

UC Berkeley

CUDARE Working Papers

Title

Generic aversion and observational learning in the over-the-counter drug market

Permalink

<https://escholarship.org/uc/item/0q03b5f2>

Authors

Villas-Boas, Sofia B
Carrera, Mariana

Publication Date

2016-10-28

Peer reviewed

Generic Aversion and Observational Learning in the Over-the-Counter Drug Market¹

Mariana Carrera² Sofia Villas-Boas³

Abstract

Through a labeling intervention at a national retailer, we test 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 their price differences, and uncertainty about generic quality that can be reduced with information on peer purchase rates. With a difference-in-differences strategy, we find that posted information on the purchases of other customers increases generic purchase shares significantly, while other treatments have mixed results. Consumers without prior generic purchases appear particularly responsive to this information. These findings have policy implications for promoting evidence-based, cost-effective choices.

¹This version: March 31, 2022. We thank David Clingingsmith, Stefano DellaVigna, Silke Forbes, Ulrike Malmendier, Owen Ozier, Silvia Prina, Gautam Rao, Mark Votruba, and seminar participants at U.C. Berkeley, Case Western, Kent State, Vanderbilt, Duke University, and Universidade Catolica Portuguesa for helpful comments. We thank a national retailer, whom we cannot name, for sharing data and allowing us to conduct a field experiment. We thank the Giannini Foundation for support.

²Department of Agricultural Economics and Economics (DAEE), Montana State University | P.O. Box 172920 | Bozeman, MT 59717-292, NBER. mariana.carrera@montana.edu

³Department of Agricultural and Resource Economics, University of California, Berkeley, Berkeley, CA 94720-3310. 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). Beliefs about efficacy and risks are acutely important when consumers choose experience goods or credence goods, which encompass most health care treatments. Mistaken beliefs can lead to treatment overuse (e.g., antibiotics) or underutilization (e.g., chronic preventative drugs) (Baicker, Mullanaithan and Schwarstein, 2015). The difficulty of making price comparisons in the health care sector may also drive the overuse of costly treatments in place of lower-cost substitutes (Carrera et al., 2018).

Despite mounting evidence of such “behavioral hazards” affecting consumers' medical decisions, little is known about how to address them. Provision of information has been shown to increase the take-up of some preventative health products (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). A better understanding of how health product preferences are formed and updated is needed to facilitate evidence-based and cost-effective choices.

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 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 brand or 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 purchase rates. For two of the tests, we designed labels with two different ways of framing the relevant information in different stores.

We randomly assigned 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 6 percentage points, or 11% relative to the pre-treatment level. This effect is more than two-thirds as large as the increase associated with a price promotion on the generic store brand, and is particularly strong among brand-loyal customers.

Our first two tests contribute to two literatures analyzing how product quality information affects consumer choices (Montgomery and Wernerfelt, 1992; Jin and Leslie, 2003; Akerberg, 2001; Akerberg, 2003) and how consumers can be inattentive to non-salient components of costs (Chetty, Looney and Kroft, 2009; Hossain and Morgan, 2006), respectively. 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 and the ways in which an individual's demand for a good can be affected by disclosure of other consumers' purchases.⁴ While other experimental studies have found positive effects of peer usage disclosure, it is often difficult to disentangle whether these effects are driven by updated quality priors or by social channels, such as status or network effects (see for example, Cai et al. (2009) in restaurants; Salganik et al. (2006) in an artificial music market).⁵ Similarly, sales rank information and volume of reviews have been shown to influence online sales of experience goods such as books (Chevalier and Mayzlin, 2006), computers (Lu et al., 2021), and video games (Cui et al., 2012).⁶

Since the consumption of pharmaceutical products is a largely private behavior, social status and network effects are arguably absent here. Also due to the private nature of drug consumption, people are unlikely to have strong priors on their peers' choices, making this a ripe setting for observational learning. However, it is not clear *a priori* whether disclosing market shares will increase the desirability of generics, given that they are typically the *less* popular choice (40-

⁴On the theoretical side, Becker (1991) and McFadden and Train (1996) formalize consumer learning about a new good's quality through their own experience and those of their peers.

⁵ 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.

⁶ Historical market shares, our focus in this paper, may provide a more coarse type of information than other customers' ratings and reviews. On the other hand, they also capture the universe of a store's customers, whereas ratings and reviews may only represent a selected subset.

50% market share on average, in the stores studied). We use in-store customer surveys to solicit priors on the purchase shares of other consumers and assess subjective expectations of influence by peer market shares. Findings suggest that peer-purchase information is more likely to sway brand-buying customers towards the generic than to sway generic-buying customers towards the brand, likely because generic-buyers are more likely to have already tried both versions.

The rest of the paper proceeds as follows. Section 2 provides background on over the counter drug regulations and a simple conceptual framework that embeds our three testable hypotheses. Section 3 describes the retail empirical setting, the experimental design, and data. Section 4 presents the empirical strategy, results and robustness checks, and 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).⁷ 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_{ibx} (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 know their private valuation v_{ibx} .⁸ Those who have not tried the generic version do not know their private valuation v_{igx} , but have an expectation over it, $E[v_{igx}]$. We do not assume this expectation is unbiased, since lack of information, prior experience in foreign markets with unregulated or unreliable generics, and differential advertising could create bias. 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, respectively⁹.

⁷ 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.

⁸ 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).

⁹ 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.

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).¹⁰ 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.

To illustrate why customers' differing priors are important, consider a hypothetical case in which the generic purchase share is 50% and, consistent with this share, the customers who see the labels are an even split of would-be generic buyers and would-be brand buyers. We would only expect observational learning to influence the decision of customers who are surprised to learn that their prospective choice is *less* popular than they thought. If would-be brand buyers and

¹⁰ 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).

would-be generic buyers all underestimate the popularity of generics, then observational learning implies an unambiguously positive effect of labels on generic purchase share. But if, instead, all customers overestimate the share of customers who make the same choice as them, then observational learning could make would-be generic buyers less likely to choose the generic just as it makes would-be brand buyers less likely to choose the brand. If generic and brand buyers are symmetrically biased in their priors, and equally responsive to the true peer purchase share, then the labels might not shift the aggregate generic purchase rate at all.

It is also possible that, regardless of their (possibly biased) priors on peer purchase shares, generic and brand buyers put different weights on this peer purchase shares when updating their expectations of the generic's private utility to them ($E[v_{igx}]$) relative to the brand (v_{ibx}). If they are pure experience goods with fixed private utilities, the purchases of other people should be irrelevant to the next choice of all consumers who have previously used both the brand and generic versions of a drug, and thus, know their private values of v_{igx} and v_{ibx} . Of course, if the drugs are not perfect experience goods, consumers 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 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. But we hypothesize it is *less* sensitive to such information after personal experience.

To summarize, if generic shares appear to increase when peer purchase shares are displayed, this could be because (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 4.5, we discuss customer surveys we conducted to explore these distinct possibilities.

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 and that pharmacists' typical responses were similar. 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 permission to post labels beneath the price tags of generic versions of products in randomly selected drug classes (see Figure 1 below).¹¹



Figure 1. Example of Label Placement

¹¹ Labels were reposted each week after price tags were updated by the retailer. We thank Yann Pannasie, Raymond Gong, Lynn Anderson, Karen Yao, Caitlin Crooks, Feyisola Shadiya, Brian Mitchell, Brian Gallo, Roni Hilel, Kyle Kennelly, Jonathan Arenas, Kathy Hua, and 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.

3.2 Experimental Intervention: Three labeling tests

For four-weeks, labels were posted beneath the price tags of generic products in treated drug classes in treated stores. The content of the labels differed across three experimental arms.

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 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].”*¹²

¹² 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 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.

