NBER WORKING PAPER SERIES

DO PHARMACISTS BUY BAYER? INFORMED SHOPPERS AND THE BRAND PREMIUM

Bart J. Bronnenberg
Jean-Pierre Dubé
Matthew Gentzkow
Jesse M. Shapiro

Working Paper 20295
http://www.nber.org/papers/w20295

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2014

We are grateful to Kevin Murphy for inspiring this project and to numerous seminar participants for comments. We thank Jeff Eastman, Grace Hyatt and Todd Kaiser at Nielsen, for their assistance with the collection of the PanelViews Survey and Art Middlebrooks, Svetlozar Nestorov, and the Marketing Data Center at the University of Chicago Booth School of Business for their help with the Homescan and RMS data. We thank the Neubauer Family Foundation, the Initiative on Global Markets at the University of Chicago Booth School of Business, the Marketing Science Institute, the National Science Foundation, and the Netherlands Foundation for Scientific Research (NWO, Vici grant) for financial support. Results in this paper are calculated based on data from The Nielsen Company (US), LLC. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w20295.ack

Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium
Bart J. Bronnenberg, Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M. Shapiro
NBER Working Paper No. 20295
July 2014, Revised August 2014
JEL No. D12,D83,L66

ABSTRACT

We estimate the effect of information on consumers' willingness to pay for national brands in physically homogeneous product categories. We measure consumer information using education, occupation, and a survey-based measure of product knowledge. In a detailed case study of headache remedies we find that more informed consumers are less likely to pay extra to buy national brands, with pharmacists choosing them over store brands only 9 percent of the time, compared to 26 percent of the time for the average consumer. In a similar case study of pantry staples such as salt and sugar, we show that chefs devote 12 percentage points less of their purchases to national brands than demographically similar non-chefs. We extend our analysis to cover 50 retail health categories and 241 food and drink categories. The results suggest that misinformation explains a sizable share of the brand premium for health products, and a much smaller share for most food and drink products. We tie our estimates together using a stylized model of demand and pricing under misinformation.

Bart J. Bronnenberg
Tilburg University and CentER
Warandelaan 2, Koopmans K-1003
5037 AB Tilburg
The Netherlands
bart.bronnenberg@uvt.nl

Jean-Pierre Dubé
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
jdube@chicagobooth.edu

Matthew Gentzkow
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
gentzkow@chicagobooth.edu

Jesse M. Shapiro
University of Chicago
Booth School of Business
5807 S. Woodlawn Avenue
Chicago, IL 60637
and NBER
jesse.shapiro@chicagobooth.edu

An online appendix is available at:
http://www.nber.org/data-appendix/w20295

1 Introduction

A 100-tablet package of 325mg Bayer Aspirin costs \$6.29 at cvs.com. A 100-tablet package of 325mg CVS store-brand aspirin costs \$1.99 (CVS 2013). The two brands share the same dosage, directions, and active ingredient. Aspirin has been sold in the United States for more than 100 years, CVS explicitly directs consumers to compare Bayer to the CVS alternative, and CVS is one of the the largest pharmacy chains in the country, with presumably little incentive to sell a faulty product. Yet the prevailing prices are evidence that some consumers are willing to pay a three-fold premium to buy Bayer.[1] Research shows that markets for automobiles (Sullivan 1998), index funds (Hortaçsu and Syverson 2004), and online books (Smith and Brynjolfsson 2001) all exhibit substantial brand premia even within groups of physically homogeneous products.

Many economists have hypothesized that consumers' willingness to pay for national brands in homogeneous product categories reflects advertising-induced misinformation.[2] Others have pointed out that branded goods may in fact produce more consumer utility, either because advertising is a complement to consumption (Becker and Murphy 1993), or because even seemingly similar brands differ in subtle ways.[3] Determining how much of the brand premium reflects misinformation has important implications for consumer welfare. We estimate that consumers spend \$196 billion annually in consumer packaged goods categories in which a store-brand alternative to the national brand exists, and that they would spend approximately \$44 billion less (at current prices) if they switched to the store brand whenever possible. If consumers are systematically misled by brand claims, this has clear implications for evaluating the welfare effects of the roughly \$140 billion spent on advertising each year in the US (Kantar Media 2013), and for designing federal regulation to minimize the potential for harm (e.g., Federal Trade Commission 1999).

In this paper, we estimate how much of the brand premium for drug-store and supermarket products

[1] Indeed, in data we introduce below, 25 percent of aspirin sales by volume (and 60 percent by expenditure) are to national-brand products.

[2] Braithwaite (1928) writes that advertisements "exaggerate the uses and merits" of national brands, citing aspirin and soap flakes as examples. Simons (1948) advocates government regulation of advertising to help mitigate "the uninformed consumer's rational disposition to 'play safe' by buying recognized, national brands" (p. 247). Scherer (1970) discusses premium prices for national-brand drugs and bleach, and writes that "it is hard to avoid concluding that if the housewife-consumer were informed about the merits of alternative products by some medium more objective than advertising and other image-enhancing devices, her readiness to pay price premiums as large as those observed here would be attenuated" (pp. 329-332). More recently, a growing body of theoretical work considers markets with uninformed or manipulable consumers (Gabaix and Laibson 2006; Ellison and Wolitzky 2012; Piccione and Spiegler 2012).

[3] In one instance, the FDA determined that a generic antidepressant performed less well than its branded counterpart, likely due to differences in their "extended release" coatings (Thomas 2012). A widely publicized 2006 recall of store-brand acetaminophen resulted from the discovery that some pills could contain metal fragments (Associated Press 2006); such risks could conceivably be lower for national brands. Hortaçsu and Syverson (2004) conclude that purchases of high-cost "brand name" index funds partly reflect willingness to pay for non-financial objective attributes such as tax exposure and the number of other funds in the same family.

results from lack of information. We match individual purchase data from the 2004-2011 Nielsen Homescan panel to a new survey containing proxies for consumer information, and to separate data on store-level quantities and prices. We estimate the effect of our information measures on the propensity to choose store brands over national brands, and study the choices of experts such as pharmacists and physicians as an approximation to behavior under perfect information. We then use these estimates, in conjunction with a stylized model of demand and pricing, to quantify how the division of surplus would change in a world in which all consumers were perfectly informed.

Our main identification challenge is to separate the effect of consumer information from other drivers of choice, such as preferences and product availability, that may be correlated with a consumer's information. With regard to preferences, we limit the scope for unmeasured heterogeneity by focusing on choices between store and national brands that are identical on all physical attributes measured by Nielsen. We further include detailed controls for income and other demographics, and compare occupations (e.g., physicians and lawyers) with similar socioeconomic status but different levels of product-specific expertise. We show that well-informed consumers look similar to other consumers in their preferences for measured product attributes, making it more plausible that they are similar in their preferences for any unmeasured attributes. We argue that whatever unmeasured preference heterogeneity remains would be likely to work against our main findings.

With regard to product availability and other store-level drivers of choice, we limit the scope for heterogeneity by comparing informed and uninformed consumers who shop in the same chain, market, and time period. We address confounds related to workplace purchases (e.g., pharmacists receiving free samples or discounts that affect their purchasing behavior) by studying experts who are no longer employed at their specialty. Though we cannot rule out all possible confounds, the pattern of evidence suggests our estimates mainly capture the causal effect of information.

We begin our analysis with a detailed case study of headache remedies. As indirect measures of information, we use the primary shopper's occupation, educational attainment, and college major. We also measure information directly through a survey of a subset of Nielsen panelists, in which we ask the panelists to name the active ingredient in various national-brand headache remedies.

The relationship among our information proxies is intuitive. The average respondent answers 59 percent of our active ingredient questions correctly. For the college-educated, this fraction rises to 62 percent. For those whose major was science or health, it is 73 percent. For registered nurses it is 85 percent, for pharmacists it is 89 percent, and for physicians and surgeons it is 90 percent. Occupational specialty is important enough to outweigh large differences in general education. For example, registered nurses are far better informed about headache remedies than lawyers, despite having completed less schooling and earning less in the labor market on average.

We find that more informed households are consistently more likely to buy store-brand headache remedies. The average household devotes 74 percent of headache remedy purchases to store brands. Controlling for household income, other demographics, and interacted fixed effects for the market, chain, and quarter in which the purchase is made, a household whose primary shopper correctly identifies all active ingredients is 19 percentage points more likely to purchase a store brand than a shopper who identifies none. Having a college-educated primary shopper predicts an increase of 4 percentage points, having a primary shopper with a healthcare occupation other than pharmacist or physician predicts an increase of 8 percentage points, and having a primary shopper who is a pharmacist or physician predicts an increase of 15 percentage points, with pharmacists buying store brands 91 percent of the time. Primary shoppers with science majors buy more store brands than those with other college degrees, and the effect of occupation is sizable among consumers not currently employed.

In a second case study of pantry staples such as salt, sugar, and baking soda, we find that chefs devote 77 percent of their purchases to store brands, as compared to 60 percent for the average consumer. The effect of being a chef is large and highly significant after including our detailed vector of controls. Food preparers who are not chefs are also significantly more likely to buy store brands than others who are demographically similar.

We find that the effects of consumer information are largely domain-specific. Neither knowledge of headache remedy active ingredients nor working in a healthcare occupation predicts store-brand purchases in pantry staple categories. Similarly, working in a food preparer occupation other than chef does not predict store-brand headache remedy purchases. We do find that chefs buy more store-brand headache remedies, possibly suggesting that some of their knowledge is transferable across domains.

We extend the approach from our two case studies to the full set of products in which there is a comparable store-brand alternative to national brands, and sufficient purchase volume to perform a reliable analysis. Among 50 health-related categories, the effects of knowledge of headache remedy active ingredients, working in a healthcare occupation other than pharmacist or physician, and working as a pharmacist or physician are positive for 43, 43, and 34 categories respectively. A substantial number of these positive coefficients—including a large share of those for over-the-counter medications—are both economically and statistically significant. On average across these categories, working as a pharmacist or physician reduces the probability of buying the national brand by roughly a fourth. Results are less consistent for the 241 food and drink categories that we study, with the effect of being a chef positive for 148 categories and negative for 93. Several of the positive coefficients are economically and statistically significant—including a number of pantry staples and other products such as baking mixes and dried fruit—but a large majority are not individually distinguishable from zero. The average effect of working as a chef is to reduce the probability of buying a national brand by 2 percent. We find suggestive evidence that the effect of information on the propensity to buy the store brand is greater the more advertising-intensive is the category and the more agreement there is among experts that store and national brands are equivalent.

Taken together, our estimates suggest that lack of information explains a sizable portion of the brand premium in many health categories, as well as in certain food categories (such as pantry staples) with little physical variation across brands. At the same time, our results suggest a smaller role for information in the many categories—including the majority of foods and beverages—in which even experts are willing to pay a premium to buy national brands.

