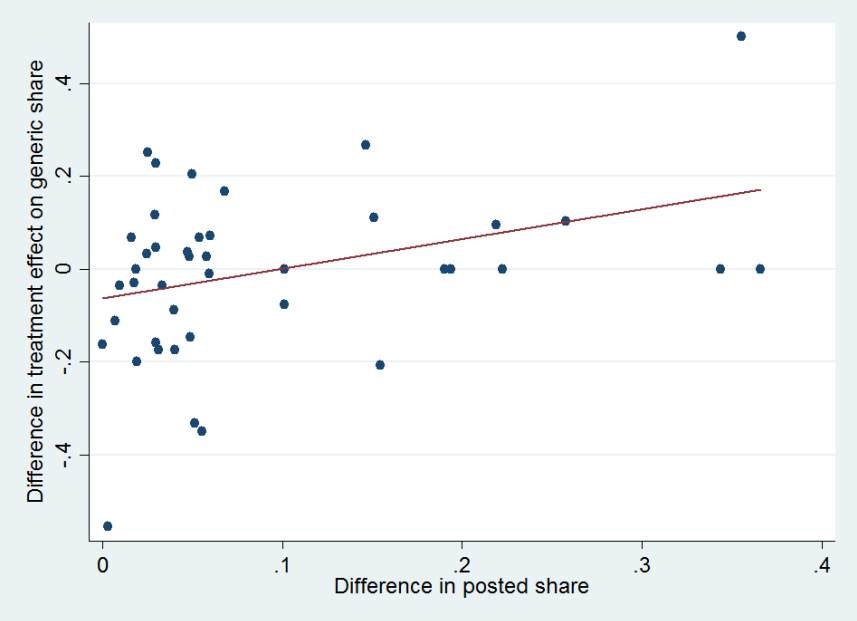


Figure 5. Effect of posted share of peers on current generic shares

4.2 Event-Study Model

The difference-in-differences model estimated in the previous section is identified under an assumption of parallel trends, i.e. assuming that sales in treated stores and untreated stores were not trending differently prior to the labeling treatment. To assess this, we estimate the following event study model over the weeks for which we have store-level data for all stores. For each of the three labeling interventions (m=1,2, and 3) and for treated and control products we estimate:

$$(3) \quad Y_{ist} = \sum_{T=-3}^{1} \beta_T D_{isT,m} + \alpha_i + \alpha_s + \alpha_t \quad + \epsilon_{ist}$$

where $D_{ist,m}$ is a dummy variable equaling one if store *s* received labeling intervention *m.* The biweek-of-sample t identifies time periods in two-week intervals relative to the first two weeks of the treatment (i.e., t=0 for the first two weeks of labeling treatment). The β_t vector contains the coefficients of interest, which we plot over time to trace out the differences between treated and untreated stores from before the treatment labeling implementation to the end of it. Importantly, if products are trending similarly in treated stores and control stores before the treatment periods, there should be no trend in the β_t coefficients in the pre-treatment period and their values should be statistically indistinguishable from zero.

Figure 6 plots the estimates we obtain from equation (3), with the $\widehat{\beta_t}$ plotted in black and the 95 percent confidence intervals, based on standard errors clustered at the active ingredient level, plotted in gray dotted lines. Vertical lines separate the sample into two weeks subperiods during the pre-treatment and treatment periods. The omitted period is the two weeks prior to the start of treatment. In the periods before the labeling treatment, none of the β_t are statistically different from zero at the 95% significance level. Our store-level dataset does not extend beyond the treatment period.