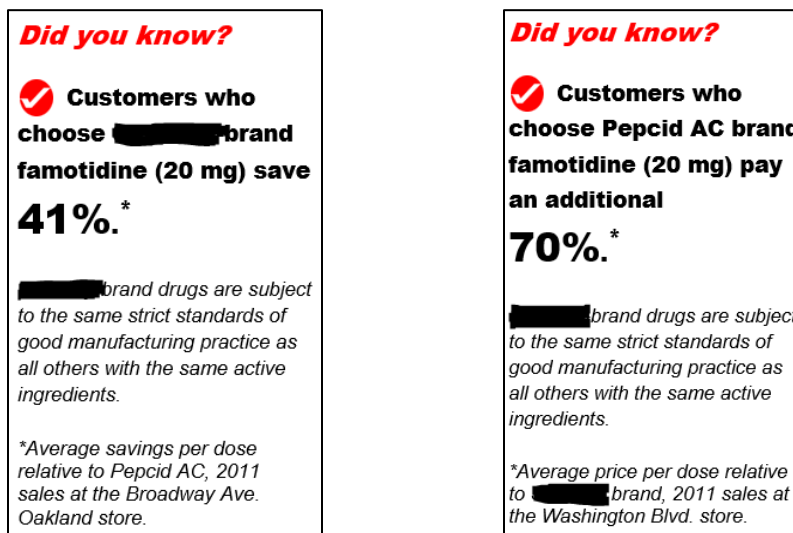


Figure 3. Labels highlighting the price difference.

[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 on consumer purchases.

In sum, the content of the labels differed across three tests corresponding to our three hypotheses, and further across two label versions for tests 2 and 3: label examples are numbered below as tests 1, 2a, 2b, 3a, and 3b. Test 2b (3b) differs from test 2a (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. We acknowledge that since all tests shared the salient new feature of a hanging label put under the product’s price tag, we cannot adjust for the potential effect of that feature alone, separate from the label content.

3.3 Data and Summary Statistics

We use two types of data: first, a panel data set of store-level weekly sales for OTC drugs at the UPC level for all 56 stores in the geographic division of our treated stores, where a week starts on a Wednesday and ends on a Tuesday at store closing. 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.

Second, we use a transaction-level dataset with household identifiers, based on 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 22 (approximately one year preceding our treatment period) to 2012 week 38 (15 weeks after our treatment ended). These data allow us to investigate how customers who shopped during the treatment period changed their behavior relative to their prior purchase patterns, whether different types of shoppers responded differently to the experimental treatments, and whether any changes persisted beyond the intervention.

OTC Product Classes

Our analysis includes twelve of the largest OTC drug classes that offer generic (store-label) versions as well as national brands, broadly grouped into the categories of pain relief, allergy/cold symptoms relief, and digestive/stomach relief. A drug class may include competing products that work in a similar way (e.g. non-sedating antihistamines) or one active ingredient alone (e.g. acetaminophen/Tylenol). We excluded children's and infants' drugs to focus on products a shopper was likely choosing for own consumption. Since our goal was to study brand-generic substitution rather than therapeutic (across product) substitution, we further combined classes that are commonly interchanged (for example, ibuprofen, acetaminophen, and naproxen were grouped as "non-aspirin pain relief," and proton pump inhibitors and H2 acid blockers were grouped as "acid reflux relief"). We then randomly chose four of the following eight class groups (hereafter referred to as "drug classes") to be treated: oral allergy, nasal allergy, acid reflux relief, laxatives, non-aspirin non-nighttime pain relief, aspirin pain relief, nighttime pain relief, and cold/ sinus/flu products, stratified by symptom category. This experimental design is visually summarized in Appendix Figures A1 and A2, and the summary statistics for the treated and untreated products across the three

symptom categories of pain relief, allergy/cold symptoms relief, and digestive/stomach relief are shown 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.¹³ Prices may be adjusted by the retailer weekly.¹⁴ 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.¹⁵

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 combine all of these UPCs into one observation, but another 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 (i.e. packages) purchased, within this set, 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

¹³ 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.

¹⁴ These adjustments are done at the same time in all stores, between Tuesday's closing and Wednesday's opening. Thus, we define weeks in our data as beginning on Wednesday and ending on Tuesday.

¹⁵ A loyalty card for this retailer is required to obtain the promotional price. Since our household level dataset 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%.

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 available at a store 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, for both treated products and untreated products. As we would expect, there are no significant differences 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, for both treated and untreated products. The generic shares of both treated and untreated products purchased are smaller in the Test 2 stores than in the control stores and this difference is marginally significant for untreated products ($p=0.09$); see Appendix Table A2 for clustered standard errors). Apart from this, and the sales quantity differences driven by store size, the differences between stores receiving treatments and control stores are not statistically significant in the pre-treatment period.

Household-level data

Longitudinal purchases at the household level are available through purchases made via loyalty cards. At this retailer, discounts posted throughout the store are only available via loyalty card and their use is frequent, 89% of OTC drug purchases. In our household-level dataset, we include households who made a purchase of any OTC drug in our sample during the pre-treatment, treatment, or post-treatment period, and who also have prior purchases (linked by loyalty card) of any OTC drug in our sample from week 22 of 2011 to week 12 of 2012. Note that the prior purchases are not required to be of the same type of drug they are purchasing during the treatment or pre-treatment period, to avoid further reducing the size of this sample. We create the following control variables: total number of OTC purchases during the period from 2011 week 22 to 2012 week 12, the share

of those purchases that were for a generic version of a product, an indicator for having previously purchased product j (brand or generic version), and an indicator equal to one if the previous purchase of product j was for the generic version, where j is the product they are observed to purchase at present. We also calculate the household's average percentage discount on total purchases at the store (including non-drug purchases), as a proxy for price sensitivity.

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 $Q = (q_{gen} + q_{brand})$.¹⁶ Using a differences-in-differences approach, we estimate 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 that compares 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. 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.

¹⁶ Note that we focus on the number of packages purchased as *quantity*, rather than the number of daily doses.

The top panel, corresponding to treated products, shows that in the pre-treatment period, mean generic shares were 46.5% in the control stores and 42.7% in the treatment stores, an insignificant difference. From the pre-treatment period to the treatment period, average generic shares of treated products increased by 4.7 percentage points in the treatment stores and decreased by 1 percentage point 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 of 5.6 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 -0.009 in control stores and -0.020 in treatment stores, leading to an insignificant difference-in-differences ($DD_{u,p}$) estimate of -1.1 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 an increase of 0.068 in the generic share of purchases and marginally statistically significant ($p=0.10$). 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.¹⁷

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

¹⁷ 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 laudable attributes than the unlabeled products.

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 generic share (gs_{ist}) or quantity (Q_{ist}) of product i purchases 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, relative to 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 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 pre-labeling period 2012wk14-2012wk19. In all specifications, standard errors are clustered at the drug class-by-store level, following Abadie et al. (2017) in defining clusters consistent with the level at which treatments were assigned.¹⁸

Table 3 reports the results of these regressions. Only Test 3 had a positive and statistically significant effect on the generic share of treated products, a 6 percentage point increase in generic share in the preferred specification of column 2. The estimated effect of Test 1, by contrast, appears close to zero.¹⁹ The estimated effect of Test 2 on generic share is somewhat smaller than Test 3's effect and statistically insignificant, but cannot be statistically distinguished from either Test

¹⁸ Our results are not sensitive to this clustering approach. In a previous version of this paper, we clustered at the product level and found similar results.

¹⁹ It is possible that the presence of a hanging tag itself negatively impacted the sale of products, perhaps counteracting a positive effect of the information stated on the tag. Since we did not test any labels without information, we cannot rule out this possibility.