To sharpen these conclusions, the final section of the paper interprets our findings through the lens of a stylized model of demand and pricing under misinformation. In the model, a set of symmetric retailers offer a store brand to compete with a single national-brand manufacturer. Households sequentially choose a retail outlet and then a brand (store or national). Prices are set simultaneously by retailers and by the manufacturer of the national brand. Households differ in their willingness to pay for the national brand. A set of informed shoppers, too small to impact market prices, perceive a different (smaller or greater) gap in utility between national and store brand than does a typical shopper. We choose the parameters of the model to match the estimated effect of information in each category and to rationalize estimated margins on store- and national-brand goods.

The estimated model implies that consumer information greatly affects the distribution of surplus in health categories. Making all consumers as informed as a pharmacist or physician, while holding prices constant at current levels, would reduce the variable profits of the national headache remedy brands by half, equivalent to 19 percent of total expenditure. The profits of store brands would increase by 5 percent of expenditure, and consumer surplus would increase by 4 percent of expenditure. If prices were to adjust to reflect the change in consumer demand, the consumer surplus gains would be even greater. In health categories other than headache remedies, the effects are smaller though still economically significant. In food and drink categories, by contrast, information effects are quantitatively small, with effects on profits and consumer surplus of a few percent in pantry staples and less than one percent in other food and drink products. Although these conclusions are contingent on the functional form and other assumptions embedded in the model, together with the coefficient estimates they paint a consistent picture of the relative importance of information in different product categories.

It is important to stress two caveats to our welfare conclusions. First, we consider the effect of consumer information only on consumer choice and product pricing. In the longer run, if consumers were to become better informed, firms would adjust their advertising expenditures and product offerings, leading to welfare consequences beyond those that we can quantify here. Second, the welfare claims we make depend on the assumption that information per se does not affect the utility a consumer receives from a product. If, for example, believing that national-brand aspirin works better actually makes national-brand aspirin more effective at reducing headaches, then informing consumers could actually make them worse off.[4]

The primary substantive contribution of this study is to use novel data and methods to quantify the importance of information in consumer choice in an important real-world market.[5] We add to existing survey and experimental evidence[6] by exploiting multiple sources of variation in consumer information, including occupational expertise.[7] Our work complements concurrent research by Carrera and Villas-Boas (2013), who use a field experiment to assess the impact of informative product labels on the propensity to purchase store-brand headache remedies. Although we focus on over-the-counter products, our findings are relevant to policy debates about substitution between branded and generic prescription medications.[8]

Methodologically, the approach of comparing the choices of demographically similar households with different levels of product information parallels that of Bartels' (1996) study of the role of information in voting, and is close in spirit to recent work in economics by Levitt and Syverson (2008), who look at real estate agents selling their own homes, and to Johnson and Rehavi (2013), who look at the frequency with which physicians give birth by caesarean section. Our model-based extrapolation of changes in prices and welfare in a world of perfect consumer information builds on recent work that uses an equilibrium

[4]This is a limitation of any revealed-preference evidence on the effect of information, but it is especially salient here as drugs are known to have brand-related placebo effects (Branthwaite and Cooper 1981; Kamenica et al. 2013).

[5]A sizable literature examines the demographic and attitudinal correlates of purchasing store-brand consumer packaged goods (e.g., Dick et al. 1995; Richardson et al. 1996; Burton et al. 1998; Sethuraman and Cole 1999; Kumar and Steenkamp 2007; Bergès et al. 2009; Steenkamp et al. 2010) and generic prescription drugs (e.g., Shrank et al. 2009). A literature on blind taste tests finds that consumers cannot distinguish among national brands (Husband and Godfrey 1934; Allison and Uhl 1964) or between national-brand and store-brand goods (Pronko and Bowles 1949), though there are exceptions (Mason and Batch 2009). Wills and Mueller (1989) and Caves and Greene (1996) use aggregate data to estimate the role of advertising and quality in brand premia. Sethuraman and Cole (1999) analyze the drivers of willingness to pay for national brands using hypothetical choices reported on a survey.

[6]Existing evidence indicates that perceptions of similarity between national- and store-brand painkillers are correlated with stated purchase intentions (Cox et al. 1983; Sullivan et al. 1994). Cox et al. (1983) find that informing consumers of active ingredient similarity does not have a discernible effect on purchase selections.

[7]We are not aware of other research on the brand preferences of healthcare professionals. An existing literature examines the health behaviors of doctors (Glanz et al. 1982), including their propensities to use certain categories of medications like sleeping pills (Domenighetti et al. 1991). Most studies of the relationship between occupation and store-brand purchases code occupation at a high level of aggregation (white collar, blue collar, etc.) without reference to specific expertise (see Szymanski and Busch 1987 for a review). An exception is Darden and Howell (1987), who study the effect of retail work experience on elements of "shopping orientation," such as attitudes toward store clerks.

[8]Purchases of branded prescription drugs in categories where generic alternatives are available are a significant component of health costs (Haas et al. 2005). A range of policies including mandatory substitution (NIHCM 2002) and financial incentives for physicians (Endsley et al. 2006) and patients (Huskamp et al. 2003) have been used in an effort to increase the generic share.

framework to evaluate the size and determinants of brand premia (Goldfarb et al. 2009).

The remainder of the paper is organized as follows. Section 2 describes our data. Section 3 lays out our empirical strategy. Section 4 presents our results for headache remedies and pantry staples. Section 5 presents our results for other health and food categories. Section 6 presents evidence on aggregate effects and welfare. Section 7 concludes.

2 Data

2.1 The Nielsen Homescan Panel

The backbone of our data is the Nielsen Homescan Panel, which we obtained through a partnership between the Nielsen Company and the James M. Kilts Center for Marketing at the University of Chicago Booth School of Business.[9] The data include purchases made on more than 77 million shopping trips by 125,114 households from 2004 to 2011. Panelist households are given optical scanners and are asked to scan the barcodes of all consumer packaged goods they purchase, regardless of outlet or store format.[10]

For each purchase, we observe the date, the universal product classification (UPC) code, the transaction price, an identifier for the store chain in which the purchase was made, and the size of the item, which we convert to equivalent units specific to a given product category (e.g., pill counts for headache remedies or ounces for salt). We compute the share of purchases going to store brand or national brand products as the share weighted by equivalent units unless otherwise noted.

Nielsen supplies household demographic characteristics including the education of the household head, a categorical measure of household income, number of adults, race, age, household composition, home ownership, and the geographic market of residence.[11]

2.2 PanelViews Surveys

We conducted two surveys of Homescan panelists as part of Nielsen's monthly PanelViews survey. The first survey was sent electronically to 75,221 households in September of 2008 with the request that each adult in the household complete the survey separately. In total, 80,077 individuals in 48,951 households responded

[9]Information on access to the data is available at http://research.chicagobooth.edu/nielsen/. See Einav et al. (2010) for a discussion of data quality in the Homescan panel.

[10]The data include purchases from supermarkets, convenience stores, mass merchandisers, club stores, drug stores, and other retail channels for consumer packaged goods.

[11]A household's geographic market is its Nielsen-defined Scantrack market. A Scantrack market can be a metropolitan area (e.g., Chicago), a combination of nearby cities (e.g., Hartford-New Haven), or a part of a state (e.g., West Texas). There are 76 Scantrack markets in the United States.

to the survey for a household response rate of 65.1 percent. The second survey was sent electronically to 90,393 households in October 2011 with the request that each adult in the household complete the survey separately. In total, 80,205 individuals in 56,258 households responded to the survey for a household response rate of 62.2 percent.

Both surveys asked for the respondent's current or most recent occupation, classified according to the 2002 Bureau of Labor Statistics (BLS) codes (BLS 2002).[12] We match these to data on the median earnings of full-time full-year workers in each occupation in 1999 from the US Census (2000). We group occupations into categories (healthcare, food preparer) using a combination of BLS-provided hierarchies and subjective judgment. The online appendix lists the occupations in these groupings.

The first survey included a set of additional questions relating to household migration patterns. These questions were used in the analysis of Bronnenberg et al. (2012). We ignore them in the present analysis.

The second survey, designed for this study, included a series of questions about households' knowledge and attitudes toward various products. In particular, for each of five national brands of headache remedy (Advil, Aleve, Bayer, Excedrin, Tylenol), we asked each respondent who indicated familiarity with a national brand to identify its active ingredient from a list of six possible choices, or state that they "don't know."[13] For each respondent we calculate the number of correct responses, treating "don't know" as incorrect. We also asked respondents whether they agreed or disagreed with a series of statements, including "Store-brand products for headache remedy / pain relievers are just as safe as the brand name products," with responses on a 1 (agree) to 7 (disagree) scale. For each respondent, we construct an indicator equal to one if the respondent chose the strongest possible agreement and zero otherwise.

The second survey also asked respondents about their college major using codes from the National Center for Education Statistics (U.S. Department of Education 2012). We define two groups of majors for analysis: health majors, which includes all majors with the word "health" in their description,[14] and non-health science majors, which includes all majors in the physical and biological sciences.

Both surveys asked respondents to indicate whether they are their household's "primary shopper" and whether they are the "head of the household." For each household we identify a single primary shopper whose characteristics we use in the analysis, following the criteria used in Bronnenberg et al. (2012). We start with all individuals within a household who respond to the survey. We then apply the following criteria

[12]In the small number of cases where an individual provided conflicting responses to the occupation question across the two surveys we use the value from the second survey.

[13]The correct active ingredients are ibuprofen (Advil), naproxen (Aleve), aspirin (Bayer), aspirin-acetaminophen-caffeine (Excedrin), and acetaminophen (Tylenol). In each case, the six possible answers were the five correct active ingredients plus the analgesic hydrocodone.

[14]Examples include "Health: medicine," "Health: nursing," and "Health: dentistry."

in order, stopping at the point when only a single individual is left: (i) keep only self-reported primary shopper(s) if at least one exists; (ii) keep only household head(s) if at least one exists; (iii) keep only the female household head if both a female and a male head exist; (iv) keep the oldest individual; (v) drop responses that appear to be duplicate responses by the same individual; (vi) select one respondent randomly.

In appendix table 1 and the online appendix, we show that our findings go largely unchanged when we incorporate data on the characteristics of secondary shoppers into our analysis.

Throughout the paper, we restrict attention to households that answered the occupation question in one or both of our PanelViews surveys.[15]

2.3 Product Classification

Nielsen provides a set of attribute variables for each UPC code purchased by a Homescan panelist. Some of these, such as size, are available for all categories. Others are category-specific. For example the data include a variable that encodes the active ingredient for each headache remedy in the data. We harmonize the codes for essentially identical descriptors (e.g., "ACET" and "ACETAMINOPHEN" both become "ACETAMINOPHEN").

We use these descriptors to aggregate UPCs into *products*. A product is a group of UPCs that are identical on all non-size attributes provided by Nielsen. For instance, in the case of headache remedies, a product is a combination of an active ingredient (e.g., aspirin, naproxen), form (e.g., tablet, gelcap), formula (e.g., regular strength, extra strength), and brand (e.g., Bayer, Aleve, store brand). We classify products as store brands using Nielsen-provided codes, supplemented with manual corrections.