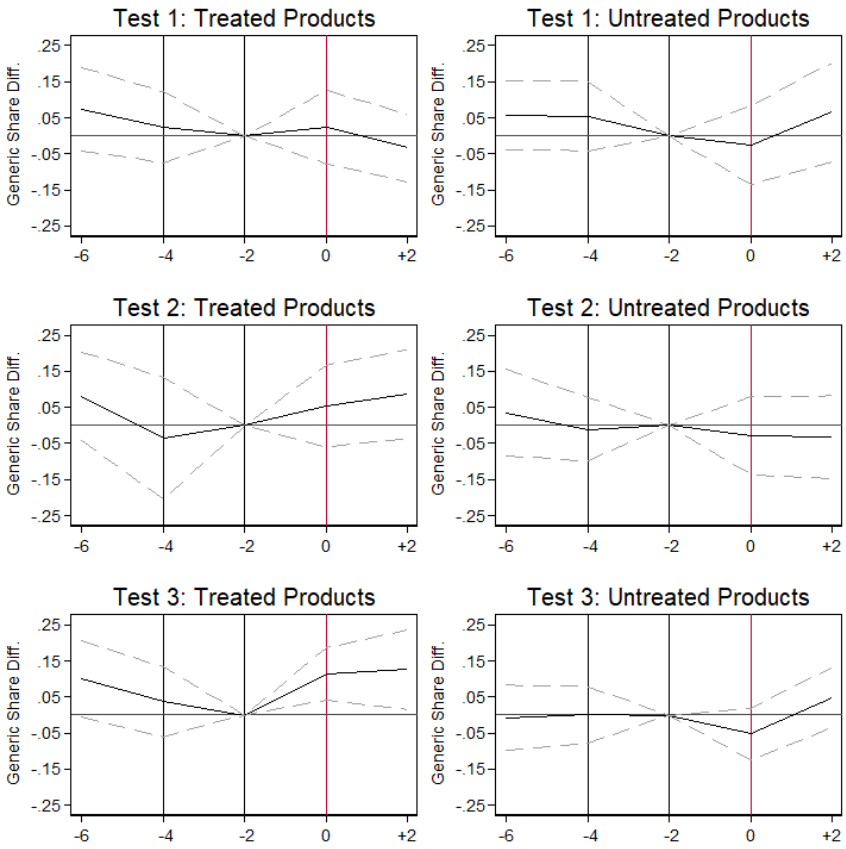


Figure 6. Event-Study Analysis of Generic Share

4.3 Customer-level analysis

An important concern arises when the analysis is restricted to store-level totals: We cannot be sure whether existing buyers of OTC products within a store are shifting their purchases towards generics, or if the labels attract different customers, who may already be buyers of generics at other retailers. Customer-level data allows us to control for past shopping choices of the customers observed during the treatment period. It also allows us to test whether the labels have different effects on consumers who have previously purchased generic versus branded OTC products.

For this analysis, we focus on the purchases made by individuals whose loyalty cards show a prior purchase of at least one OTC drug in our sample from week 48 of 2011 to week 12 of 2012, although it is not necessarily the same type of drug they are purchasing during the treatment or pre-treatment period. We estimate a linear probability difference-in-differences model similar to the store-level equation (1) in Columns 1 and 4 of Table 5, for treated and untreated products, separately.

The equation is:

$$(4)\quad Generic_{ijst} = XB + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \delta_t + \delta_j + \delta_s + t_t + \varepsilon_{ijst}$$

where $Generic_{ijst}$ is equal to 1 if the store-brand version of product j is purchased by customer i in store s in time week t. Controls in X include dummies for whether brand, generic, or both versions of the product were on sale during a given week in store s. As in the store-level analysis, t_t identifies the four-week treatment period; T_1, T_2, and T_3 are the treatment period interactions with the stores used in tests 1-3; δ_i are product fixed effects and δ_s are store fixed effects.

Table 5 shows the results of this estimation, which qualitatively matches those of the store-level analysis. In the first column, the point estimate of the effect of Test 1 is negative and insignificant, the estimate for Test 2 suggests a borderline-significant 4.4 percentage point increase in the probability of choosing a generic OTC drug over its brand-name counterpart, and the estimate for Test 3 suggests a 6.8 percentage point increase which is significant at the 5% level. Column 4 shows that none of the three tests had significant effects on the probability of generic choice for untreated products.

Next, we augment this model with individual-level controls defined by past shopping behavior during the six-month period preceding the start of the intervention. These are the total number of OTC purchases, the share of those purchases that were for generic products, an indicator for having purchased product *j* (brand or generic version), an indicator for having purchased any generic version of product *j*, and the average percentage discount on total purchases at the store. These covariates have significant explanatory power, raising the R-squared of the regression from .12 to .37 for both treated and untreated products. If the type of customer making purchases changed between the pre-test period and the labeling test period in a different way, in treated stores versus control stores, we would expect the estimated treatment effects to shrink between columns 1 and 2, but they are virtually unchanged.

In columns 3 and 6, we test for heterogeneity of treatment effects by prior purchasing experience. Given that symptom-treating over-the-counter drugs are experience goods, customers who have already purchased this store's generic version of a certain drug may be much be less responsive to the information provided by all of the labeling tests. To test this, we add an interaction term between each test and the indicator variable any_samegen.[24] The baseline effect of Test 3, now representing the effect only on customers who have never before purchased the generic version of a particular product, increases slightly to 0.082, representing a 8.2 percentage

[24] We also add interactions between any_samegen and treat_time, and any_samegen with indicators for the Test 1, Test 2, and Test 3 stores.

point increase. Since the baseline probability that customer of this type will choose the generic is 32 percent, this is a sizeable effect.[25] Consistent with the hypothesis that customers with prior purchases of the generic version are less responsive, the coefficient of the interaction term between Test 3 and the dummy variable for making a prior generic purchase is large and negative, albeit imprecisely estimated. The point estimate, combined with the fact that customers who have previously purchased the generic version of product j chose the generic 82 percent of the time, suggests a smaller relative increase in generic purchasing for this type of customer. The results for Tests 1 and 2 indicate no significant effect for either type of customer.

4.4 Post-treatment effects on customers exposed to labels

In Table 6, we use household-level data on purchases made after the end of the labeling period to test whether customers exposed to the labels showed any persistent changes in their generic purchasing rates. For three consecutive four-week intervals following the four-week treatment period, we estimate the following model, using data from the four-week pre-test period and the four-week period of interest. The model matches equation (2) except PT_1, PT_2 and PT_3 represent interactions between treated stores and a post-time dummy in place of the treat-time dummy.

$$(4)\quad \text{Generic}_{ijst} = XB + \beta_1 PT_1 + \beta_2 PT_2 + \beta_3 PT_3 + \delta_t + \delta_j + \delta_s + \text{post\_time}_t + \varepsilon_{ijst}$$

The estimation coefficients of this equation, run separately for treated products (Col. 1) and untreated products (Col. 3) are shown in Table 6. In this specification, however, it is possible that any true learning or persistence is masked by the fact that only twenty percent of the shoppers in any subsequent four-week period were present in the OTC drug aisles during the four weeks in which the labels were posted. To hone in on the effect among customers directly exposed to the labels, we also calculate a triple-difference estimate for the post-intervention effects of each test, i.e. the difference between the effects of the post_time * treat_store interactions between customers who made a purchase of *any* OTC drug (not necessarily one that received a label) during the treatment time period, relative to customers who did not make any purchase of OTC drugs during the treatment time period. Note that this approach assumes the labels did not change the propensity of shoppers in the store at a given time for purchasing an OTC product.[26]

[25] The given probability is the average for all customers with any_samegen=0 purchasing treated products in the pre-test period across all six treated stores.