3's positive effect or Test 1's null effect in either specification. For untreated products, Column 4 shows that the estimated effects of all three tests are close to zero, suggesting that the labels did not influence the purchasing of products that did not receive labels.²⁰ Columns 5-8 show the same specifications estimated for the outcome of total item quantity for treated and untreated products. The reason we test for effects on quantity is that labels might draw the attention of consumers to the treated products, possibly leading more consumers to purchase them. Results show positive but statistically insignificant estimates for the treated products in all three tests.

Note that in Table 3, we estimate the average effect of Test 1 across three different comparability statements used on the information labels, and the average effects of Test 2 and Test 3 across the two framing variations that were used in different treated stores. In Table 4, we disaggregate these results. While we continue clustering at the drug class-by-store level, the fact that some framing variations were only tested in one store, and that Test 1's variations were each tested on a subset of treated products, means we have a smaller number of treated clusters (at most 4 per store) for estimating the effects of each treatment variation. Since cluster-robust test statistics may over-reject the null in the case of a small number of "effective" clusters, we also report p-values and confidence intervals from a wild cluster bootstrap procedure (Cameron and Miller, 2015; Roodman *et al.*, 2019).²¹

Among the different comparability statements used on information labels, none had a significant effect on purchases. 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 for both generic purchase share and quantity purchased, but the bootstrapped p-values are $p=0.09$ for generic share and $p=0.12$

²⁰As mentioned previously, in theory, labels could have spillover effects on other generic products in the same store.

²¹We use the Stata *boottest* package with 2000 replications applying Webb weights (Roodman *et al.* 2019; Webb, 2014).

for quantity. The other framing method (“pay an additional $y\%$ ” relative to the generic), by contrast, has small and insignificant point estimates for both outcomes, but the imprecision of the estimates does not allow us to reject the null hypothesis that these two framing variations had the same effect.

For Test 3, we varied the framing to test whether statements like “*45% of customers buy the generic*” have a stronger effect on generic share than statements like “*55% of customers buy the brand.*” If they do, then we might infer that part of Test 3’s impact was driven either by pointing out the existence of the generic product, or through an implicit perceived encouragement to buy it. Results in Table 4 show no evidence of this. The point estimate for the second framing, highlighting the brand share, is actually larger than the point estimate for the first framing, but we cannot reject that they have the same effect. This evidence is consistent with the labels working through observation learning: in a binary choice situation, “*45% of customers buy the generic product*” conveys the same information as “*55% 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 exogenous variation in the share displayed. Appendix Figure A3 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, because 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. A linear regression (Appendix Table A3) estimates a coefficient of 0.68, suggesting that holding the product and store constant, a 10 percentage point increase in the

posted generic share is associated with an additional 6.8 percentage point increase in the predicted generic purchase rate.

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 including the prior weeks for which we have store-level data for all stores. For each of the three labeling interventions ($m = 1, 2, \text{ and } 3$) and for treated and control products separately, we estimate:

$$(3) \quad GenericShare_{ist} = \sum_{t \neq -1} \beta_t D_{sm} + \alpha_i + \alpha_s + \alpha_t + \epsilon_{ist}$$

where D_{sm} is a dummy variable equaling one if store s received labeling intervention m , α_i are product fixed effects and α_s are store fixed effects. Fixed effects for bi-week-of-sample t represent 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, and $t=1$ for the last two weeks of treatment.). The β_t vector contains the coefficients of interest, capturing the difference between the treated stores and the untreated stores during each specific period relative to the excluded biweek of sample, $t=-1$. If the generic shares of products are trending similarly in treated stores and control stores before the treatment periods, there should be no trend in the β_T coefficients during the pre-treatment period and their values should be statistically indistinguishable from zero.

Figure 5 plots the estimates we obtain from equation (3), with the $\hat{\beta}_t$ plotted in black and the 95 percent confidence intervals, based on standard errors clustered at the drug class-by-store level, shown as gray dotted lines. Vertical lines separate the sample into two weeks sub-periods during the pre-treatment and treatment

periods; our store-level dataset does not extend beyond the treatment period. 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 estimates are statistically different from zero at the 95% significance level, consistent with the assumption of pre-existing parallel trends for all treatment stores relative to the controls.²²

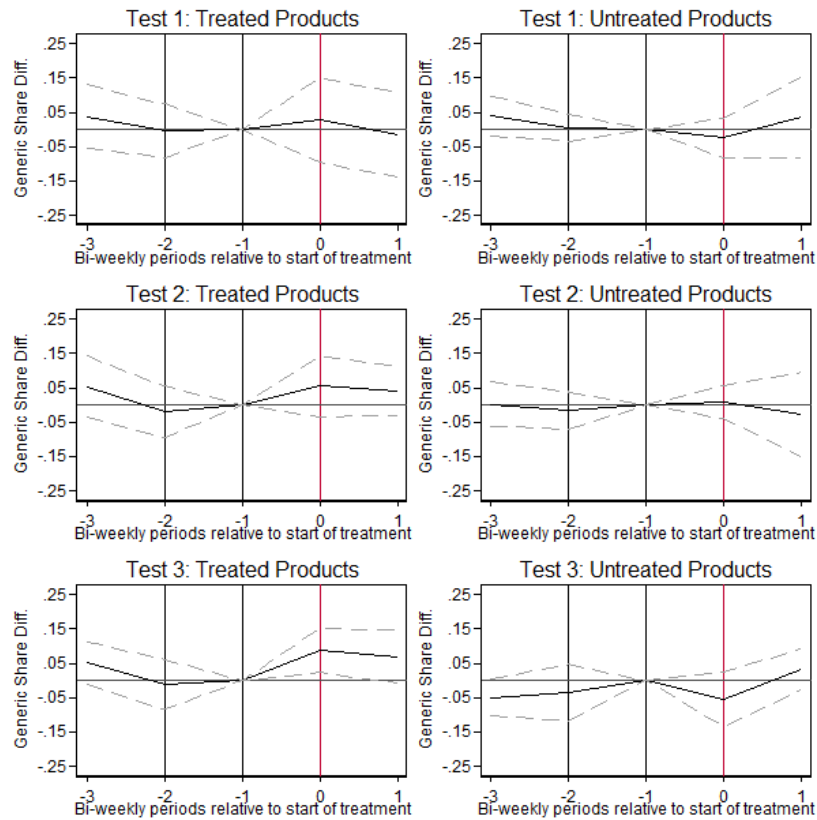


Figure 5. Event-Study Analysis of Generic Share

4.3 Customer-level analysis

²² The increase in the size of the confidence intervals in period +1 is driven by generally smaller purchase quantities in those two weeks due to seasonal trends, making generic shares noisier. As Tests 1 and 2 were conducted in fewer stores than Test 3, their confidence intervals are also generally wider.

An important concern arises when the analysis is restricted to store-level totals: We cannot confirm whether existing buyers of OTC products within a store are shifting their purchases towards generics, or if instead the labels attract the attention of different customers, who may already be buyers of generics at other retailers. Household-level data allows us to control for the past shopping choices of the customers observed during the treatment period. This enables us to rule out that the effects are driven by the composition of shoppers changing, and furthermore to test whether the labels have different effects on consumers who have previously purchased generic versus branded OTC products.

An initial look at this data reveals strong habit persistence in the choice of brand versus generic formulations. For example, of the 13,640 household-drug combinations that we observe with two or more purchases of the same drug during the pre-intervention period, 86% make the same choice on both purchase occasions. Of those who first purchase the brand, 87% buy the brand again in their second purchase, and of those who first purchase the generic, 85% buy the generic again. Appendix Table A4 shows transition probabilities between first and second, as well as second and third purchases.