To compare store brands and national brands we aggregate products into *comparable product groups*, which are sets of products that are identical on all product attributes except for brand and item size.[16] We will use the abbreviated term *comparable* to stand in for *comparable product group* throughout the paper.

We restrict attention to comparables in which we observe at least 500 purchases with at least some purchases going to both store-brand and national-brand products.[17] We eliminate categories in which the available attribute descriptors do not provide sufficient information to identify comparable products.[18] We also eliminate categories in which the average retail price per equivalent unit for national-brand products is

[15] Nielsen provides projection factors to aggregate their panelists into a representative population. As these projection factors are not designed for the subpopulation we study we do not use them in our main analysis. In appendix table 1 we show our core results in specifications that weight by the projection factors.

[16] In appendix table 1 we show the robustness of our main results to conditioning on item size.

[17] We further eliminate comparable product groups in which fewer than 50 retail chains ever sell a store brand according to the retail scanner data we discuss in section 2.4 below.

[18] These are: deli products, fresh produce, nutritional supplements, miscellaneous vitamins, and anti-sleep products.

lower than store-brand products.[19] This leaves us with a universe of 420 comparables.

For our case study of headache remedies we consider the subset of these comparables classified by Nielsen as adult, non-migraine, daytime headache remedies.

For our case study of pantry staples we consider the subset of these comparables classified by Nielsen as table salt, sugar, or baking soda.

We restrict our sample to transactions such that at least one comparable national-brand purchase and at least one comparable store-brand purchase are observed in the Homescan data in the same retail chain and quarter as the given transaction. This restriction limits the likelihood that a national-brand product is purchased because no store-brand alternative is available (or vice versa).

Although we compute summary statistics for the universe of 420 comparables, we conduct regression analysis using only those comparables with at least $5,000$ sample purchases. We do this to ensure sufficient data to estimate models with a rich set of controls. With this restriction, there are 332 comparables available for regression analysis, including 6 headache remedies, 44 other health-related products, 6 pantry staples, 235 other food and drink products, and 41 remaining products. The online appendix lists all comparables that we use in our regression analysis.

2.4 Retail Scanner Data

To estimate prices and aggregate expenditure, we use 2008 store-level scanner data from the Nielsen Retail Measurement Services (RMS) files, which we obtained through a partnership between Nielsen and Chicago Booth's Kilts Center. These data contain store-level revenue and volume by UPC and week for approximately 38,000 stores in over 100 retail chains. We use our product classification to aggregate UPCs into products.

For each comparable, we compute average price per equivalent unit for national and store brands respectively as the ratio of total expenditure to total equivalent units across all grocery, drug, and mass merchandise stores across all weeks in 2008. We also estimate total US expenditure on national and store brands respectively by multiplying the number of equivalent units purchased in the Homescan data by (i) the ratio of total equivalent units for the comparable in RMS and Homescan, (ii) the average price per equivalent unit, (iii) the ratio of 2008 US food, drug, and mass merchandise sales to total 2008 expenditure measured in RMS.[20]

The sum of estimated total US expenditure across the comparables in our sample is \$196 billion. If all

[19] Retail prices are from retail scanner data we discuss in section 2.4 below. We exclude 34 comparables based on this condition.

[20] The Annual Retail Trade Survey of the United States Census Bureau reports 2008 annual sales in grocery stores, pharmacies and drug stores, and warehouse clubs and superstores of \$512 billion, \$211 billion, and \$352 billion, respectively, totaling \$1,075 billion (U.S. Census 2013).

observed equivalent units were purchased at the average price per equivalent unit of store brands, this sum would fall by \$44 billion or 22 percent.

2.5 Wholesale Price Data

We estimate retail margins by brand using data from National Promotion Reports' PRICE-TRAK product, obtained through Chicago Booth's Kilts Center. These data contain wholesale price changes and deal offers by UPC in 48 markets from 2006 until 2011, along with associated product attributes such as item and pack sizes. The data are sourced from one major wholesaler in each market, which is representative due to the provisions of the Robinson-Patman (Anti-Price Discrimination) Act.

We compute the average wholesale price of each product as the unweighted average post-deal price across markets. We compute retail margins by matching wholesale prices with retail prices by UPC, item size, and year. We then compute the median retail margin of national-brand and store-brand products within each comparable.[21]

3 Empirical Strategy

Let there be a set of households indexed by i. Each household must choose between a national brand and a store brand of some product. Household i believes that the national brand delivers $\Delta v_i \geq 0$ more money-metric utility than the store brand, but the true difference in utility is $\Delta\tilde{v}_i \geq 0$. The difference between the price of the national brand and the price of the store brand at the store where i shops is $\Delta p_i > 0$. We let y_i be an indicator for i choosing the store brand, and assume $y_i = 1$ if and only if $\Delta p_i \geq \Delta v_i$.

To illustrate the intuition for our empirical strategy, consider a set of households who face the same prices Δp and have the same true utility $\Delta\tilde{v}$. Suppose there is an index $\phi_i \in [0,1]$ of household i's information such that $\Delta v_i = \phi_i \Delta\tilde{v} + (1-\phi_i)\Delta v$, where Δv is the utility difference perceived by an uninformed household ($\phi_i = 0$), and $\Delta\tilde{v}$ is the utility difference perceived by a perfectly informed household ($\phi_i = 1$). By looking at how y_i varies with ϕ_i, we can learn the sign of $(\Delta v - \Delta\tilde{v})$: if y_i is increasing in ϕ_i, willingness to pay for national brands is too high ($\Delta v > \Delta\tilde{v}$); if y_i is decreasing in ϕ_i, it is too low ($\Delta v < \Delta\tilde{v}$); if y_i is independent of ϕ_i, we learn perceived willingness to pay equals true utility ($\Delta v = \Delta\tilde{v}$). In addition, if we can identify a set of expert households for whom $\phi_i \approx 1$, we can evaluate the null hypothesis that national and store brands are in fact the same ($\Delta\tilde{v}_i = 0$) by asking whether $y_i = 1$ for almost all such i.

[21] We compute the median rather than the mean retail margin to avoid the influence of outlier observations that arise due to mismatch in item size etc.

To implement this strategy, we must overcome three challenges. First, we do not directly measure information ϕ_i. We therefore form a vector K_i of proxies for ϕ_i, including knowledge of active ingredients, completed schooling, college major, and occupation.[22] These measures are proxies in the sense that the correlation of K_i with choice y_i reflects both a direct causal effect (e.g., knowing that Tylenol's active ingredient is acetaminophen directly affects choice) and an indirect effect of information correlated with K_i (e.g., consumers who know Tylenol's active ingredient also tend to be well informed about other characteristics of headache remedies).

Second, we must hold constant prices Δp_i as well as other contextual drivers of choice such as in-store displays, product positioning on store shelves, etc. We do this by assuming that all such drivers are a function of observable store and time characteristics Z_i. In our preferred specifications, Z_i will include interacted indicators for market, chain, and calendar quarter. In appendix table 1, we show that our results survive even richer controls for the timing and location of purchases.

Third, we must hold constant true preferences $\Delta\tilde{v}_i$. We focus on the choice of brand within comparable product groups that are homogeneous on measured attributes, so that variation in preferences for such attributes cannot explain variation in brand choice. We assume that any remaining preference heterogeneity can be parametrized as a function of a set of observable household characteristics X_i such as age and income. We find that controlling for income strengthens our results in many cases, and we show that a relationship between information and choice is present even among occupational groups that are similar in socioeconomic status (e.g., lawyers and physicians).[23] We also show empirically that preferences for measured attributes (e.g., regular vs. extra strength, tablet vs. caplet) do not correlate with our information proxies K_i. Finally, we expect that any remaining preference heterogeneity is likely to work against our main findings: if national brands are of higher quality and more informed households have a stronger preference for quality (physicians have if anything a greater taste for high-quality medicine, and chefs have if anything a greater taste for high-quality food), our estimates will tend to understate the effect of information on choice.

To describe the relationships among choice y_i, information K_i, household characteristics X_i and choice

[22] Past purchase experience may also serve as a proxy for a household's knowledge of the category. As past purchases are endogenous both to preferences and to the choice environment, we do not include this proxy in our main analysis. In appendix table 1 we show that our core findings are unchanged if we estimate specifications that control for average annual purchase volume. In these specifications, higher purchase volume is consistently associated with a statistically significant increase in the propensity to buy store brand.

[23] In our main specifications, we proxy for income using the categorical household income variable supplied by Nielsen. Appendix table 1 presents specifications that additionally control for average annual grocery spending and median occupational income.

environment Z_i, we will estimate linear probability models of the following form:

$$\Pr(y_i = 1 | K_i, X_i, Z_i) = \alpha + K_i\beta + X_i\gamma + Z_i\rho \tag{1}$$

where α, β, γ, and ρ are vectors of parameters.[24] Although for notational ease we have written the model at the level of the household, a given household can make multiple purchases. We therefore estimate the model at the level of the purchase occasion, reporting standard errors that allow for correlation at the level of the household, and weighting transactions by purchase volume. Appendix table 1 shows that our main conclusions are unaffected if we estimate binary logit models instead of linear probability models.

In sections 4 and 5 we present extensive descriptive evidence that variation in information across households affects brand choice. In section 6 we further parametrize Δv_i and $\Delta\tilde{v}_i$ and add an explicit model of price setting in order to quantify effects of information on consumer surplus and profits.

4 Case Studies

4.1 Headache Remedies

We begin our analysis with a case study of adult, non-migraine, daytime headache remedies. The first rows of table 1 show summary statistics for the six comparables in this category. These products span four active ingredients, each associated with a familiar national brand: aspirin (Bayer), acetaminophen (Tylenol), ibuprofen (Advil), and naproxen (Aleve). We estimate total annual expenditure on these comparables to be \$2.88 billion. Store-brand purchases account for 74 percent of pills and 53 percent of expenditures.[25]

On average, the per-pill price of a store brand is 40 percent of the price of a comparable national brand. For aspirin, a mature product that has been off patent since 1917, the per-pill price of store brands is 22 percent of the national-brand price. These price differences are not due to differences in where these products are sold or to volume discounts: among cases in our panel in which we observe at least one national-brand and one store-brand purchase for the same active ingredient and package size in the same market, chain, and week, the per-pill price paid for store brands is on average 26 percent of the price of an equivalent national brand. The median gap is 31 percent, and the national brand is cheaper in only 5 percent of cases.

Store-brand alternatives for national-brand headache remedies are widely available. Using our store-level data, we estimate that 85 percent of national-brand headache remedy purchase volume is purchased

[24] When we pool data across multiple comparables, we will allow the intercept α to differ by comparable.