[26] In Appendix Table A2, we show that the "exposed" customers, using this definition, did not appear to have different propensities to choose generic products prior to the treatment period, compared against customers at the

(5) $Generic_{ijst}$

$$= XB + \beta_1 PT_1 + \beta_2 PT_2 + \beta_3 PT_3 + \beta_4 PT_1 * ShoppedTreatTime$$
$$+ \beta_5 PT_2 * ShoppedTreatTime + \beta_6 PT_3 * ShoppedTreatTime$$
$$+ \beta_7 T1Store * ShoppedTreatTime + \beta_8 T2Store * ShoppedTreatTime$$
$$+ \beta_9 T3Store * ShoppedTreatTime + \beta_{10} ShoppedTreatTime * post_t$$
$$+ \beta_{11} ShoppedDuringTreatTime + \delta_j + \delta_s + post_t + \varepsilon_{ijst}$$

The results from this equation are shown in Columns 2 and 4 of Table 6 for treated and untreated products, respectively. Table 6 shows estimates for each of three four-week periods following the labeling experiment. Estimated effects appear noisy and generally insignificant, so we cannot draw firm conclusions about the existence of any persistent effects following our labeling tests.

4.4 Robustness

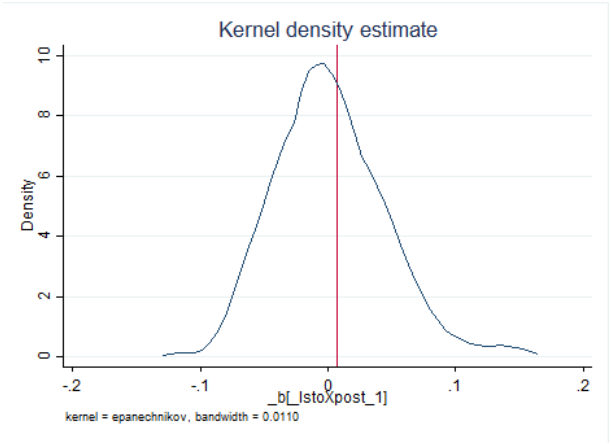
We tested for serial correlation in store-product observations at the week level, and could not reject the null hypothesis that there is no serial correlation in generic share. Although the test detected serial correlation in quantity purchased by store-item combination, our results using quantity as an outcome do not change when correcting for serial correlation at this level, and our results for generic share do not change when including store-item quantity as a regressor (see Appendix Table A3).

To verify that our results for generic share are not driven by short-term seasonal trends specific to the stores that we treated, we conducted a placebo test. We used sales data from the same months in 2010 and 2011 to estimate the placebo difference-in-differences effect on generic share in the stores we treated in 2012. Results are shown in Appendix Table A4, revealing no discernible placebo effect.

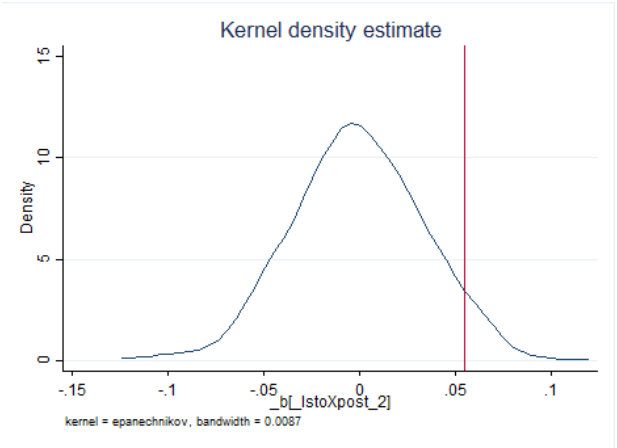
Finally, we used a randomization inference test to nonparametrically estimate p-values for the main estimates of our three treatment effects. We used the data from the pre-treatment and treatment periods in all 60 stores. In each of 500 permutations, we randomly drew one store to be assigned Treatment 1, two stores to be assigned Treatment 2, three stores to be assigned Treatment 3, and six stores to be used as control stores. The regression based on Equation (2) was re-run for each of these draws, and we plot in the three panels of Figure 7 how the point estimates for the three treatment effects compare to the point estimates obtained over 500 draws. These results show

same store who did not make an OTC purchase during the treatment period. In addition, we generally find no significant effect of the labels on total quantity purchased (see Table 3).

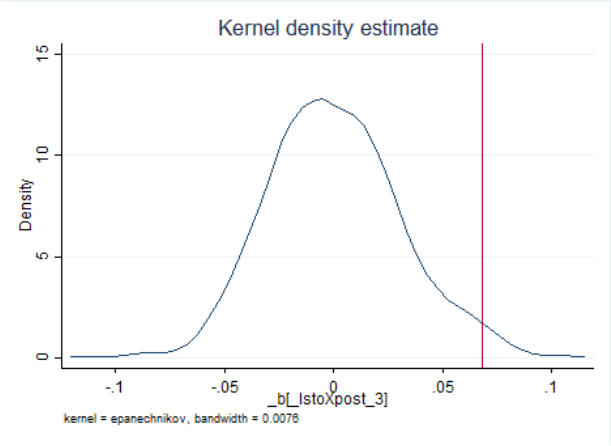
that for Treatment 3, only 4.2% of randomly selected permutations yield a treatment effect as large as the one we obtain.