To estimate treatment effects of the labeling interventions, we focus on the purchases made by individuals whose loyalty cards show a prior purchase of at least one OTC drug from week 22 of 2011 to week 12 of 2012, our past purchases observation period. Each purchase of a treated or untreated product is an observation, and the dependent variable, *generic*, is equal to one if the purchase made was for a generic version.²³ 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. The equation is:

²³ For the cases in which the same household made more than one purchase of the same product in the same week (4%), *Generic* represents the generic share of these purchases and the regressions weight observations by quantity. In only 5% of these cases (less than 0.2% of the sample), the generic share lies between 0 and 1.

$$(4) \text{ Generic}_{ijst} = XB + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + t_t + \delta_j + \delta_s + \varepsilon_{ijst}$$

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 always include dummies for whether brand, generic, or both versions of the product were on sale during a given week in store s , and add household level controls in subsequent columns. 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_j are product fixed effects, δ_s are store fixed effects, and we show both standard errors clustered at the drug class-by-store level and p-values calculated with the wild cluster bootstrap.

Table 5 shows the results of this estimation, which largely matches those of the store-level analysis. In the first column (before adding household-level control variables), the point estimate of the effect of Test 1 is slightly negative and insignificant, the estimate for Test 2 shows a 6.5 percentage point increase in the probability of choosing a generic OTC drug over its brand-name counterpart, and the estimate for Test 3 shows a 7.4 percentage point increase. Both the Test 2 and Test 3 effects are statistically significant (5% and 1% levels, respectively) under both classical and wild cluster bootstrapped p-values. Tests of coefficient equality, under the wild cluster bootstrap, reject that Test 1 and Test 3 have the same effect ($p=0.01$) and marginally reject that Tests 1 and 2 have the same effect ($p=.07$). Column 4 shows that none of the three tests had significant effects on the probability of generic choice for untreated products, with point estimates remarkably close to zero.

Next, in columns 2 and 5, we add household-level controls based on the past purchases observation period: the generic share of OTC purchases, an indicator for any previous purchase of the product now being purchased, an indicator for whether the last previous purchase of the product was for the generic version, and the past average percentage discount obtained on total spending, a proxy for price

sensitivity. These covariates have significant explanatory power, raising the R-squared of the regression from .12 to .39 for both treated and untreated products. If the type of customer making purchases changed between the pre-test period and the labeling test period, differently in treated stores versus control stores, we would expect the estimated treatment effects to shrink between columns 1 and 2. When controlling for past purchasing behavior, the estimated effects of both Test 2 and Test 3 are somewhat smaller, and the Test 2 estimate's level of significance is reduced to $p=0.055$ (classical cluster-robust) or $p=0.08$ (bootstrapped). Test 3's estimated effect remains highly significant.

In columns 3 and 6, we test for heterogeneity of treatment effects based on previous generic purchase. 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 less responsive to the information provided in certain labeling tests. To test this, we add an interaction term between each test and the indicator for previously purchasing the generic version of the same product.²⁴ The baseline effect of Test 3, now representing the effect only on customers who did not purchase the generic version previously, shows a percentage point increase of 8.1 ($p<.001$). Since the baseline probability that a customer of this type will choose the generic is 31 percent, this is a 26% increase in likelihood of choosing the generic. The point estimate of the interaction between Test 3 and the dummy variable for making a prior generic purchase is large and negative. Combined with the fact that customers who previously purchased the generic version of product j chose the generic 86 percent of the time, this large and negative point estimate suggests a much smaller relative increase, 2.4%, in generic purchasing for customers not previously tied to the brand. This coefficient is borderline significant (classical $p=.066$, bootstrapped $p=.078$). For Test 2, the

²⁴ We also add interactions between previous purchase of the generic and the treatment time period, and between previous purchase of the generic and indicators for the Test 1, Test 2, and Test 3 stores.

baseline treatment effect estimate is 0.061 and less significant ($p=.067$), but the interaction term does not suggest a different impact on previous generic users. The results for Test 1 indicate no significant treatment effect for either type of customer, and results in columns 4-6 show no significant effects of any treatment on untreated products.

Appendix Table A5 shows household-level results disaggregated by labeling statement, as done previously using store-level data in Table 4, and also shows results for the set of households with no observed prior purchases of OTC drugs. Disaggregating the effects by the framing of Tests 2 and 3 confirms two findings described earlier, in the store-level analysis: Only the “Save X%” framing variation of Test 2 (Test 2a) significantly increased generic purchase shares, while both ways of framing the peer purchase shares in Test 3 had equivalent effects. Also, Test 3 appears to have equally strong effects on customers who have made prior OTC purchases at this store and those who have not, while Test 2 had a smaller and statistically insignificant effect in the latter group.

Among households with no prior purchases of OTC drugs, point estimates suggest a large positive effect of Test 1, in particular for the labels noting FDA approval of the generic and its confirmed bioequivalence, when applicable. This would suggest that customers who have never purchased an OTC drug at this retailer may be uninformed about the comparability of store-label and national brands (while others who have made prior purchases are not), but the results are not significant under the wild cluster bootstrapped standard errors.

4.4 Post-treatment effects on customers exposed to labels

To explore whether the information provided in the labels leads to persistent changes in purchasing behavior, we turn to purchase data from the period following label removal, which we divide into three consecutive four-week intervals. Our access to household identifiers is crucial for this analysis, because OTC products

are purchased infrequently: only twenty percent of the shoppers in any subsequent four-week period made a purchase of any OTC drug during the treatment period. Thus, it is unsurprising that testing for a different change in generic purchase rates between treatment and control stores, from the pre-treatment period to the post-treatment period, yields no significant effects (Appendix Table A6, odd-numbered columns), because the majority of customers shopping in the post-treatment period may not have seen the labels at all.²⁵

To test for persistent changes among shoppers who were present in the OTC drug aisles during the four-week labeling period, we adopt another difference-in-differences approach. We compare pre- to post-intervention changes in generic purchase rates between customers who were or were not exposed to the labels, using any purchase during the treatment period (of a treated or untreated product) as a proxy for label exposure. We estimate the following model within treated stores alone. For convenience, in this equation we represent the terms that are separately estimated for Test 1, Test 2, and Test 3, as the summation of terms over $k = 1 - 3$.

(5) $Generic_{ijst}$

$$\begin{aligned}
&= XB + \sum_{k=1}^3 \beta_k PostPeriod_t * TS_{ks} * Exposed_i \\
&+ \sum_{k=1}^3 \gamma_k PostPeriod_t * TS_{ks} + \sum_{k=1}^3 \theta_k TS_{ks} Exposed_i + \delta_j + \delta_s \\
&+ \varepsilon_{ijst}
\end{aligned}$$

$PostPeriod$ indicates that week t is part of the post-intervention period, while the intervention period is excluded from the regression. TS_{1s} , TS_{2s} , and TS_{3s} are indicators for whether store s was treated with labels for test 1, 2, or 3, respectively.

²⁵ The estimated equation for the regression results in Table A6 matches equation (4) except PT_1 , PT_2 and PT_3 represent interactions between treated stores, and a post-labeling period replaces the treat-time dummy:

$$Generic_{ijst} = XB + \beta_1 PT_1 + \beta_2 PT_2 + \beta_3 PT_3 + \delta_t + \delta_j + \delta_s + post_t + \varepsilon_{ijst}$$

$Exposed_i$ is equal to one if consumer i was exposed to the labels, *i.e.*, if she purchased any of the products included in the sample during the treatment period, and zero otherwise. We include interactions between *PostPeriod* and TS_k to capture the difference in the purchases of unexposed consumers between the pre-intervention and post-intervention period, and we include interactions between *Exposed* and each set of treated stores ($TS_k * Exposed$) to control for any pre-existing differences in the purchases of exposed and unexposed consumers.