[25] Among households with multiple headache remedy purchases, 31 percent bought only store brands and 16 percent bought only national brands. The remaining 52 percent bought both store brands and national brands.

when a store brand with the same active ingredient and form and at least as many pills is sold in the same store and year at a lower price. In our PanelViews survey data, only 3.6 percent of households report that no store-brand alternative was available at their last purchase.

In figure 1 we look at the relationship between knowledge of active ingredients and our indirect knowledge proxies—completed schooling, occupation, and college major. The relationships are as expected. Panel A shows that shoppers with a college education correctly identify the active ingredient in 62 percent of cases, as against 52 percent for those with a high school degree or less. Panel B shows that nurses correctly identify the active ingredient in 85 percent of cases, pharmacists in 89 percent, and physicians and surgeons in 90 percent. Panel C shows that shoppers whose college major is health or science related are more informed than other shoppers. In the online appendix, we confirm these relationships in a regression framework, showing that they remain strong even after controlling for a rich set of household characteristics, including income.

Having validated our proxies, we turn to our main question of interest: the impact of information on the share of purchases that go to store brands. Figure 2 shows that greater knowledge of active ingredients predicts more purchases of store brands. Those who can name no active ingredients buy just over 60 percent store brands. Those who can name all five active ingredients buy nearly 85 percent store brands. Though these differences are large, they could be due to reverse causality: those interested in saving money buy store brands and also take the time to read ingredient labels. We turn next to variation in information induced by exogenous household characteristics in part to alleviate this concern.

Figure 3 shows the relationship between store-brand share and completed schooling. With no controls, we see that those with education beyond high school buy more store brands than those with a high school degree or less, but that there is no clear difference between those with some college, a college degree, or more than a college degree. The main confound here is income, which is strongly negatively correlated with store-brand purchases (see appendix figure 1). After controlling for income, we find a monotonic positive relationship between completed schooling and store-brand share.

Figure 4 shows the relationship between store-brand share and occupation. Here we see a negative relationship between store-brand share and median occupational income among non-healthcare occupations. Households whose primary shopper is a healthcare professional buy far more store brands than others of similar income. Pharmacists, physicians, and nurses buy more store brands than lawyers, who have high levels of schooling but different occupational expertise.

Pharmacists, who stand out in the survey data in figure 1 as among the most informed about active ingredients, also stand out for having the largest store-brand share among large healthcare occupations.

Only 8.5 percent of volume bought by pharmacists are national-brand headache remedies, an amount small enough to be explained by the occasional stock outs of store brands, and the fact that some purchases are made by the non-pharmacist member of a pharmacist's household.[26]

Table 2 presents the relationship between store-brand share and knowledge of active ingredients in a regression framework. The table presents estimates of equation 1, where the information variables of interest K_i are the share of active ingredients known and an indicator for college education. All specifications allow the intercept α to differ by comparable. Columns (1) and (2) include in Z_i market and calendar quarter fixed effects; column (3) adds interacted indicators for the market, chain, and calendar quarter. Column (1) includes in X_i controls for demographic characteristics other than income; column (2) adds income controls. In the preferred specification, column (3), college education increases the propensity to buy store brand by 2.6 percentage points, and going from knowledge of no active ingredients to knowledge of all increases the store-brand share by 19 percentage points. The estimated effect of education gets larger when income controls are added; the effect of active ingredient knowledge is fairly stable across specifications.

Column (4) of table 2 augments the specification in column (3) by adding to K_i an indicator for whether the shopper reports that store brands are "just as safe" as national brands. This is a less convincing measure of information than active ingredient knowledge, as the correct answer is arguably unclear. Still, it is worth noting that it is a very strong correlate of brand choice: believing store brands are just as safe as national brands has an additional effect of 21 percentage points over and above the effect of active ingredient knowledge. The effect of having this belief *and* being able to name all active ingredients correctly is 35 percentage points.

Table 3 presents regression evidence on the effect of occupation. The model and controls in the first three columns are the same as in table 2, but now the vector K_i of information proxies consists of an indicator for college education, an indicator for being a pharmacist or physician, and an indicator for being in a healthcare occupation other than pharmacist or physician. The estimated occupation effects remain stable as we add

[26]The fact that 8.5 percent of purchases by households whose primary shopper is a pharmacist are to national-brand goods suggests at first that 8.5 percent of the time a pharmacist is willing to pay a significant price premium to buy a national brand.

There are three main reasons to interpret the finding differently.

First, the primary shopper need not be the only shopper in the household. In the small number of cases (12 households, 37 transactions) in which a household with both a primary shopper and a secondary shopper who are pharmacists buy a headache remedy, only 1.6 percent of purchases are to national brands. In the case of single-person households in which the only person is a pharmacist (22 households, 109 transactions), only 5 percent of purchases are to national brands.

Second, although we have focused on transactions in retailers who stock both national brands and store brands, some stockouts may nevertheless occur. Matsa (2011) estimates the stockout rate for over-the-counter drugs to be 2.8 percent. In the face of a stockout of the store brand, pharmacists who are unable to delay their purchase may switch to buying a national-brand good.

Third, although the average price premium for national brands is very large in this category, there is some price variation, and pharmacists may be buying when the price difference is unusually small. In the Homescan data, we find that the ratio of the average store-brand price to the average national-brand price is 6 percent greater when we focus on purchases by households whose primary shopper is a pharmacist, and 14 percent greater when we focuses on cases where the only person in the household is a pharmacist.

controls. In the preferred specification of column (3) we find that being a pharmacist or physician increases the propensity to buy store brands by 15 percentage points; being in another healthcare occupation increases the propensity by 8 percentage points.

Column (4) of table 3 presents evidence on the role of college major. We restrict the sample to respondents who completed college and who reported their college major in our survey. We find that non-health science majors are 5 percentage points more likely to buy store brand. Column (5) of table 3 presents occupation results for the subsample of respondents who are not currently employed for pay. (Recall that our occupation variables are defined based on the most recent employment spell.) The coefficients on the occupation indicators remain large in magnitude and statistically significant, though less precisely estimated than in the full sample. Taken together, columns (4) and (5) suggest our results are unlikely to be driven by factors specific to current employment in a healthcare profession, such as the availability of employee discounts or free samples. As further evidence, in the online appendix we use data from the Bureau of Labor Statistics to show that the propensity to buy store brand is greater among shoppers whose occupations require medical knowledge. This holds true even if we exclude shoppers who we have classified as having occupations in healthcare.

Table 4 presents evidence on the extent to which our direct and indirect knowledge measures capture the same underlying variation. Column (1) repeats the preferred specification of table 3 column (3), this time restricting to respondents who participated in the wave of our survey in which we assessed active ingredient knowledge. Column (2) restricts the sample to shoppers who named all active ingredients correctly. Column (3) adds the additional restriction that the respondent believes store brands are "just as safe" as national brands. Restricting attention to well-informed consumers reduces the estimated effect of education and occupation substantially, while only slightly reducing precision. In the final column, the occupation coefficients are reduced by more than 70 percent and are statistically indistinguishable from zero. These findings are consistent with the interpretation that all of our measures capture variation along a common dimension, which we interpret as information.

As further support for our identifying assumptions, appendix figure 2 shows that healthcare professionals and non-healthcare professionals look similar in their choices over observed product attributes such as active ingredient and physical form. Appendix figures 3 and 4 show similar results for average annual purchase volume and item size, respectively.

4.2 Pantry Staples

We now turn to the analysis of food purchases. Here our proxies for knowledge are indicators for whether the primary shopper is a chef ("chef or head cook") or other food preparer.[27] We begin with a case study of pantry staples: salt, sugar, and baking soda. We choose these products because they are uniform in chemical composition and purpose, and thus analogous to headache remedies in being relatively homogeneous.

The lower portion of table 1 includes summary statistics for the six comparables we classify as pantry staples: baking soda; regular iodized and plain salt (sold in boxes); and regular granulated, light brown, and powdered sugar (sold in bags). Collectively, these comparables account for \$1.81 billion of expenditure. Store-brand purchases account for 60 percent of volume and 57 percent of expenditure. On average, the ratio of store-brand to national-brand price per equivalent volume is 0.92, with a range from 0.75 (plain salt) to 0.92 (granulated sugar).

Figure 5 shows the relationship between store-brand share and occupation. As with headache remedies, there is a strong negative relationship between store-brand share and median occupational income. Households whose primary shopper is a food preparer or food service manager buy more store brands than others of similar occupational income. Chefs—the occupational group we would have expected *ex ante* to be most informed about the quality of food products—buy more than 77 percent store brands in these categories, more than any other occupation of meaningful size.

Table 5 shows the relationship with occupation in a regression framework. The specifications in the five columns are the same as in table 3, with the information proxies of interest K_i now consisting of an indicator for college education, an indicator for being a chef, and an indicator for being a food preparer but not a chef. In our preferred specification of column (3), we estimate that being a chef increases the probability of buying store brands by 12 percentage points, and working in a non-chef food preparation occupation increases this probability by 2 percentage points. These effects are somewhat smaller in magnitude than those we estimate when we do not include our preferred set of controls. In contrast to headache remedies, we do not find any clear effect of college education. Column (4) shows that non-health science majors and health majors are not statistically different from other college graduates. Column (5) shows that the coefficient on being a chef goes largely unchanged when we focus on shoppers who are not currently employed. The coefficient on being a non-chef food preparer falls and becomes statistically insignificant, but its confidence interval

[27] Our second survey wave asked respondents to identify the most common additive to table salt (iodine), the scientific name for baking soda (sodium bicarbonate), and the most common ingredient of granulated sugar (sucrose). The share of these questions answered correctly is positively correlated with working as a chef but not with being a non-chef food preparer, and is positively correlated (but not statistically significantly so) with the propensity to buy store-brand pantry staples. Results for these knowledge measures are presented in the online appendix.

includes the magnitude of our preferred estimate. These findings suggest that the effects we estimate are not driven by mechanical effects of employment in the food industry.

4.3 Evidence on Domain Specificity

We find that health experts purchase more store-brand health products and that food experts purchase more store-brand food products. A natural follow-up question is to what extent experts' knowledge is transferable outside of their domain of expertise. Perhaps pharmacists' understanding of the equivalence of national-brand and store-brand headache remedies leads them to also recognize the likely equivalence of national-brand and store-brand baking soda. Or perhaps their understanding does not translate beyond the categories with which they are directly familiar.

Table 6 presents evidence on domain specificity. The first two columns look at the effect of healthcare expertise on pantry staple purchases. Column (1) shows that the share of headache remedy active ingredients known has no significant effect on the probability of purchasing store-brand pantry staples, with a confidence interval that rules out effects greater than 1.2 percentage points. Column (2) shows that pharmacists, physicians, and other healthcare professionals are also not significantly more likely to buy store-brand pantry staples. The confidence intervals on the pharmacist-physician and other healthcare occupation coefficients rule out effects greater than 5.2 percentage points and 2.2 percentage points respectively. We can confidently reject the hypothesis that these effects are as large as the effects we estimate for headache remedy purchases. The evidence thus suggests that healthcare expertise does not translate to behavior outside the health domain, consistent with past evidence on the domain specificity of expertise (Levitt et al. 2010).