Panel A: Labeling Treatment 1



Panel B: Labeling Treatment 2



Panel C: Labeling Treatment 3

Figure 7. Distribution of Point Estimates Drawn from Randomly Chosen Assignments.

Note: Point estimate from the actual treatment/control assignment shown as a red vertical line.

4.5. Mechanisms for the effect of information on peer purchases

As discussed in Section 2.2, a very simple model of observational learning, in which all shoppers similarly weight the choices of others as a signal of a good's quality, would predict that the posted share should have opposite effects on people whose priors of their peer shoppers' generic purchase shares are above, versus below, the posted shares. This could lead to a near-zero net effect of the labels on generic share when the current share of generic customers is around 50% and priors are unbiased. We described two plausible reasons why posted shares close to 50% could lead to an increase in the generic purchase share: (1) generic-buying customers may put less weight on the choice of their peers as a signal of drug quality, perhaps because they are more likely to have tried both the brand and generic already, or (2) brand-buying customers may have priors of the generic purchase share that are further from the truth than the priors of generic-buying customers, on average.

To explore the potential relevance of these two explanations, we surveyed 298 customers in three of the retailer's locations.[27] Customer participants were first asked to make a hypothetical choice between the brand and generic versions of an over-the-counter painkiller, with typical price values shown. Then, they were asked to guess what share of other shoppers at the store would make the same choice that they did. Next, respondents were asked to consider hypothetical information stating that the share of other shoppers making the same choice they made was smaller than they had guessed, and asked whether they would still choose the brand (or generic) if this hypothetical information were true, or if they would consider switching to the generic (or brand) product.

Before discussing the findings that pertain to the two hypothesized explanations for the effect observed in the field experiment, we note that the survey responses generally support the assumptions of our conceptual framework: 94% of respondents have purchased the brand version of the painkiller at some point in the past, confirming that most shoppers have experience with (and thus, less uncertainty about the therapeutic value of) the brand-name product. By comparison, only 65% of respondents (and 36% of brand-choosing respondents) have ever purchased a generic version of the painkiller. Among consumers who have tried both the brand and the generic, 81%

[27] All surveys were conducted on one Saturday, via tablets handed to respondents. As compensation, respondents received a $5 gift card to the retailer. We avoided stores that were treated during the treatment period to reduce the probability of surveying customers who had seen the posted labels.

make a hypothetical choice of buying the generic at the typical list prices shown to them, similar to the patterns observed among repeat consumers in the household-level dataset.

Interestingly, it is not the case that most brand-buying consumers believe the brand works better; only 26% report this belief. The plurality of brand-buying consumers, 46%, chose "Don't Know" or refused to answer whether they believe that the brand works better than the generic at relieving pain, that the generic works better than the brand, or that they worked equally well (see Appendix Table A5). Furthermore, of the brand-buyers who believe that the brand and generic work equally well to relieve pain (26%), fifty-nine percent answered "Don't know" or refused to answer when asked which one is safer. These responses support the notion that many brand buyers have imprecise priors regarding the efficacy and/or safety of generic drugs, relative to the brand, especially those who have never before tried the generic. People who choose the generic are less likely to convey uncertainty by choosing "Don't Know" or "Refuse to Answer" when asked how the brand and generic compare in terms of efficacy, safety, or taste (21% vs 31%, p=0.056).

The survey responses also reveal diffuse and imprecise priors regarding the share of consumers who buy the brand or generic: 50% is a modal answer, accounting for 24% of responses, and the remaining responses range from 5 to 95%.[28] As expected, we see that consumers who choose the brand, on average, believe that fewer consumers choose the generic (49%) than consumers who choose the generic themselves (58%, p<.001). Interestingly, the guesses of those who choose the brand themselves are closer, on average, to the true proportions. That is, we find no evidence that the beliefs of brand-buying consumers are less accurate than those of generic-buying consumers.

To assess how the predisposition for "observational learning" might differ between generic and brand buyers, we analyzed responses about how likely one would be to change their choice (i.e. consider switching from the brand to the generic, or consider switching from the generic to the brand) if they learned that the share of customers making the same choice as them was smaller than what they guessed. Of consumers who had chosen the brand version, 19% said they would "probably" or "definitely" buy the generic if this was the case. By contrast, only 8.7% of generic buyers said would "probably" or "definitely" buy the brand if they learned that buying the generic was less common than they thought (p<.01).[29] We interpret this as suggestive evidence in favor of

[28] Although the survey did not allow for non-response on this question, survey enumerators noted that many respondents wanted to skip this question, stating "I have no idea." Such respondents, when pressed to make a guess, would typically answer "half."

[29] In both cases, X% was engineered to be a smaller percentage than what they gave as their prior, indicating that more people than they expect choose the opposite product. Details of this process are in the Appendix.

our hypothesis that brand-buying customers are more likely to be swayed towards the generic by information on the purchases of other customers, than generic-buying customers are to be swayed towards the brand. It is also consistent with over-the-counter drugs being an experience good; those who have already tried a product are far less uncertain about its therapeutic value.