Results for the treated stores are shown in Table 6. Since our number of clusters is further reduced in this analysis, we focus solely on the wild cluster bootstrap p-values. Estimated effects are noisy, but suggest that the Test 3 labels continued to influence the choice of products among customers who had made purchases during the label period. Appendix Table A4 (even numbered columns) shows qualitatively similar results of a triple-difference specification that includes control stores, similarly defining exposed customers in control stores as those who purchased any OTC product during the treatment time period.

Note that this approach assumes the labels did not change the composition of shoppers making an OTC purchase at a given time, *i.e.* being “exposed” to the labels is quasi-random. The fact that we find no significant changes in the quantity of purchases during the labeling period is consistent with the assumption that making a purchase is primarily driven by the need to purchase an OTC drug during the treatment time period. Also, the coefficients of the interaction terms between Test 1, 2, and 3 stores and *Exposed*, shown at the bottom of Table 6, indicate that in the period prior to the labeling intervention, the purchases of soon-to-be-exposed customers did not significantly differ from those of the non-exposed customers.

4.4 Robustness

To assess the robustness of our store-level results, we tested for serial correlation in store-product observations at the week level, and could not reject the

null hypothesis of no serial correlation in generic share. Although the test detected serial correlation in quantity purchased, our results using quantity as an outcome do not change when correcting for serial correlation (Appendix Table A7). We also estimated treatment effects on quantity using a conditional fixed effects Poisson model (Appendix Table A7) to account for the variation in baseline quantity across distinct products. Results match the effect sizes in Tables 3 and 4, with greater precision: Test 3 increases the quantity sold of treated products by 11% ($p=0.024$) and Test 2a, where the percentage price difference is highlighted as “savings,” increases their quantity by 13.5% ($p=0.046$).

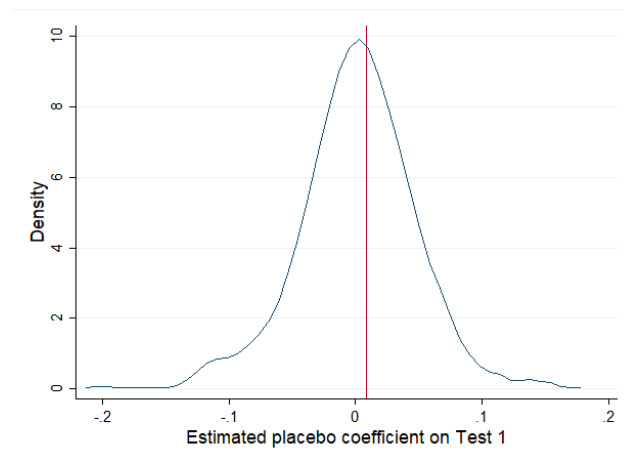
Next, we show in the appendix that our main store-level results are robust to the wild cluster bootstrap procedure. The first columns in Appendix Tables A8 and A9 show bootstrapped p -values and 95% confidence intervals for the effects of each treatment on generic share and quantity that are very similar to Table 3.

We also show that our results are not dependent on our choice of control stores. The second panels of Appendix Tables A8 and A9 show that we estimate similar treatment effects for generic share and quantity when using all other stores from the geographic division ($N=56$) as untreated.

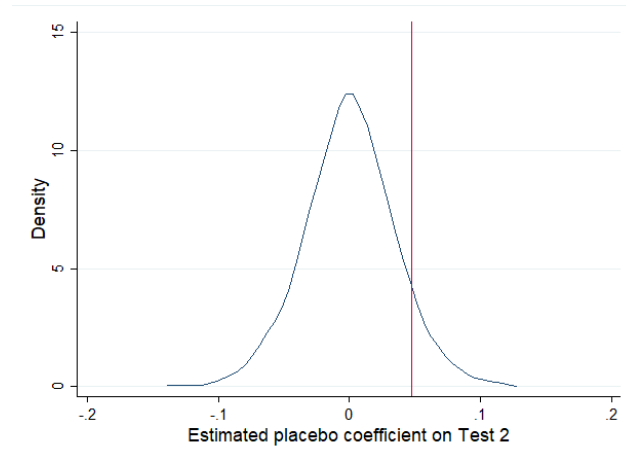
To verify that our results for generic share are not driven by short-term seasonal trends specific to the stores that we treated, we conducted a falsification test. We used sales data from 2010 and 2011 to test for any placebo difference-in-differences effects on generic share or quantity in the stores and weeks we treated in 2012, relative to untreated stores, but during the prior years. Results are shown using both our six designated control stores and all untreated division stores as the control group, alongside results from 2012 in Table A8 (generic share) and Table A9 (quantity). None of the placebo estimates for generic share or quantity has a bootstrapped p -value smaller than 0.10.

Finally, we used a randomization inference test to nonparametrically estimate p -values for the main estimates of our three tests’ treatment effects. We

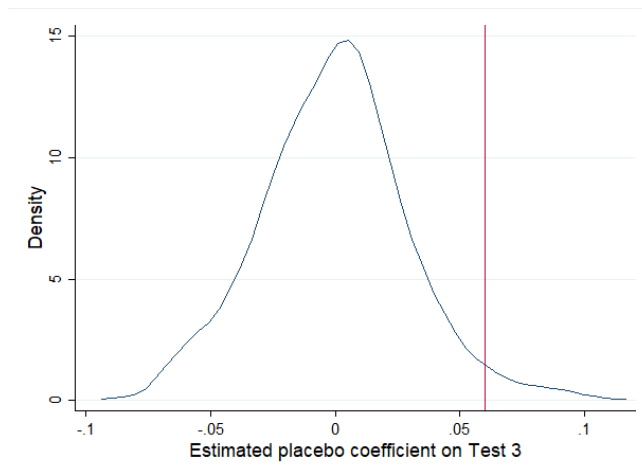
used the data from the pre-treatment and treatment periods in all 56 stores. In each of 1,500 permutations, we randomly drew one store to be assigned Treatment 1, two stores to be assigned Treatment 2, three stores to be assigned Treatment 3, and six stores to be used as control stores. We also randomly re-drew which of the drug class groups were assigned treatment within each symptom category (see Appendix Figure A1). 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 (shown as red vertical lines) compare to the point estimates obtained over 1,500 draws. These results show that for Treatment 3, 5.5% of randomly selected permutations yield a treatment effect as large as the one we obtain. This is consistent with the p-value of 0.06 obtained with the wild cluster bootstrap using all stores in the control group (Appendix Table A8, Panel B).



Panel A: Labeling Treatment 1



Panel B: Labeling Treatment 2



Panel C: Labeling Treatment 3

Figure 6. 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 simple model of observational learning, in which shoppers place a constant weight on the choices of others, 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 null net effect of the labels on generic share when the current share of generic customers is around 50% and priors are either unbiased or symmetrically biased towards what one personally buys. There are two reasons why posted shares close to 50% could, however, lead to a net increase in the generic purchase share: (1) generic-buying customers may put less weight on the choices of others as a signal of drug quality, because they have already tried both the brand and generic, or (2) brand-buying customers may have priors of the generic purchase share that are further from the true purchase ratios than the priors of generic-buying customers, on average.

To explore these two possible explanations, we surveyed 298 customers in three of the retailer's locations.²⁶ Customer participants were first asked to make a hypothetical choice between the brand and generic versions of an over-the-counter painkiller, with prices shown. They were asked to guess what share of other shoppers at the store would make the same choice that they did. Then, 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 information were true, or if they would consider switching to the generic (or brand) product.