The final column of table 6 looks at the effect of food preparation expertise on headache remedy purchases. Here, we do see some evidence of transferability: chefs are a statistically significant 11 percentage points more likely to buy store-brand headache remedies than other consumers. There is no significant effect for food preparers other than chefs.

5 Cross-category Comparisons

5.1 Health Products

We turn next to analyzing a broad set of health products. We restrict attention to the 6 headache remedy comparables that we study above, and 44 additional comparables for which we observe at least $5,000$ purchases by households with non-missing values of our demographic controls. These include other medications such as cold remedies, first aid products such as bandages, and miscellaneous products such as vitamins and contact lens solution. Non-painkiller health categories account for \$8.94 billion of expenditure per year. Store-brand purchases account for 56 percent of volume. Store-brand prices are half of national-brand prices on average.

For each comparable, we run one regression to estimate the effect of knowing headache remedy active ingredients (using the specification in column (3) of table 2) and one to estimate the effect of occupation (using the specification in column (3) of table 3). Figures 6, 7, and 8 present coefficients on these information proxies along with 95 percent confidence intervals.[28] We present an analogous plot for the coefficients on college education in the online appendix. In order to test joint hypotheses about the coefficients in these plots, we conduct 10 bootstrap replications of our estimates. In each bootstrap we draw a random subset of households with replacement.

Figure 6 shows that the coefficient on active ingredient knowledge is positive in 43 out of 50 cases. The share of coefficient estimates that are positive is thus 0.86, which has a bootstrap standard error of 0.04, and is therefore highly statistically distinguishable from the null hypothesis of no effect (half of coefficients positive). Consistent with the evidence on domain specificity that we present above, if we estimate analogous models for non-health comparables, the coefficient on active ingredient knowledge is positive in only 168 out of 282 cases, which is much closer to the null hypothesis and highly statistically distinguishable from the number for health categories. Figure 9 illustrates the contrast visually, plotting the distribution of t-statistics separately for health and non-health comparables.

The differences among the coefficients in figure 6 are instructive. The coefficients tend to be larger and more significant for medications and relatively smaller for first aid and eye care products, suggesting that in the latter group informed shoppers perceive true quality differences. Indeed, contact lens solutions are the only healthcare product we have identified where some medical professionals recommend patients buy national brands due to quality concerns with store brands (Secor 2002). In the online appendix, we show that the estimated effects of information proxies tend to be larger (though not statistically significantly so) in comparables in which Consumer Reports considers store brands and national brands to be equivalent. We also examine whether the effect of information is greater in the comparables for which the price gap between national and store brands is greatest. Finally, we show that the effect of information tends to be greater in comparables in which advertising is more intensive, consistent with the idea that perceptions of product quality by the uninformed may be driven by advertising on the part of national-brand manufacturers.

Figures 7 and 8 present coefficients for the effect of being a pharmacist or physician and the effect

[28] Although knowledge of headache remedy active ingredients is obviously most relevant to headache remedy purchases, we expect it to also be a good proxy for more general knowledge relevant to the other health categories.

of other healthcare professions respectively. We see broadly similar patterns to the coefficients on active ingredient knowledge, though with somewhat less precision. The effect of being a pharmacist or physician is positive in share 0.68 of cases (bootstrap standard error = 0.05), and the effect of being in another healthcare occupation is positive in share 0.86 of cases (bootstrap standard error = 0.04). In the online appendix we present plots analogous to figure 9 for these two sets of coefficients.

5.2 Food and Drink Products

Next we consider the remaining food and drink comparables in our data. We restrict attention to the 6 pantry staples that we study above, plus 235 additional comparables for which we observe at least $5,000$ purchases by households with non-missing values of our demographic controls. These comprise a broad cross-section of supermarket products, from milk and eggs, to carbonated beverages, to ready-to-eat cereal. Excluding pantry staples, these categories account for $123 billion of expenditure. Store-brand purchases account for 43 percent of volume. On average, the price-per-equivalent-volume for store brands is 69 percent of that for national brands.

For each comparable, we run a separate regression to estimate the effect of working as a chef or other food preparer on store-brand purchases (using the specification in column (3) of table 5). Figure 10 summarizes the estimated coefficients and 95 percent confidence intervals. Rather than try to present all coefficients in a single figure, we aggregate comparables other than pantry staples into what Nielsen calls "product groups," weighting the individual comparables by precision and computing the aggregate confidence interval as if the individual coefficients are statistically independent. Thus, for example, the comparables for cola, diet cola, lemon-lime soda, and so forth are combined into the Nielsen product group "carbonated beverages."

The estimated effects of knowledge on store-brand purchases in these categories are weaker than what we saw for health products. The coefficients on working as a chef are positive for 148 comparables and negative for 93. The share of coefficient estimates that are positive is thus 0.61, with a bootstrap standard error of 0.04. The coefficients that are individually statistically significant are generally small in magnitude. The pantry staples categories stand out as having among the most positive and significant coefficients: granulated sugar has the third largest coefficient in the figure, and three of the top six coefficients are pantry staples. In the online appendix we present plots analogous to figure 10 for working in other food preparation occupations and for having a college education.

6 Aggregate Effects of Consumer Information

In this section we view our data through the lens of a stylized model of household demand for brands and price-setting by manufacturers and retailers. We combine the estimated coefficients from the preceding analysis with additional data moments to estimate the model. Using the estimated model, we compute the effect of consumer information on the distribution of consumer and producer surplus, and on prices and market shares.

The purpose of this analysis is twofold. First, we wish to aggregate the coefficients estimated in section 5 to learn how expenditures, market shares, and profits would change in the drug store and the grocery store if all households behaved like expert shoppers. This aggregation does not rely on details of the model: it amounts to an expenditure-weighted aggregation of the coefficients presented in figures 7 and 10, along with information on retail and wholesale prices.

Second, we wish to predict how consumer welfare and firm pricing would change in a world of informed shoppers. Our predictions are contingent on a set of strong parametric, symmetry, and conduct assumptions. These assumptions allow us to solve the model in closed form for a large set of product categories, and to show transparently how the various empirical moments determine our estimates. Because the resulting model is highly stylized, our welfare and pricing results should be taken more as suggestive illustrations of the economic forces at work than as realistic empirical predictions.

6.1 Model

For each comparable, consider a market with R retailers indexed by r and households indexed by i. Each retailer sells a store brand with price $p(0,r)$ and a national brand with price $p(1,r)$. Each household must make a single purchase from the choice set $\{0,1\} \times \{1,...,R\}$. Both the store brand and the national brand are manufactured at constant marginal cost c. A single manufacturer captures all profits from the sale of the national brand. Each retailer captures profits from the sale of its own store brand. The market consists of a large number of uninformed households—which we define as consumers who are not pharmacists or physicians for health products and consumers who are not chefs for food products—as well as a small number of informed households. We assume the latter are few enough that firms ignore them in making pricing decisions.

Each household maximizes utility $u_i(b,r)$ given by

$$u_i(b,r) = v_i(b) - p(b,r) + \tau_i(r), \tag{2}$$

where $b \in \{0,1\}$ is an indicator for purchasing the national brand, $v_i(b)$ is an idiosyncratic perceived brand preference, and $\tau_i(r)$ is an idiosyncratic travel cost distributed standard type-I extreme value up to a scale parameter. Each household has a true brand preference $\tilde{v}_i(b)$.

We specify brand preferences as follows. We normalize $v_i(0) = \tilde{v}_i(0) = 0$. For each household, we let $\tilde{v}_i(1) = \lambda \xi_i$ where λ is a parameter and ξ_i is a preference shock distributed i.i.d. logistic across households. For uninformed households, $v_i(1) = \xi_i$; for informed households, $v_i(1) = \tilde{v}_i(1)$.

The parameter $\lambda \geq 0$ indicates the similarity between true and perceived brand preference for uninformed households. When $\lambda = 1$, perceived and true brand preference agree; when $\lambda = 0$, national and store brand are truly identical but are perceived to be different. Throughout our analysis, we define consumer welfare with respect to true brand preference.

The game proceeds in three stages. First, the manufacturer and retailers simultaneously announce all prices $p(b,r)$. Second, each household learns its travel cost $\tau_i(r)$ and chooses which retailer r to visit. Third, each household learns its perceived brand preference $v_i(b)$ and chooses which brand b to purchase. We restrict our attention to a symmetric equilibrium in which $p(0,r) = p(0)$, and hence $p(1,r) = p(1)$, for each retailer r.

6.2 Estimation

Estimation is in closed form. Here we outline the key steps; an appendix provides additional details.

We match $p(0)$ and $p(1)$ to the average store-brand and national-brand prices, respectively, and we choose c to match the median retail margin of store brands. We choose the scale of $\tau_i(r)$ to match the retailer's markup on the store brand: greater dispersion in $\tau_i(r)$ implies less competition among retailers and hence greater retail margins. Similarly, we choose the scale of ξ_i to match the manufacturer's markup on the national brand. Given scale parameters, we can then choose the location of ξ_i to match the overall market share of the national brand: a high market share for the national brand implies a high mean value of ξ_i.

Having pinned down the preferences of the uninformed, we choose λ to match the difference in store-brand purchase probability between informed and uninformed consumers shown in figures 7 and 10. When informed households purchase more store brand than uninformed households, $\lambda < 1$. When informed household purchase more national brands than uninformed households, $\lambda > 1$.

The online appendix presents point estimates for all parameters for all comparables, with bootstrapped standard errors.

6.3 Results

Tables 7 and 8 present summaries of our findings, aggregated across groups of comparables, for health and food products, respectively. For each set of products we present the change relative to baseline from two counterfactuals in which households choose according to true rather than perceived brand preference. In the first counterfactual, prices are held constant at observed levels; in the second, prices adjust to reflect the change in consumer demand. We measure changes in consumer expenditure and surplus, and changes in retailer and manufacturer profit, relative to baseline expenditure levels.

The left panel of table 7 presents results for headache remedies. Holding prices constant at baseline levels, if all consumers became as informed as pharmacists or physicians, the market share of national-brand headache remedies would fall by half, total expenditure on headache remedies would fall by 14 percent, and consumer surplus would increase by 4 percent relative to baseline expenditure. The national-brand manufacturer would lose profits equivalent to 19 percent of baseline expenditure, and retailers would gain profits equivalent to 5 percent. Note that total surplus falls even though we evaluate consumer welfare with respect to true preferences. The reason is that prices do not equal marginal costs; hence improvements in consumer information necessarily improve consumer welfare but do not necessarily improve social surplus.