5. Conclusion

Unlike prescription drugs, OTC drugs are purchased by consumers with free choice among available competing products and direct access to price information and standardized "drug facts." Nevertheless, a low responsiveness to price differences between near substitutes may result from biased beliefs about the differences between brand and generic drugs. We implemented a labeling experiment at six locations of a national retailer to test three hypotheses for consumer aversion to generic OTC drugs: (1) lack of information regarding their similarity to the brand, (2) inattention to the price difference, and (3) biased beliefs regarding generic quality, that can be updated with information on their peers' purchases.

We found no evidence for the first hypothesis and weak evidence for the second hypothesis. In the stores in which we displayed price differences in percentage terms, we documented a significant increase in generic purchase share. However, this increase did not appear to be driven by changing the behavior of customers, but rather, by attracting different customers.

We found the strongest evidence for the third hypothesis. Given that OTC drugs are consumed privately, individuals cannot easily observe the drugs that their peers purchase. Thus, it is perhaps not surprising that this information impacted behavior. In the two stores in which generic purchase shares of treated products averaged above 50%, labels that provided these shares, by product, led to a further increase of 10 percentage points. We conclude that some consumers do not buy generic products because they are unaware that these products are commonly purchased by shoppers like them, who apparently find them safe and effective.

The customer-level analysis rejects the possibility that our treatment effects are purely driven by new customers who already buy generic drugs elsewhere, or by existing buyers of generic drugs simply buying more of them. Among existing customers who were previously loyal to brands, we find that information on peer purchases of OTC drugs is the only treatment that increases the probability of buying a generic for the first time. This suggests that these customers may be wary of the overall quality of the generic products, but respond to the information that many other customers find them acceptable.

References

Ackerberg, Daniel A., 2001. “Empirically Distinguishing Informative and Prestige Effects of Advertising,” The RAND Journal of Economics, vol. 32: 316- 333.

Ackerberg, Daniel A., 2003. “Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination,” International Economic Review, vol. 44: 1007-1040.

Akerlof, George A, 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," The Quarterly Journal of Economics, MIT Press, vol. 84(3), pages 488-500, August.

A.C. Nielsen. “The power of private label 2005: a review of growth trend around the world” 2005.

Banerjee, A. (1992). A simple model of herd behavior. Quarterly Journal of Economics, 107(3), 797-817.

Becker, G.S. (1991). A note on restaurant pricing and other examples of social influences on price. Journal of Political Economy, 99(5), 1109-1116.

Bikhchandani, S, Hirshleifer, D. and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. Journal of Political Economy, 100(5), 992-1026.

Bronnenberg, Bart, Jean-Pierre Dubé, Matthew Gentzkow and Jesse Shapiro (2015) “Do pharmacists buy Bayer? Informed shoppers and the brand premium” Quarterly Journal of Economics.

Bursztyn, Leonardo, Ederer, Florian P., Ferman, Bruno and Yuchtman, Noam (2013) “Understanding Peer Effects in Financial Decisions: Evidence from a Field Experiment.” Available at SSRN: http://ssrn.com/abstract=2101391.

Cai, Hongbin, Yuyu Chen, and Hanming Fang. 2009. "Observational Learning: Evidence from a Randomized Natural Field Experiment." American Economic Review, 99(3): 864-82.

Carrera, Mariana, Dana Goldman, Geoffrey Joyce, and Neeraj Sood 2018. "Do Physicians Respond to the Costs and Cost-Sensitivity of their Patients?" *American Economic Journal- Economic Policy,* 10 (1), 113-152.

Chen, S S, Monroe, K B and Lou, Y (1998). The effects of framing price promotion messages on consumers' perceptions and purchase intentions. Journal of Retailing, 74(3), 353-372.

Chetty, Raj, Looney, Adam, and Kory Kroft, 2009. “Salience and Taxation: Theory and Evidence” American Economic Review 99(4): 1145-1177.

Eliashberg, Jehoshua, and Steven M. Shugan, 1997. “Film Critics: Influencers or Predictors?,” Journal of Marketing, vol. 61: 68-78.

Gendall, P., Hock, J., Pope, T., & Young, K. (2006).Message framing effects on price discounting.Journal of Product & Brand Management,15(7), 458-465.

Hibbard, Judith and Jessica Greene, 2013, “What The Evidence Shows About Patient Activation: Better Health Outcomes And Care Experiences; Fewer Data On Costs” Health Affairs, vol. 32:207-214

Hilger, James, Greg Rafert, Sofia B. Villas-Boas, 2011“Expert Opinion and the Demand for Experience Goods: An Experimental Approach,” Review of Economics and Statistics, 93 (4): 1289-1296

Hossain, T. and Morgan, J., 2006: “…Plus Shipping and Handling : Revenue (Non) Equivalence in Field Experiments on eBay,” B.E. Journals in Economic Analysis and Policy: Advances in Economic Analysis and Policy, 6, 2, 1-27.

Jin, Ginger Z., and Phillip Leslie, 2003 “The Effects of Information on Product Quality: Evidence from Restaurant Hygiene Cards,” Quarterly Journal of Economics, vol. 118: 409-451.

Kesselheim AS, Misono AS, Lee JL, et al. Clinical Equivalence of Generic and Brand-Name Drugs Used in Cardiovascular Disease: A Systematic Review and Meta-analysis. JAMA. 2008;300(21):2514-2526. doi:10.1001/jama.2008.758.

Khan, Romana, Kanishka Misra, and Vishal Singh. "Ideology and Brand Consumption." Psychological science (2013).

Lacetera, Nicola, Devin Pope, and Justin Sydnor, "Heuristic Thinking and Limited Attention in the Car Market" American Economic Review (2012) 102(5), 2206-2236.