Before presenting the findings related to the two possible explanations described above, 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 personal experience with the brand-name product. By comparison, only 65% of respondents have ever purchased a generic version of the painkiller. Among consumers who have tried both the brand and the generic, 81% make a hypothetical choice of buying the generic at the typical list prices shown to them,

²⁶ Details are provided in Appendix B. We avoided stores that were treated during the treatment period to reduce the probability of surveying customers who had seen the posted labels.

similar to the patterns observed among repeat consumers in the household-level dataset. Of those who have never tried the generic, only 37% make the hypothetical choice of the generic in the survey. Consistent with being less likely to have tried the generic, more brand-choosing respondents than generic-choosing respondents answered “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 work equally well (16% vs. 6% of generic-choosing customers, $p=0.018$; see Appendix Table A10). These responses are consistent with the hypothesis that many brand buyers have imprecise priors regarding the efficacy of generic drugs, relative to the brand.

The survey responses also reveal diffuse 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%.²⁷ We find that consumers who choose the brand, on average, believe that fewer consumers choose the generic (49%) than consumers who choose the generic themselves (58%, $p<.001$). The guesses of those who choose the brand are actually closer, on average, to the true proportions in sales data. 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, 20% said they would “probably” or “definitely” buy the generic if this was the case. By contrast, only

²⁷ 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.”

8.9% of generic buyers said they would “probably” or “definitely” buy the brand if they learned that buying the generic was less common than they thought ($p < .01$).²⁸ With the caveat that these questions were framed as hypotheticals, we interpret this as suggestive evidence 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. This is consistent with over-the-counter drugs being an experience good; those who have already tried a product are less uncertain about its therapeutic value.

5. Conclusion

Unlike prescription drugs, OTC drugs are purchased by consumers with unrestricted choice over competing products and direct access to prices and standardized drug facts. Nevertheless, entrenched brand preferences among near substitutes may result from biased beliefs or uncertainty 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 or uncertainty about the generic that can be addressed with information on their peers’ purchases.

We found no evidence for the first hypothesis. Labels providing comparability information had no overall impact on purchases. Labels that displayed price differences as “savings” in percentage terms increased generic purchasing, but primarily among customers who had already bought OTC drugs at this retailer. Labels that highlighted the same price difference as a price premium charged by the national brand did not influence purchasing. While our results on Test 2 overall were mixed, we believe interventions that increase the salience of

²⁸ 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 as they chose. Details of this process are in the Appendix.

price-savings deserve more testing, as we cannot reject the possibility of effects as strong as those of Test 3.

We found the strongest evidence for the third hypothesis. Labels showing the share of customers at a given store who buy the generic (or brand) version of each product increased generic purchase rates by about 6 percentage points overall, and 8 percentage points among customers who did not previously buy the generic. This effect is more than three-fifths of the size of the average increase associated with price promotions, suggesting it is equivalent to reducing the generic's price by an additional 3.4 percent,²⁹ and is particularly strong among brand-loyal customers. With the average consumer's generic share for treated products increasing from 45% to 51%, a back-of-the-envelope calculation implies a reduction of 2.8% in total spending on these products. A complete conversion from 45% to 100% generic, by comparison, would reduce spending by about 25%.³⁰

The fact that the effect was similar regardless of whether the labels provided the share buying the generic or the share buying the brand suggests that the effect was driven by observational learning rather than an implicit encouragement to choose the generic. Our survey evidence also suggests that customers who have never yet tried the generic may be more responsive to the information that other customers frequently choose it. We conclude that some brand-loyal consumers are wary of the quality or safety of generic products but respond to the information that other customers find them acceptable. Posting market shares of generic products proved effective in our setting, a grocery store chain, but future work could explore the effects in different retail environments and for other types of experience goods. Other interventions that publicize customer ratings and reviews may also be fruitful avenues for future research.

²⁹ A price promotion of the generic reflects, on average, a 5.5% price reduction.

³⁰ On average, our treated generic products cost 62% as much as brand equivalents.

These findings have implications for policies aimed at promoting evidence-based and cost-effective choices. Public health interventions and employer wellness initiatives using peer promotion may be more effective than simply advertising facts. Given the growing emphasis in U.S. healthcare on giving patients an active role in the choice of their treatments, this paper provides a first stab at better understanding what types of new information can shift their choices.

References

- Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Woolridge. 2017. "When Should You Adjust Standard Errors for Clustering?" NBER Working Paper 24003.
- 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.
- Allcott, Hunt., 2011. Social norms and energy conservation. *Journal of Public Economics* 95 (9): 1082–1095.
- 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.
- Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50 (2): 317–72.
- 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.
- Chetty, Raj, Looney, Adam, and Kory Kroft, 2009. "Salience and Taxation: Theory and Evidence" *American Economic Review* 99(4): 1145-1177.

Chevalier, J. A., & Mayzlin, D., 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, **43**(3), 345– 354.

Cui, G., Lui, H. -K., & Guo, X. , 2012. The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce*, **17**(1), 39– 58.

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.

Khare, A., L. I. Labrecque, A. K. Asare, 2011, The Assimilative and Contrastive Effects of Word-of-Mouth Volume: An Experimental Examination of Online Consumer Ratings, *Journal of Retailing*, 87 (1): 111-126.

Lu, J., X. Su, Y. Diao, N. Wang, B. Zhou, 2021. Does online observational learning matter? Empirical evidence from panel data, *Journal of Retailing and Consumer Services*, 60, 102480.

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.

Roodman, David, Morten Ørregaard Nielsen, James G. MacKinnon, and Matthew D. Webb. 2019. "Fast and Wild: Bootstrap Inference in Stata Using Boottest." *Stata Journal* 19 (1): 4–60.

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.

Webb, M. D. 2014. *Reworking wild bootstrap based inference for clustered errors*. QED Working Paper 1315, Queen's University, Department of Economics.

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.

Table 1. Descriptive Statistics, Pre-treatment period*Panel A: Treated Products*

	Control stores	Test 1 stores	Test 2 stores	Test 3 stores	All
Brand price per unit	0.44 (0.01)	0.43 (0.03)	0.44 (0.02)	0.43 (0.02)	0.44 (0.01)
Generic price, as a share of brand price	0.62 (0.01)	0.62 (0.01)	0.62 (0.01)	0.62 (0.01)	0.62 (0.01)
Weekly quantity sold per product	11.30 (0.56)	13.48 (1.20)	8.37 (0.67)	8.36 (0.47)	10.26 (0.34)
Generic share	0.47 (0.01)	0.48 (0.03)	0.39 (0.02)	0.43 (0.02)	0.45 (0.01)

Panel B: Untreated Products

	Control stores	Test 1 stores	Test 2 stores	Test 3 stores	All
Brand price per unit	0.38 (0.01)	0.40 (0.03)	0.39 (0.02)	0.38 (0.02)	0.38 (0.01)
Generic price, as a share of brand price	0.54 (0.01)	0.52 (0.02)	0.52 (0.02)	0.56 (0.01)	0.54 (0.01)
Weekly quantity sold per product	12.06 (0.73)	17.32 (2.53)	9.90 (1.11)	9.89 (1.05)	11.60 (0.53)
Generic share	0.47 (0.01)	0.47 (0.03)	0.40 (0.03)	0.49 (0.02)	0.46 (0.01)

Notes: This table shows means and their standard errors across all product-week-store observations for pre-treatment period weeks 2012wk14-2012wk19. Weekly quantity is the number of packages sold per product (same active ingredient but may vary in units (count of doses), brand, pill type or inactive ingredients). Prices are in dollars per unit, inclusive of discounts, averaged over the different UPCs sold for each active ingredient, weighted by purchase share, 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, by week.