Allowing prices to adjust softens the blow for the national-brand manufacturer by allowing the manufacturer to lower the relative price of the national brand. This harms retailers but increases the gains to the consumer. Because prices come to better reflect manufacturing costs, total surplus rises relative to the case in which prices are held constant, and there is no aggregate efficiency loss relative to baseline.

The right panel of table 7 shows that for other health categories we find effects that are similar directionally to those for headache remedies, smaller in magnitude, and still economically significant. Allowing for price adjustment, consumers would gain surplus equivalent to 4 percent of baseline expenditure in health categories other than headache remedies, were they to choose according to their true preferences. Additional results presented in the online appendix show that much of this gain would come from (non-headache-remedy) medication categories.

Table 8 examines food and drink categories. Here, the small price differences between national and store brands and the relatively modest effects of information combine to imply fairly small impacts. The greatest effect is found in pantry staples, where allowing for both price adjustment and greater consumer information would improve consumer welfare by an amount equal to 3 percent of baseline expenditure.

7 Conclusions

Across a range of products we find strong evidence that more informed shoppers buy more store brands and fewer national brands. Consumer information plays a large quantitative role in health categories, where our estimates imply that expenditures and market shares would change significantly if all households behaved like expert shoppers. By contrast, the role of consumer information is smaller in food and drink categories, where our estimates suggest much smaller gaps between expert and non-expert shopping behavior.

Our study is limited to examining the effects of information on quantities and prices. If consumers were to become more informed, markets would adjust on other margins as well. In particular, a more informed population of consumers might change whether and how much firms choose to advertise their products, as well as which products are introduced to the market. Taking account of these forms of dynamic adjustment, and examining their implications for welfare, is an important priority for future work.

Appendix

Details of Model Estimation and Computation

We estimate parameters separately for each comparable product group. Let ξ_i be distributed logistic with location parameter μ and scale parameter σ_{brand}. Define σ_{retail} so that $\tau_i(r)/\sigma_{\text{retail}}$ is distributed standard type I extreme value. The parameters to be estimated are $\{\mu, \sigma_{\text{brand}}, \sigma_{\text{retail}}, R, \lambda\}$.

Let S be the population market share of the store brand for uninformed households. From the properties of the logistic distribution, it is immediate that

$$S = \text{logit}^{-1}\left((\Delta p(r) - \mu)/\sigma_{\text{brand}}\right), \tag{A1}$$

where $\Delta p \equiv p(1) - p(0)$.

Begin with estimation of μ and σ_{brand}. It is possible to show that in a symmetric interior equilibrium the manufacturer's first-order condition is

$$p(1) - c = (1-S)/\frac{dS}{dp(1)}, \tag{A2}$$

where

$$\frac{dS}{dp(1)} = \frac{S(1-S)}{\sigma_{\text{brand}}}. \tag{A3}$$

Given $p(0)$, $p(1)$, and c, equations (A1), (A2), and (A3) imply unique values of μ and σ_{brand} for a given S. We estimate μ and σ_{brand} by substituting the sample analogue of S into the resulting expressions.

Turn next to estimation of σ_{retail} and R. These are not separately identified but for our purposes it is sufficient to identify $\tilde{\sigma}_{\text{retail}} \equiv \frac{R}{R-1}\sigma_{\text{retail}}$. To do this we observe that in a symmetric interior equilibrium the price of the store brand must satisfy

$$p(0) - c = \left[\frac{S}{\tilde{\sigma}_{\text{retail}}} + \frac{dS}{dp(1)}\frac{1}{S}\right]^{-1}. \tag{A4}$$

Given $p(0)$, $p(1)$, c, equations (A2) and (A4) define a unique $\tilde{\sigma}_{\text{retail}}$ as a function of S. We estimate $\tilde{\sigma}_{\text{retail}}$ by substituting the sample analogue of S into the resulting expression.

The final parameter to estimate is λ. Let S_λ be the population market share of the store brand for informed households:

$$S_{\lambda} \quad = \quad \text{logit}^{-1}\left(\left(\Delta p\left(r\right)/\lambda-\mu\right)/\sigma_{\text{brand}}\right) \tag{A5}$$

It follows that:

$$\lambda = \frac{\Delta p}{\sigma_{\text{brand}}\left(\text{logit}\left(S_{\lambda}\right)-\text{logit}\left(S\right)\right)+\Delta p}. \tag{A6}$$

We estimate λ by substituting sample analogues of S and S_{λ} into this expression.

We use the linear probability models reported in figures 7 (for health categories) and 10 (for food categories) to define the sample analogues of S_{λ} and S. We denote these as $\hat{S}_{\lambda}$ and $\hat{S}$, respectively. We define an expert to be a pharmacist or physician for health categories and a chef for non-health categories. We define $\hat{S}_{\lambda}$ to be the mean predicted probability of choosing store brand if each purchaser i were an expert with the average expert's level of education and the purchaser's own demographics X_i and choice environment Z_i. We define $\hat{S}$ so that the average of $\hat{S}_{\lambda}$ and $\hat{S}$, weighted by the sample shares of experts and non-experts, is equal to the overall share choosing store brand.

A few exceptional cases are worth noting. When we do not observe the retail margin (or it is estimated to be negative), we use the expenditure-weighted average retail margin across other comparables in the same group (health/food). When our linear probability model implies that $\hat{S}_{\lambda} \geq 1$, we impute $\lambda = 0$. When our linear probability model implies that $\hat{S}_{\lambda} \leq 0$, or when no value of $\lambda \in \left[0, \overline{\lambda}\right]$ explains $\hat{S}_{\lambda}$, we set λ equal to an upper bound $\overline{\lambda}$. We use the threshold $\overline{\lambda} = 3$ in our estimates. Finally, when no value of $\tilde{\sigma}_{\text{retail}}$ solves equation (A4), we assume in computing counterfactuals that prices are fixed at $p\left(0\right)$ and $p\left(1\right)$. We summarize the frequency of these cases in the online appendix.

To compute counterfactual prices under informed choice, we solve equations (A2) and (A4) numerically assuming that demand is governed by informed rather than uninformed preferences. Exact expressions for the change in consumer welfare under informed choice are readily derived from the assumed preference structure.

References

Allison, Ralph I. and Kenneth P. Uhl. 1964. Influence of beer brand identification on taste perception. *Journal of Marketing Research* 1(3): 36-39.

Associated Press. 2006. Contaminated with metal, a painkiller is recalled. November 10. Accessed at http://www.nytimes.com/2006/11/10/health/10recall.html on March 29, 2013.

Bartels, Larry M. 1996. Uninformed votes: Information effects in presidential elections. *American Journal of Political Science* 40(1): 194-230.

Becker, Gary S. and Kevin M. Murphy. 1993. A simple theory of advertising as a good or bad. *The Quarterly Journal of Economics* 108(4): 941-964.

Bergès, Fabian, Daniel Hassen, Sylvette Monier-Dilhan, and Hélène Raynal. 2009. Consumers' decision between private labels and national brands in a retailer's chain: A mixed multinomial logit application. *Gestion 2000* 3: 41-57.

BLS. 2002. *Census Occupation Classification*. Accessed at http://www.bls.gov/cps/cenocc.pdf in 2010.

Braithwaite, Dorothea. 1928. The economic effects of advertisement. *The Economic Journal* 38(149): 16-37.

Branthwaite, Alan and Peter Cooper. 1981. Analgesic effects of branding in treatment of headaches. *British Medical Journal* 282(6276): 1576-1578.

Bronnenberg, Bart J., Jean-Pierre Dubé, and Matthew Gentzkow. 2012. The evolution of brand preferences: Evidence from consumer packaged goods. *American Economic Review* 102(6): 2472-2508.

Burton, Scot, Donald R. Lichtenstein, Richard G. Netemeyer, and Judith A. Garretson. 1998. A scale for measuring attitude toward private label products and an examination of its psychological and behavioral correlates. *Journal of the Academy of Marketing Science* 26(4): 293-306.

Carrera, Mariana and Sofia Villas-Boas. 2013. Generic aversion and observational learning in the over-the-counter drug market. Case Western mimeo.

Caves, Richard E. and David P. Greene. 1996. Brands' quality levels, prices, and advertising outlays: Empirical evidence on signals and information costs. *International Journal of Industrial Organization* 14(1): 29-52.

Cox, Steven R., Kenneth A. Coney, and Peter F. Ruppe. 1983. The impact of comparative product ingredient information. *Journal of Public Policy & Marketing* 2, 57-69.

CVS. 2013. *CVS Shop*. Accessed at http://www.cvs.com/shop/product-detail/Bayer-Aspirin-Tablets-Easy-Open-Cap?skuId=100073 on March 29, 2013.

Darden, William R. and Roy D. Howell. 1987. Socialization effects of retail work experience on shopping orientations. *Journal of the Academy of Marketing Science* 15(3): 52-63.

Dick, Alan, Arun Jain, and Paul Richardson. 1995. Correlates of store brand proneness: Some empirical observations. *Journal of Product & Brand Management* 4(4): 15-22.

Domenighetti, Gianfranco, Michele Tomamichel, Felix Gutzwiller, Silvio Berthoud, and Antoine Casabianca. 1991. Psychoactive drug use among medical doctors is higher than in the general population. *Social Science & Medicine* 33(3): 269-274.

Einav, Liran, Ephraim Leibtag, and Aviv Nevo. 2010. Recording discrepancies in Nielsen Homescan data: Are they present and do they matter? *Quantitative Marketing and Economics* 8(2): 207-239.

Ellison, Glenn and Alexander Wolitzky. 2012. A search cost model of obfuscation. *RAND Journal of Economics* 43(3): 417-441.

Endsley, Scott, Geof Baker, Bernard A. Kershner, and Kathleen Curtin. 2006. What family physicians need to know about pay for performance. *Family Practice Management* 13(7): 69-74.

Federal Trade Commission. 1999. *Final Order In the Matter of Novartis Corporation, et al., Docket 9279*. Accessed at http://www.ftc.gov/os/decisions/docs/Volume127.pdf on January 18, 2013.

Gabaix, Xavier and David Laibson. 2006. Shrouded attributes, consumer myopia, and information suppression in competitive markets. *The Quarterly Journal of Economics* 121(2): 505-540.

Glanz, Karen, Stanley B. Fiel, Lawrence R. Walker, and Marvin R. Levy. 1982. Preventive health behavior of physicians. *Journal of Medical Education* 57(8): 637-639.

Goldfarb, Avi, Qiang Lu, and Sridhar Moorthy. 2009. Measuring brand value in an equilibrium framework. *Marketing Science* 28(1): 69-86.

Haas, Jennifer S., Kathryn A. Phillips, Eric P. Gerstenberger, and Andrew C. Seger. 2005. Potential savings from substituting generic drugs for brand-name drugs: Medicare expenditure panel survey, 1997-2000. *Annals of Internal Medicine.* 142(11): 891-897.

Hortaçsu, Ali and Chad Syverson. 2004. Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *The Quarterly Journal of Economics* 119(2): 403-456.