Lee, Y.H. and Malmendier, U., 2007: "The Bidder's Curse," NBER Working Paper 13699.

Ma, Juan, Zhaoning Wang and Tarun Khanna, "Why Advertising Safety Isn't Safe? Reminder Effect and Consumers' Negative Response to Information about Product Quality" (January 26, 2017).
INSEAD Working Paper No. 2017/08/STR/EMI. Available at SSRN: https://ssrn.com/abstract=2906179

McFadden, D.L. and Train, K. (1996). Learning from self and others. Journal of Political Economy, 104(4), 684-703.

Montgomery, Cynthia A & Wernerfelt, Birger, 1992. "Risk Reduction and Umbrella Branding," The Journal of Business, University of Chicago Press, vol. 65(1), pages 31-50, January.

Moretti, E. (2011). Social learning and peer effects in consumption: evidence from movie sales. The Review of Economic Studies, 78(1), 356-393.

Nease, Robert, Sharon Frazee, Larry Zarin, and Steven B. Miller (2013), "Choice Architecture is a Better Strategy than Engaging Patients to Spur Behavior Change," Health Affairs, vol. 32: 242-249.

Nelson, Phillip, 1970. "Information and Consumer Behavior," Journal of Political Economy, University of Chicago Press, vol. 78(2), pages 311-29, March-Apr.

Reinstein, David A., and Christopher M. Snyder, 2005. "The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics," Journal of Industrial Economics, vol. 53: 27-51.

Roehr B. Potential benefits of "well engaged patients" are akin to those from a blockbuster drug, say experts. BMJ 2013;346:f886

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

Saxell, Tanja. "Private Experience and Observational Learning in Pharmaceutical Demand." Working Paper https://sites.google.com/site/tsaxell/ 2014.

Shrank, William H., et al. "Physician perceptions about generic drugs." The Annals of pharmacotherapy 45.1 (2011): 31-38.

Shrank, William H., et al. "Patients' perceptions of generic medications." Health Affairs 28.2 (2009): 546-556.

Train K.E. Discrete Choice Methods with Simulation. Cambridge: University Press, 2002.

Wiggins, Stephen. and W.J. Lane, 1983. "Quality Uncertainty, Search, and Advertising," American Economic Review, vol. 73: 881-894.

Wolinsky, Asher, 1995 "Competition in Markets for Credence Goods," Journal of Institutional and Theoretical Economics, vol. 151, 117-131.

Zhang, Juanjuan, 2010 "The Sound of Silence: Observational Learning in the U.S. Kidney Market," Marketing Science, vol 29, 315-335.

Table 1. Descriptive Statistics

Panel A: Treated Products

	Control stores	Test 1 stores	Test 2 stores	Test 3 stores	All
Brand price per unit	0.41	0.41	0.42	0.41	0.41
	0.26	0.27	0.29	0.27	0.27
Generic price, as a share of brand price	0.60	0.58	0.60	0.58	0.59
	0.29	0.16	0.22	0.20	0.25
Weekly quantity sold per product	11.30	13.48	8.37	8.36	10.26
	12.51	10.96	8.68	7.43	10.83
Generic share	0.49	0.51	0.41	0.46	0.47
	0.30	0.26	0.31	0.30	0.30

Panel B: Untreated Products

	Control stores	Test 1 stores	Test 2 stores	Test 3 stores	All
Brand price per unit[a]	0.42	0.43	0.40	0.42	0.42
	0.29	0.30	0.29	0.32	0.30
Generic price, as a share of brand price[a]	0.52	0.51	0.52	0.55	0.53
	0.21	0.21	0.22	0.23	0.22
Weekly quantity sold per product	9.60	13.85	8.28	7.98	9.33
	13.32	18.88	11.64	13.42	13.71
Generic share	0.43	0.44	0.39	0.46	0.43
	0.32	0.26	0.32	0.33	0.32

[a]Liquid and nasal products are excluded from the price calculations since units could not be determined.
Notes: The averages represent pre-treatment weeks 2012wk16-2012wk19. "Weekly quantity" is the number of packages sold (which may vary in dose counts, brand, flavor, and pill type). Prices are in dollars per unit (count), inclusive of discounts, averaged over the different UPCs sold for each active ingredient and then averaged across the different products in the treated and untreated groups. "Generic price as share of brand price" is the per-unit price of generic formulations divided by the per-unit price of brand formulations. "Generic share" is the number of generic packages of each product divided by the total number of packages sold for each product, averaged over the pre-treatment weeks.

Table 2: Difference-in-Differences in Generic Share, Pooled Treatments

Panel A: Treated Products

	Control Stores	Treatment Stores		Differences
Pre-treatment means	0.488	0.455		-0.033
s.e.	0.011	0.022		0.023
N	485	484		969
Treated Time means	0.476	0.501		0.025
s.e.	0.015	0.012		0.016
N	319	322		641
Difference over time	-0.012	0.046*	$DD_{t.p.}$ =	0.058*
s.e.	0.018	0.024		0.027
N	804	806		1610

Panel B: Untreated Products

	Control Stores	Treatment Stores		Differences
Pre-treatment	0.426	0.433		0.007
s.e.	0.004	0.014		0.017
N	592	576		1168
Treated Time means	0.426	0.419		-0.008
s.e.	0.016	0.007		0.022
N	390	374		764
Difference over time	0.001	-0.014	$DD_{u.p.}$ =	-0.015
s.e.	0.015	0.015		0.026
N	982	950		1932
			DDD =	0.073*
				0.039
				3542

Notes: Stores treated with Test 1, Test 2, and Test 3 are all included as the pooled treatment stores. Six treated stores and six control stores are included. Pre-treatment time is weeks 14-19 in 2012; treated time is weeks 20-23 in 2012. Standard errors are clustered by week.