Table 2: Difference-in-Differences in Generic Share, Pooled Treatments*Panel A: Treated Products*

	Control Stores	Treatment Stores	Difference:
Pre-Treatment Period means	0.465	0.427	-0.039
s.e.	(0.022)	(0.018)	(0.029)
N	461	460	921
Treated Period means	0.455	0.473	0.018
s.e.	(0.022)	(0.027)	(0.035)
N	307	305	612
Difference over time	-0.010	0.047	DD _{t.p.} = 0.056
s.e.	(0.021)	(0.023)	(0.031)
N	768	765	1,533

Panel B: Untreated Products

	Control Stores	Treatment Stores	Difference:
Pre-Treatment Period	0.465	0.459	-0.007
s.e.	(0.026)	(0.024)	(0.034)
N	421	414	1,168
Treated Period means	0.456	0.439	-0.018
s.e.	(0.027)	(0.028)	(0.036)
N	283	271	764
Difference over time	-0.009	-0.020	DD _{u.p.} = -0.011
s.e.	(0.015)	(0.015)	(0.027)
N	704	685	1,389
Triple difference			DDD = 0.068
s.e.			(0.041)
N			2,922

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 at the drug class-by-store level.

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	-0.03 (0.05)	0.01 (0.04)	-0.05 (0.05)	-0.01 (0.04)	1.26 (1.14)	1.09 (1.04)	0.53 (1.25)	0.40 (1.16)
Test 2: Price comparison	0.06 (0.05)	0.05 (0.03)	-0.03 (0.04)	0.01 (0.02)	1.32 (0.86)	1.18 (0.84)	0.64 (0.60)	0.58 (0.60)
Test 3: Observational Learning	0.08 (0.04)	0.06 (0.02)	0.02 (0.03)	0.02 (0.02)	1.30 (0.78)	1.30 (0.75)	0.64 (0.39)	0.64 (0.41)
Generic on promotion		0.09 (0.03)		0.08 (0.02)		0.45 (0.75)		0.05 (0.62)
Brand on promotion		-0.07 (0.02)		-0.11 (0.02)		1.42 (0.56)		0.37 (0.45)
Both on promotion		-0.04 (0.04)		-0.01 (0.03)		0.31 (0.68)		1.93 (0.79)
Weighted by quantity	No	Yes	No	Yes	--	--	--	--
N, observations	1533	1533	1389	1389	1560	1556	1440	1438
N, drug class X store clusters	48	48	48	48	48	48	48	48
Dependent variable mean	0.45	0.45	0.46	0.46	10.9	10.9	11.6	11.6
Tests of equality, p -values								
H_0 : Test 1 = Test 2	0.13	0.34	0.76	0.70	0.96	0.93	0.93	0.89
H_0 : Test 1 = Test 3	0.03	0.17	0.24	0.51	0.96	0.81	0.93	0.84
H_0 : Test 2 = Test 3	0.66	0.65	0.22	0.69	0.98	0.83	1.00	0.92
H_0 : Test 1 = Test 2 = Test 3	0.08	0.38	0.32	0.78	1.00	0.96	1.00	0.98

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 during the same four week period. In six control stores, no labeling tests were conducted. "Generic on promotion" and "Brand on promotion" are dummy variables indicating that any package of a generic or brand product are on a price promotion during that week in that store. Store and product fixed effects are included in all specifications. Standard errors clustered at the drug class-by-store level are in parentheses; significance stars are omitted.

Table 4: Treatment Effects by Label Content and Framing

	Generic share		Quantity	
	Treated Products	Untreated Products	Treated Products	Untreated Products
	(1)	(2)	(3)	(4)
Test 1: Comparability Statement				
a. "Same active ingredient"	0.000 (0.036) {0.99} [-0.343, 0.211]		3.049 (2.027) {0.19} [-14.273, 19.952]	
b. "... and approved by the FDA"	-0.007 (0.050) {0.90} [-0.759, 0.869]		-0.209 (1.176) {0.86} [-21.842, 16.368]	
c. "FDA determined bioequivalence"	0.037 (0.032) {0.61} [-0.087, 0.479]		-0.851 (1.306) {0.59} [-14.663, 2.987]	
No Label (untreated)		-0.008 (0.037) {0.85} [-0.215, 0.061]		0.398 (1.159) {0.85} [-1.755, 4.293]
Test 2: Price Comparison				
a. Framing: Save X%	0.069 (0.037) {0.09} [-0.040, 0.265]	0.002 (0.039) {0.96} [-0.155, 0.071]	1.675 (0.945) {0.12} [-0.616, 4.601]	0.569 (1.058) {0.75} [-1.885, 2.687]
b. Framing: Pay Y% More	0.025 (0.034) {0.50} [-0.149, 0.174]	0.014 (0.017) {0.56} [-0.049, 0.104]	0.674 (0.972) {0.56} [-2.070, 3.533]	0.590 (0.417) {0.31} [-0.613, 1.468]
Test 3: Observational Learning				
a. Framing: X% choose generic	0.051 (0.026) {0.10} [-0.008, 0.110]	0.027 (0.025) {0.37} [-0.051, 0.083]	1.527 (0.779) {0.07} [-0.133, 3.159]	0.570 (0.502) {0.31} [-0.691, 1.644]
b. Framing: Y% choose brand	0.079 (0.026) {0.10} [-0.033, 0.202]	-0.002 (0.014) {0.91} [-0.064, 0.057]	0.838 (0.789) {0.33} [-1.800, 2.797]	0.780 (0.517) {0.17} [-0.591, 2.400]
Weighted by quantity	Yes	Yes	--	--
N, observations	1533	1389	1556	1438
N, drug class X store clusters	48	48	48	48
Dependent variable mean	0.45	0.46	10.9	11.6
Tests of equality, <i>p</i> -values				
Test 1 Statements	0.26		0.43	
Test 2 Framing Variations	0.35	0.78	0.38	0.99
Test 3 Framing Variations	0.42	0.32	0.19	0.77

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 label statements for Test 1 were varied at the level of the product, based on generic product's FDA status, and tested at one store. The framing of the information presented in Tests 2 and 3 was varied at the store level. Each framing variation for Tests 2 and 3 was tested at one store, with the exception of the first framing of Test 3 ("X% choose generic") which was tested at two stores. Controls for price promotions are included as in Table 3, and the generic share regressions are weighted by purchase quantity. Standard errors, clustered at the drug class-by-store level, are in parentheses. Braces and square brackets below contain p-values and 95% confidence intervals from a wild cluster bootstrap procedure (Stata boottest, 2000 replications, Webb weights) which is also used to conduct tests of coefficient equality.