Husband, Richard W. and Jane Godfrey. 1934. An experimental study of cigarette identification. *Journal of Applied Psychology* 18(2): 220-223.

Huskamp, Haiden A., Patricia A. Deverka, Arnold M. Epstein, Robert S. Epstein, Kimberly A. McGuigan, and Richard G. Frank. 2003. The effect of incentive-based formularies on prescription-drug utilization and spending. *New England Journal of Medicine* 349(23): 2224-2232.

Johnson, Erin M. and M. Marit Rehavi. 2013. Physicians treating physicians: Information and incentives in childbirth. NBER Working Paper No. 19242.

Kamenica, Emir, Robert Naclerio, and Anup Malani. 2013. Advertisements impact the physiological efficacy of a branded drug. *Proceedings of the National Academy of Sciences* 110(32): 12931-12935.

Kantar Media. 2013. Kantar Media reports U.S. advertising expenditures increased 3 percent in 2012. March 11. Accessed at http://kantarmediana.com/intelligence/press/us-advertising-expenditures-increased-3-percent-2012 on April 23, 2013.

Kumar, Nirmalya and Jan-Benedict E.M. Steenkamp. 2007. *Private Label Strategy: How to Meet the Store Brand Challenge*. Cambridge, MA: Harvard Business School Press.

Levitt, Steven D., John A. List, and David H. Reiley. 2010. What happens in the field stays in the field: Exploring whether professionals play minimax in laboratory experiments. *Econometrica* 78(4): 1413-1434.

Levitt, Steven D. and Chad Syverson. 2008. Market distortions when agents are better informed: The value of information in real estate transactions. *Review of Economics and Statistics* 90(4): 599-611.

Matsa, David A. 2011. Competition and Product Quality in the Supermarket Industry. *The Quarterly Journal of Economics* 126(3): 1539-1591.

Mason, Kevin and Alice Batch. 2009. What's in a (brand) name? *Proceedings of American Society of Business and Behavioral Sciences* 16(1).

National Institute for Health Care Management (NIHCM). 2002. *A Primer: Generic Drugs, Patents, and the Pharmaceutical Marketplace*. Washington DC: National Institute for Health Care Management Foundation.

Piccione, Michele and Ran Spiegler. 2012. Price competition under limited comparability. *The Quarterly Journal of Economics* 127(1): 97-135.

Pronko, Nicholas H. and J.W. Bowles Jr. 1949. Identification of cola beverages. III. A final study. *Journal of Applied Psychology* 33(6): 605-608.

Richardson, Paul S., Arun K. Jain, and Alan Dick. 1996. Household store brand proneness: A framework. *Journal of Retailing* 72(2): 159-185.

Scherer, Frederic M. 1970. *Industrial Market Structure and Economic Performance*. Chicago, IL: Rand McNally.

Secor, Glenda B. 2002. Educate patients about inconsistency of generic contact lens solutions. *Primary Care Optometry News*. May. Accessed at http://www.healio.com/optometry/contact-lenses-eyewear/news/print/primary-care-optometry-news/%7Ba6af06d9-ffe3-4a21-88a2-10f963172c5a%7D/educate-patients-about-inconsistency-of-generic-contact-lens-solutions on May 24, 2013.

Sethuraman, Raj and Catherine Cole. 1999. Factors influencing the price premiums that consumers pay for national brands over store brands. *Journal of Product & Brand Management* 8(4): 340-351.

Shrank, William H., Emily R. Cox, Michael A. Fischer, Jyostna Mehta, and Niteesh K. Choudhry. 2009. Patients' perceptions of generic medications. *Health Affairs* 28(2): 546-556.

Simons, Henry. 1948. *Economic Policy for a Free Society*. Chicago, IL: University of Chicago Press.

Smith, Michael D. and Erik Brynjolfsson. 2001. Consumer decision-making at an Internet shopbot: Brand still matters. *Journal of Industrial Economics* 49(4): 541-558.

Steenkamp, Jan-Benedict E.M., Harold J. van Heerde, and Inge Geyskens. 2010. What makes consumers willing to pay a premium for national brands over private labels. *Journal of Marketing Research* 47(6): 1011-1024.

Sullivan, Donald L., Stephen W. Birdwell, and Suzan N. Kucukarslan. 1994. An assessment of consumer purchasing behavior for private-label vs. brand-name over-the-counter products in chain pharmacies. *Journal of Pharmaceutical Marketing & Management* 8(1): 85-108.

Sullivan, Mary W. 1998. How brand names affect the demand for twin automobiles. *Journal of Marketing Research* 35(2): 154-165.

Szymanski, David M. and Paul S. Busch. 1987. Identifying the generics-prone consumer: A meta-analysis. *Journal of Marketing Research* 24(4): 425-431.

Thomas, Katie. 2012. An increase in scrutiny for generics. *New York Times*. October 31. Accessed at http://www.nytimes.com/2012/11/01/business/fda-increases-scrutiny-of-some-generic-drugs.html on March 29, 2013.

U.S. Census. 2000. Earnings by occupation and education, 1999. Accessed at http://www.census.gov/hhes/www/income/earnings/earnings.html on December 17, 2009.

U.S. Census. 2013. Monthly retail trade report: Estimates of monthly retail and food services sales by kind of business: 2008. Accessed at http://www.census.gov/retail/ on March 31, 2013.

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics (NCES). 2012. CIP user site. Accessed at http://nces.ed.gov/ipeds/cipcode/Default.aspx?y=55 on Decemeber 28, 2012.

Wills, Robert L. and Willard F. Mueller. 1989. Brand pricing and advertising. *Southern Economic Journal* 56(2): 383-395.

Table 1: Summary statistics

	Total expenditure ($bn / year)	Store-brand share (volume)	Store-brand share ($)	Price ratio (store brand / national brand)
Headache remedies				
Acetaminophen gelcaps	$0.39	0.51	0.38	0.58
Ibuprofen gelcaps	$0.50	0.29	0.22	0.69
Acetaminophen tablets	$0.44	0.81	0.60	0.36
Aspirin tablets	$0.24	0.75	0.40	0.22
Ibuprofen tablets	$0.94	0.81	0.61	0.36
Naproxen sodium tablets	$0.37	0.57	0.44	0.61
Total (6)	$2.88	0.74	0.53	0.40
Other health products (82)	$10.87	0.58	0.47	0.54
Pantry staples				
Baking soda	$0.14	0.33	0.27	0.75
Salt (iodized)	$0.07	0.53	0.47	0.76
Salt (plain)	$0.04	0.47	0.40	0.75
Sugar (brown)	$0.17	0.70	0.65	0.81
Sugar (granulated)	$1.27	0.60	0.59	0.92
Sugar (powdered)	$0.13	0.72	0.70	0.88
Total (6)	$1.81	0.60	0.57	0.92
Other food & drink products (256)	$134.90	0.39	0.33	0.71
Remaining products (70)	$45.05	0.26	0.20	0.58

Notes: Total expenditure is 2008 annual expenditure in all grocery, drug, and mass merchandise stores in the US, estimated as described in section 2.4. Store-brand share (volume) is the share of equivalent quantity units (pills for headache remedies, pounds for pantry staples) in each comparable devoted to store brands in our sample of the Nielsen Homescan Panel. Store-brand share ($) is the share of expenditure devoted to store brands in our sample of the Nielsen Homescan Panel. Price ratio is the average price per equivalent quantity unit observed in the Nielsen RMS data for store brands divided by the analogous average price for national brands. Rows for "headache remedies" and "pantry staples" each correspond to a single comparable product group. Rows for "other health products," "other food & drink products," and "remaining products" aggregate over multiple comparable product groups, with the number of such groups shown in parentheses. In columns two through four, these aggregates average over comparable product groups weighting by expenditure, except for headache remedies, where we weight by number of pills.

Table 2: Knowledge and headache remedy purchases

Dependent variable: Purchase is a store brand

Primary shopper characteristics:	(1)	(2)	(3)	(4)
College education	0.0094	0.0212	0.0255	0.0214
	(0.0072)	(0.0075)	(0.0073)	(0.0068)
Share of active ingredients known	0.1792	0.1805	0.1898	0.1463
	(0.0111)	(0.0111)	(0.0108)	(0.0105)
Believe store brands are "just as safe"				0.2058
				(0.0070)
Demographic controls?	X	X	X	X
Market & quarter fixed effects?	X	X		
Income controls?		X	X	X
Market-chain-quarter fixed effects?			X	X
Sample	Second survey wave	Second survey wave	Second survey wave	Second survey wave
Mean of dependent variable	0.7392	0.7392	0.7392	0.7392
R^2	0.1331	0.1365	0.3561	0.3934
Number of households	26530	26530	26530	26530
Number of purchase occasions	195268	195268	195268	195268

Notes: Unit of observation is a purchase of a headache remedy by a household. Observations are weighted by equivalent volume (number of pills). Standard errors in parentheses are clustered by household. Income controls are indicators for 16 household income categories. Demographic controls are indicators for categories of race, age, household composition, and housing ownership. "Believe store brands are 'just as safe'" means the primary shopper chose "agree" (1) on a 1-7 agree/disagree scale in response to the statement "Store-brand products for headache remedies/pain relievers are just as safe as the brand name products." All models include fixed effects for the comparable product group.

Table 3: Occupation and headache remedy purchases

Dependent variable: Purchase is a store brand

Primary shopper characteristics:	(1)	(2)	(3)	(4)	(5)
College education	0.0171	0.0288	0.0351		0.0431
	(0.0061)	(0.0064)	(0.0061)		(0.0100)
Pharmacist or physician	0.1527	0.1683	0.1529	0.1667	0.1445
	(0.0296)	(0.0294)	(0.0295)	(0.0380)	(0.0493)
Other healthcare occupation	0.0792	0.0834	0.0790	0.0624	0.0489
	(0.0099)	(0.0098)	(0.0102)	(0.0172)	(0.0224)
Health major				0.0096	
				(0.0165)	
Non-health science major				0.0507	
				(0.0245)	
Demographic controls?	X	X	X	X	X
Market & quarter fixed effects?	X	X			
Income controls?		X	X	X	X
Market-chain-quarter fixed effects?			X	X	X
Sample	All	All	All	College major reported	Not currently employed
Mean of dependent variable	0.7424	0.7424	0.7424	0.7536	0.7390
R^2	0.1166	0.1195	0.3037	0.4401	0.4330
Number of households	39555	39555	39555	14190	13479
Number of purchase occasions	279499	279499	279499	92020	103624

Notes: Unit of observation is a purchase of a headache remedy by a household. Observations are weighted by equivalent volume (number of pills). Standard errors in parentheses are clustered by household. Occupation is defined as of the primary shopper's most recent employment spell. "Health major" and "non-health science major" refer to primary shopper's reported college major. Income controls are indicators for 16 household income categories. Demographic controls are indicators for categories of race, age, household composition, and housing ownership. All models include fixed effects for the comparable product group.