Table 3: Treatment Effects of Each Labeling Intervention, Store-level data

	Generic share				Quantity			
	Treated Products		Untreated Products		Treated Products		Untreated Products	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Test 1: Comparability Statement	-0.037 (0.032)	0.0067 (0.026)	-0.017 (0.045)	0.0038 (0.024)	1.22 (0.77)	0.91 (0.74)	0.092 (1.54)	0.17 (1.53)
Test 2: Price comparison	0.053 (0.038)	0.055 (0.033)	-0.037 (0.044)	-0.0060 (0.021)	1.29** (0.55)	1.05** (0.47)	0.29 (0.41)	0.23 (0.41)
Test 3: Observational Learning	0.073** (0.028)	0.068** (0.027)	0.000076 (0.033)	0.0071 (0.020)	1.21 (0.80)	1.24 (0.81)	0.39 (0.32)	0.37 (0.29)
Generic on promotion		0.11*** (0.023)		0.085*** (0.026)		0.86** (0.38)		1.01** (0.37)
Brand on promotion		-0.13*** (0.041)		-0.15*** (0.039)		1.64*** (0.41)		0.73 (0.55)
Both on promotion		-0.030 (0.032)		0.013 (0.049)		0.68 (0.56)		2.60** (1.05)
Weighted by quantity	No	Yes	No	Yes	--	--	--	--
N	1610	1610	1932	1932	1680	1680	2040	2040
Mean dependent variable	0.47	0.47	0.43	0.43	10.3	10.3	9.33	9.33

Notes: Observations are at the week-store-drug level. Quantity is total products purchased of both brand and generic versions. Generic share is the share of generic purchases within that quantity. Test 1 was conducted at one store, test 2 was conducted at two stores, and test 3 was conducted at three stores. All tests were conducted in the same six week period. In six control stores, no labeling tests were conducted. "Generic on promotion" and "Brand on promotion" are dummy variables indicating that any of the applicable products are on a price promotion during that week in that store. Store, product, and week fixed effects are included in all specifications. Standard errors clustered at the product level are in parentheses; * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 4: Treatment Effects by Framing of Information

	Generic share				Quantity			
	Treated Products		Untreated Products		Treated Products		Untreated Products	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Test 1: Comparability Statement	-0.037	0.007	-0.017	0.004	1.22	0.91	0.092	0.17
	(0.032)	(0.026)	(0.045)	(0.024)	(0.77)	(0.74)	(1.54)	(1.53)
Test 2: Price Comparison								
Framing: Save X%	0.114	0.075*	-0.073	-0.028	1.71**	1.54*	-0.080	0.007
	(0.074)	(0.040)	(0.063)	(0.036)	(0.77)	(0.75)	(0.65)	(0.56)
Framing: Pay Y% More	-0.011	0.019	0.006	0.038	0.87	0.55	0.67*	0.45
	(0.049)	(0.049)	(0.041)	(0.024)	(0.63)	(0.62)	(0.37)	(0.43)
Test 3: Observational Learning								
Framing: X% choose generic	0.069**	0.064*	-0.002	0.019	1.41	1.41	0.31	0.29
	(0.032)	(0.031)	(0.037)	(0.020)	(0.83)	(0.84)	(0.39)	(0.37)
Framing: Y% choose brand	0.083	0.081**	0.005	-0.029	0.82	0.91	0.54	0.53
	(0.052)	(0.032)	(0.047)	(0.032)	(0.89)	(0.90)	(0.42)	(0.45)
Generic on promotion		0.110***		0.085***		0.86**		1.01**
		(0.023)		(0.026)		(0.39)		(0.37)
Brand on promotion		-0.135***		-0.152***		1.63***		0.73
		(0.041)		(0.039)		(0.41)		(0.55)
Both on promotion		-0.030		0.014		0.69		2.59**
		(0.032)		(0.050)		(0.56)		(1.05)
Weighted by quantity	No	Yes	No	Yes	--	--	--	--
N	1610	1610	1932	1932	1680	1680	2040	2040
Mean dependent variable	0.47	0.47	0.43	0.43	10.3	10.3	9.33	9.33

Notes: Observations are at the week-store-drug level. Quantity is total products purchased of both brand and generic versions. Generic share is the share of generic purchases within that quantity. The framing of the information presented in Tests 2 and 3 was varied at the store level. Each framing variation was tested at one store, with the exception of the first framing of Test 3 ("X% choose generic") which was tested at two stores. "Generic on promotion" and "Brand on promotion" are dummy variables indicating that any of the applicable products are on a price promotion during that week in that store. Standard errors clustered at the product level are in parentheses; * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 5: Treatment Effect, Heterogeneity at the customer level