Table 5: Treatment Effects and Heterogeneity at the Household Level

<i>Y</i> = Generic purchased	Treated Products			Untreated Products		
	(1)	(2)	(3)	(4)	(5)	(6)
Test 1: Comparability Statement	-0.011 (0.026) {0.73} [-0.25, 0.0462]	-0.007 (0.020) {0.75} [-0.15, 0.0336]	-0.002 (0.037) {0.97} [-0.40, 0.0777]	-0.009 (0.027) {0.79} [-0.43, 0.295]	-0.003 (0.033) {0.88} [-0.56, 0.550]	-0.028 (0.031) {0.46} [-0.53, 0.491]
Test 2: Price comparison	0.065 (0.031) {0.06} [-0.0073, 0.163]	0.054 (0.028) {0.07} [-0.0052, 0.143]	0.061 (0.031) {0.07} [-0.0056, 0.160]	-0.000 (0.023) {0.98} [-0.13, 0.0356]	-0.007 (0.041) {0.89} [-0.16, 0.0640]	-0.021 (0.045) {0.69} [-0.19, 0.0621]
Test 3: Observational Learning	0.074 (0.015) {0.002} [0.037, 0.108]	0.056 (0.013) {0.001} [0.024, 0.0840]	0.081 (0.014) {0.000} [0.048, 0.112]	-0.003 (0.020) {0.88} [-0.060, 0.0393]	0.013 (0.015) {0.47} [-0.031, 0.0438]	-0.002 (0.025) {0.97} [-0.092, 0.0491]
Test 1 x Previous Use of Generic			0.002 (0.080) {0.98} [-0.17, 0.280]			0.086 (0.030) {0.24} [-0.090, 0.229]
Test 2 x Previous Use of Generic			-0.008 (0.057) {0.89} [-0.17, 0.171]			0.050 (0.045) {0.28} [-0.040, 0.222]
Test 3 x Previous Use of Generic			-0.060 (0.032) {0.08} [-0.13, 0.00615]			0.045 (0.066) {0.64} [-0.12, 0.228]
Previous Use of Generic		0.477 (0.017) {0.000} [0.44, 0.512]	0.498 (0.019) {0.000} [0.46, 0.536]		0.465 (0.017) {0.000} [0.43, 0.502]	0.474 (0.020) {0.000} [0.43, 0.523]
Generic share of all OTC drugs previously purchased		0.304 (0.016) {0.000} [0.27, 0.337]	0.304 (0.015) {0.000} [0.27, 0.338]		0.264 (0.019) {0.000} [0.22, 0.304]	0.263 (0.019) {0.000} [0.22, 0.304]
Previous use of purchased product (brand or generic version)		-0.203 (0.013) {0.000} [-0.23, -0.176]	-0.203 (0.013) {0.000} [-0.23, -0.175]		-0.197 (0.013) {0.000} [-0.22, -0.170]	-0.196 (0.013) {0.000} [-0.22, -0.170]
N	8809	8809	8809	8256	8256	8256
N, drug class X store clusters	48	48	48	48	48	48
R-sq	0.115	0.388	0.389	0.118	0.389	0.390
Dependent variable mean	0.45	0.45		0.48	0.48	
Previous Choice of Brand			0.31			0.34
Previous Choice of Generic			0.86			0.83
Tests of equality, <i>p</i> -values						
H ₀ : Test 1 = Test 2	0.07	0.09	0.24	0.84	0.96	0.93
H ₀ : Test 1 = Test 3	0.01	0.02	0.09	0.88	0.68	0.58
H ₀ : Test 2 = Test 3	0.80	0.95	0.59	0.94	0.67	0.78
H ₀ : Test 1 = Test 2 = Test 3	0.04	0.06	0.18	0.97	0.87	0.81

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. The sample is limited to households with at least one prior purchase of an OTC product during the first thirteen weeks of 2012 (prior to the start of the pre-treatment time 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 household having purchased the generic version of the OTC drug currently being purchased, in their last observed purchase prior to the pre-treatment time period. All models include store and product fixed effects, controls for price promotions, and a dummy for the treatment period. Models 2, 3, 5, and 6 include household-level controls such as total number of prior OTC drug purchases and past average percentage discount on total purchases at the retailer. 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, clustered at the drug class-by-store level, are in parentheses. Braces and square brackets below contain *p*-values and 95% confidence intervals from a wild cluster bootstrap procedure (Stata *boottest*, 2000 replications, Webb weights), which is also used to conduct tests of coefficient equality. Significance stars are omitted.

Table 6: Post-treatment Effects on Choice of Generic

<i>Y = Generic purchased</i>	Treated Products within Treated Stores			Untreated Products within Treated Stores		
	Weeks 24-27 (1)	Weeks 28-31 (2)	Weeks 32-36 (3)	Weeks 24-27 (4)	Weeks 28-31 (5)	Weeks 32-36 (6)
Test 1 x Post x Exposed	0.065 (0.044) {0.42} [-0.057, 0.261]	-0.017 (0.043) {0.64} [-0.173, 0.120]	0.033 (0.029) {0.35} [-0.079, 0.282]	-0.019 (0.015) {0.59} [-0.188, 0.237]	0.002 (0.052) {0.95} [-1.062, 0.859]	0.012 (0.042) {0.84} [-0.500, 0.191]
Test 2 x Post x Exposed	-0.039 (0.047) {0.47} [-0.123, 0.119]	-0.018 (0.051) {0.73} [-0.176, 0.083]	0.008 (0.027) {0.79} [-0.056, 0.081]	-0.069 (0.041) {0.33} [-0.154, 0.049]	-0.086 (0.039) {0.13} [-0.212, 0.088]	-0.049 (0.031) {0.35} [-0.124, 0.082]
Test 3 x Post x Exposed	0.065 (0.031) {0.08} [-0.010, 0.138]	0.050 (0.032) {0.18} [-0.029, 0.121]	0.036 (0.031) {0.32} [-0.040, 0.104]	-0.056 (0.030) {0.09} [-0.138, 0.012]	-0.022 (0.045) {0.69} [-0.099, 0.120]	0.021 (0.037) {0.62} [-0.097, 0.095]
Baseline effect: Post period						
Test 1 x Post	0.012 (0.017) {0.55} [-0.063, 0.089]	0.064 (0.016) {0.05} [-0.009, 0.136]	0.027 (0.020) {0.40} [-0.057, 0.143]	-0.004 (0.034) {0.88} [-0.605, 0.768]	0.029 (0.042) {0.62} [-0.203, 0.676]	0.013 (0.043) {0.71} [-0.171, 0.534]
Test 2 x Post	-0.021 (0.032) {0.53} [-0.129, 0.062]	0.029 (0.035) {0.54} [-0.038, 0.144]	0.027 (0.020) {0.18} [-0.015, 0.085]	0.031 (0.044) {0.79} [-0.106, 0.108]	0.031 (0.024) {0.35} [-0.079, 0.102]	0.013 (0.021) {0.60} [-0.094, 0.052]
Test 3 x Post	-0.004 (0.029) {0.88} [-0.067, 0.066]	0.025 (0.028) {0.41} [-0.039, 0.092]	0.034 (0.023) {0.22} [-0.021, 0.088]	0.043 (0.011) {0.02} [0.015, 0.073]	0.035 (0.033) {0.46} [-0.078, 0.101]	-0.011 (0.033) {0.78} [-0.106, 0.082]
Baseline effect: Exposed						
Test 1 stores x Exposed	-0.036 (0.030) {0.45} [-0.253, 0.079]	-0.030 (0.027) {0.48} [-0.300, 0.074]	-0.038 (0.028) {0.39} [-0.123, 0.055]	-0.016 (0.034) {0.62} [-0.532, 0.564]	-0.019 (0.029) {0.60} [-0.458, 0.487]	-0.023 (0.031) {0.57} [-0.480, 0.463]
Test 2 stores x Exposed	-0.005 (0.029) {0.85} [-0.065, 0.076]	0.002 (0.027) {0.95} [-0.055, 0.078]	-0.010 (0.025) {0.69} [-0.061, 0.062]	-0.007 (0.029) {0.80} [-0.112, 0.055]	-0.008 (0.029) {0.80} [-0.109, 0.056]	-0.009 (0.030) {0.78} [-0.114, 0.063]
Test 3 stores x Exposed	-0.034 (0.018) {0.10} [-0.076, 0.006]	-0.024 (0.017) {0.24} [-0.062, 0.019]	-0.033 (0.019) {0.17} [-0.076, 0.017]	-0.010 (0.022) {0.63} [-0.078, 0.048]	-0.010 (0.023) {0.66} [-0.088, 0.061]	-0.020 (0.026) {0.47} [-0.117, 0.056]
N	3440	3715	4242	3551	3887	4597
N, drug class X store clusters	0.45	0.45	0.45	0.48	0.48	0.48
Dependent variable mean	24	24	24	24	24	24

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, capturing the difference in generic purchase share overall between the pre-treatment period and the specified post-treatment period for that set of stores. "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. The interactions "Test X x Exposed" capture the average difference between generic purchase rates among exposed and unexposed customers within Test X store(s) in the pre-treatment period. Standard errors, clustered at the drug class-by-store level, are in parentheses. Braces and square brackets below contain p-values and 95% confidence intervals from a wild cluster bootstrap procedure (Stata boottest, 2000 replications, Webb weights). Significance stars are omitted.