	Treated Products			Untreated Products		
Y = Generic purchased	(1)	(2)	(3)	(4)	(5)	(6)
Test 1: Comparability Statement	-0.028	-0.018	-0.015	0.0050	-0.0074	-0.023
	(0.033)	(0.026)	(0.036)	(0.032)	(0.029)	(0.035)
Test 2: Price comparison	0.044*	0.044	0.039	0.0013	-0.013	-0.039
	(0.020)	(0.028)	(0.031)	(0.033)	(0.025)	(0.028)
Test 3: Observational Learning	0.068**	0.065**	0.082**	0.0037	0.016	0.0013
	(0.028)	(0.023)	(0.028)	(0.027)	(0.028)	(0.036)
Test 1 x Previous Use of Generic			-0.010			0.052
			(0.090)			(0.065)
Test 2 x Previous Use of Generic			0.023			0.094
			(0.084)			(0.057)
Test 3 x Previous Use of Generic			-0.053			0.044
			(0.040)			(0.042)
Previous Use of Generic		0.45***	0.48***		0.41***	0.43***
		(0.021)	(0.025)		(0.034)	(0.038)
Generic share of all OTC drugs previously purchased		0.32***	0.32***		0.30***	0.30***
		(0.0092)	(0.0090)		(0.018)	(0.019)
Previous use of purchased product (brand or generic version)		-0.21***	-0.21***		-0.19***	-0.19***
		(0.025)	(0.025)		(0.032)	(0.032)
N	8847	8847	8847	9398	9398	9398
R-sq	0.115	0.373	0.373	0.117	0.366	0.366
Dependent variable mean	0.45	0.45	0.45	0.46	0.46	0.46
No Previous Use of Generic			0.32			0.33
Previous Use of Generic			0.82			0.76

Notes: Linear probability models for the choice of a generic. Observations represent each individual purchase of a treated or untreated drug in the pre-treatment or treatment period. "Test 1," "Test 2," and "Test 3" treatment indicators are interactions for treated store and treatment time period. "Previous Use of Generic" is an indicator for the individual having made any purchase of a generic OTC drug. Models 3 and 6 include interaction terms between treatment time and previous use of generic, and between previous use of generic and indicators for stores receiving test 1, test 2, and test 3. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 6: Post-treatment effects on choice of generic

Y = Generic purchased	Treated Products within Treated Stores						Untreated Products within Treated Stores					
	Weeks 24-27		Weeks 28-31		Weeks 32-36		Weeks 24-27		Weeks 28-31		Weeks 32-36	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Test 1 x Post x Exposed		0.051		-0.016		0.032		-0.044		-0.023		-0.049
		(0.073)		(0.058)		(0.055)		(0.038)		(0.041)		(0.084)
Test 2 x Post x Exposed		-0.050		0.0088		0.019		-0.094**		-0.046		-0.075
		(0.062)		(0.056)		(0.049)		(0.043)		(0.050)		(0.063)
Test 3 x Post x Exposed		0.078*		0.058		0.020		-0.077*		-0.028		-0.0041
		(0.040)		(0.043)		(0.037)		(0.041)		(0.027)		(0.032)
Test 1 x Post	0.012	-0.0042	0.047	0.053	0.015	0.0048	-0.025	-0.0096	-0.0096	-0.0015	-0.033	-0.017
	(0.035)	(0.042)	(0.035)	(0.039)	(0.031)	(0.029)	(0.037)	(0.030)	(0.033)	(0.037)	(0.042)	(0.031)
Test 2 x Post	-0.0063	0.0093	0.029	0.026	0.035	0.029	0.0091	0.039	-0.012	0.0031	-0.018	0.0051
	(0.034)	(0.035)	(0.032)	(0.037)	(0.035)	(0.031)	(0.027)	(0.029)	(0.029)	(0.036)	(0.033)	(0.041)
Test 3 x Post	0.027	0.0028	0.045	0.026	0.037	0.031	0.022	0.046	0.022	0.032	-0.00095	0.00042
	(0.030)	(0.039)	(0.030)	(0.035)	(0.027)	(0.029)	(0.025)	(0.030)	(0.021)	(0.020)	(0.026)	(0.024)
Test 1 stores x Exposed		-0.020		-0.044		-0.016		-0.045		-0.013		-0.029
		(0.047)		(0.026)		(0.047)		(0.027)		(0.048)		(0.033)
Test 2 stores x Exposed		-0.0065		0.0072		-0.0081		0.012		-0.0083		0.014
		(0.070)		(0.032)		(0.069)		(0.033)		(0.069)		(0.036)
Test 3 stores x Exposed		-0.020		-0.0070		-0.017		-0.0053		-0.017		-0.0035
		(0.040)		(0.036)		(0.040)		(0.034)		(0.038)		(0.035)
Post x Exposed		0.0018		-0.066***		0.026		-0.020		0.035*		-0.046***
		(0.024)		(0.018)		(0.023)		(0.013)		(0.018)		(0.015)
Exposed		-0.0086		0.0037		-0.0051		0.0049		-0.0085		0.0086
		(0.017)		(0.014)		(0.016)		(0.013)		(0.016)		(0.016)
N	7857	7857	8819	8819	7628	7628	8789	8789	9449	9449	11459	11459
R-sq	0.358	0.359	0.332	0.334	0.355	0.356	0.348	0.349	0.364	0.364	0.342	0.343
cdepmean	0.45	0.45	0.46	0.46	0.45	0.45	0.46	0.46	0.45	0.45	0.46	0.46

Notes: Linear probability models for the choice of a generic. Observations represent each individual purchase of a treated or untreated drug during the specified post-period combined with the pre-treatment period. For each test, "Test X x Post" is an interaction for a store treated with labeling test X and the specified post-treatment time period. "Exposed" is an indicator for the individual having made any purchase of a (treated or untreated) OTC drug during the treatment time period, indicating their presence in the OTC aisles of the store. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.