Table 4: Occupation and headache remedy purchases by well-informed consumers

Dependent variable: Purchase is a store brand

Primary shopper characteristics:	(1)	(2)	(3)
College education	0.0313	0.0148	0.0133
	(0.0074)	(0.0129)	(0.0123)
Pharmacist or physician	0.1578	0.1083	0.0304
	(0.0331)	(0.0365)	(0.0379)
Other healthcare occupation	0.0732	0.0466	0.0198
	(0.0130)	(0.0153)	(0.0160)
Sample	Second survey wave	Second survey wave	Second survey wave
Primary shopper survey response:			
Know all active ingredients		X	X
Believe store brands are "just as safe"			X
Mean of dependent variable	0.7392	0.8054	0.8732
R^2	0.3440	0.5412	0.6049
Number of households	26530	6887	4274
Number of purchase occasions	195268	52808	33373

Notes: Unit of observation is a purchase of a headache remedy by a household. Observations are weighted by equivalent volume (number of pills). Standard errors in parentheses are clustered by household. Occupation is defined as of the primary shopper's most recent employment spell. All specifications include demographic controls, income controls, comparable product group fixed effects, and market-chain-quarter fixed effects as in column (3) of table 3. "Know all active ingredients" means the primary shopper correctly identified the active ingredient in all five headache remedies. "Believe store brands are 'just as safe'" means the primary shopper chose "agree" (1) on a 1-7 agree/disagree scale in response to the statement "Store-brand products for headache remedies/pain relievers are just as safe as the brand name products."

Table 5: Occupation and pantry staple purchases

Dependent variable: Purchase is a store brand					
Primary shopper characteristics:	(1)	(2)	(3)	(4)	(5)
College education	-0.0230	-0.0060	-0.0062		-0.0023
	(0.0050)	(0.0052)	(0.0039)		(0.0063)
Chef	0.1383	0.1298	0.1175	0.2079	0.1403
	(0.0204)	(0.0197)	(0.0189)	(0.0513)	(0.0367)
Other food preparer	0.0438	0.0344	0.0227	0.0529	0.0112
	(0.0132)	(0.0127)	(0.0101)	(0.0204)	(0.0157)
Health major				0.0013	
				(0.0101)	
Non-health science major				0.0243	
				(0.0167)	
Demographic controls?	X	X	X	X	X
Market & quarter fixed effects?	X	X			
Income controls?		X	X	X	X
Market-chain-quarter fixed effects?			X	X	X
Sample	All	All	All	College major reported	Not currently employed
Mean of dependent variable	0.5987	0.5987	0.5987	0.5801	0.5931
R^2	0.0885	0.0922	0.3862	0.4453	0.4613
Number of households	44502	44502	44502	15948	15286
Number of purchase occasions	588484	588484	588484	192026	222918

Notes: Unit of observation is a purchase of a pantry staple by a household. Observations are weighted by equivalent volume (pounds). Standard errors in parentheses are clustered by household. Occupation is defined as of the primary shopper's most recent employment spell. "Health major" and "non-health science major" refer to primary shopper's reported college major. Income controls are indicators for 16 household income categories. Demographic controls are indicators for categories of race, age, household composition, and housing ownership. All models include fixed effects for the comparable product group.

Table 6: Evidence on domain specificity

Dependent variable: Purchase is a store brand

Primary shopper characteristics:	(1)	(2)	(3)
College education	-0.0048	-0.0072	0.0430
	(0.0048)	(0.0039)	(0.0061)
Share of active ingredients known	-0.0012		
	(0.0067)		
Pharmacist or physician		0.0018	
		(0.0256)	
Other healthcare occupation		0.0056	
		(0.0084)	
Chef			0.1095
			(0.0340)
Other food preparer			0.0081
			(0.0168)
Products	Pantry Staples	Pantry Staples	Headache Remedies
Mean of dependent variable	0.5978	0.5987	0.7424
R^2	0.4059	0.3860	0.3017
Number of households	29561	44502	39555
Number of purchase occasions	404372	588484	279499

Notes: Unit of observation is a purchase of a pantry staple (first two columns) or headache remedy (third column) by a household. Observations are weighted by equivalent volume (pounds or number of pills). Standard errors in parentheses are clustered by household. Occupation is defined as of the primary shopper's most recent employment spell. All specifications include demographic controls, income controls, comparable product group fixed effects, and market-chain-quarter fixed effects as in column (3) of tables 3 and 5.

Table 7: Health categories purchases under full information

	Headache remedies (6)			***Other health categories (44)***		
	Baseline	Informed consumers at baseline prices	Informed consumers at equilibrium prices	Baseline	Informed consumers at baseline prices	Informed consumers at equilibrium prices
National-brand quantity share	0.258	0.117	0.286	0.435	0.370	0.506
		(0.040)	(0.107)		(0.021)	(0.026)
National-brand price (relative to cost)	6.036	—	3.798	3.639	—	3.181
			(1.245)			(0.290)
Store-brand price (relative to cost)	2.047	—	1.996	1.949	—	1.809
			(0.150)			(0.068)
Change as a share of baseline expenditure:						
Manufacturer profit		-0.188	-0.141		-0.053	-0.011
		(0.057)	(0.066)		(0.014)	(0.019)
Retailer profit		0.053	-0.010		0.019	-0.034
		(0.017)	(0.035)		(0.005)	(0.012)
Consumer expenditure		-0.135	-0.151		-0.034	-0.044
		(0.040)	(0.089)		(0.009)	(0.026)
Consumer surplus		0.038	0.150		0.033	0.037
		(0.029)	(0.088)		(0.007)	(0.027)
Total surplus		-0.097	-0.000		-0.001	-0.007
		(0.027)	(0.008)		(0.007)	(0.005)
Baseline consumer expenditure ($bn / year):	$2.88			$8.94		

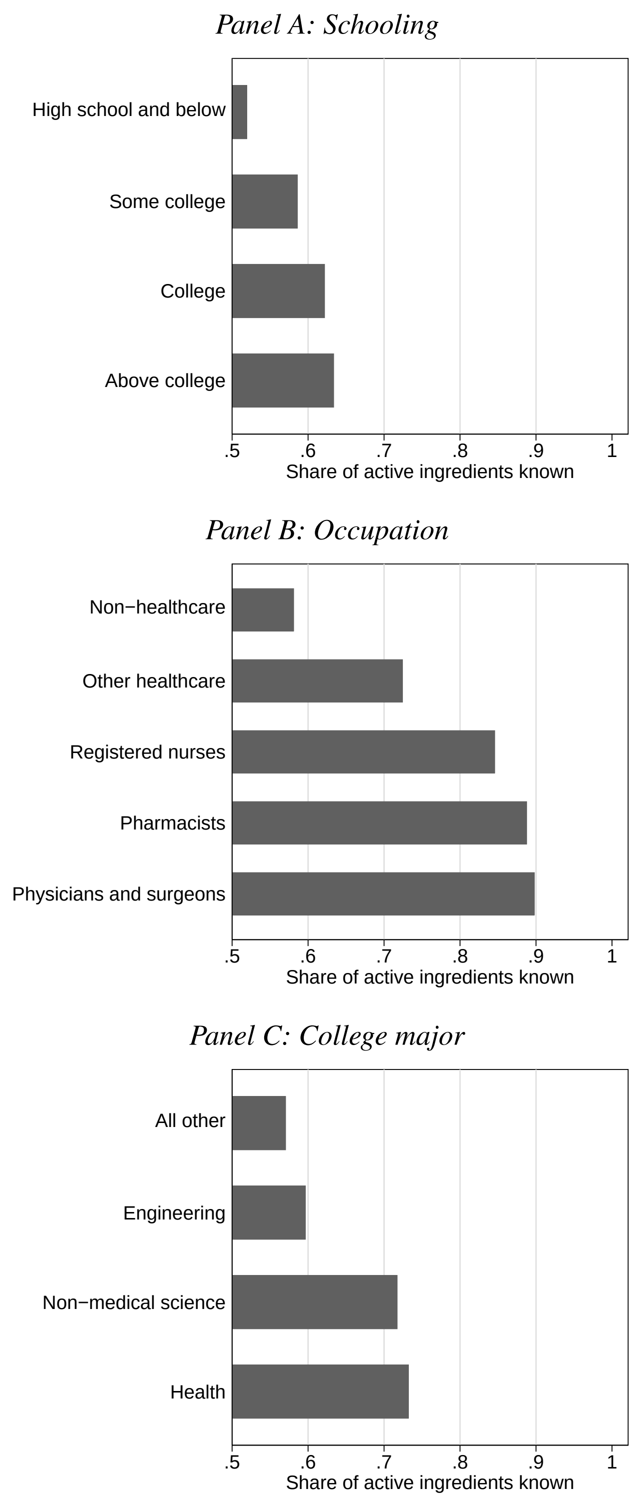
Notes: The two panels report results for headache remedy comparables and other health comparables, respectively, with the number of comparables in parentheses. The "baseline" column reports average prices relative to estimated manufacturing costs and repeats summary information from table 1. Total expenditure are estimated 2008 totals for all grocery, drug, and mass merchandise stores in the US. Headache remedy relative prices and national-brand shares are averaged over comparable product groups weighting by equivalent units sold, while other health category relative prices and national-brand shares are averaged over comparable product groups weighting by expenditure. The "informed consumers at baseline prices" counterfactual computes the effect of all households choosing according to true rather than perceived brand preference, holding prices constant at baseline levels. The "informed consumers at baseline prices" counterfactual further allows prices to adjust to reflect the change in consumer demand. Standard errors in parentheses are from 10 bootstrap replications in which we draw households at random with replacement and recompute all estimates. These standard errors thus account for correlation in sampling error across comparables. See section 6 for details of model specification and estimation.

Table 8: Food and drink purchases under full information

	Pantry staples (6)			***Other food and drink categories (235)***		
	Baseline	Informed consumers at baseline prices	Informed consumers at equilibrium prices	Baseline	Informed consumers at baseline prices	Informed consumers at equilibrium prices
National-brand quantity share	0.404	0.311	0.434	0.575	0.565	0.574
		(0.018)	(0.006)		(0.010)	(0.006)
National-brand price (relative to cost)	1.312	—	1.250	1.962	—	2.010
			(0.008)			(0.041)
Store-brand price (relative to cost)	1.146	—	1.134	1.346	—	1.343
			(0.002)			(0.004)
Change as a share of baseline expenditure:						
Manufacturer profit		-0.018	-0.016		-0.005	0.005
		(0.003)	(0.003)		(0.005)	(0.008)
Retailer profit		0.008	-0.008		0.003	-0.001
		(0.002)	(0.002)		(0.002)	(0.002)
Consumer expenditure		-0.010	-0.024		-0.002	0.004
		(0.002)	(0.004)		(0.004)	(0.008)
Consumer surplus		0.002	0.026		0.004	-0.004
		(0.001)	(0.004)		(0.001)	(0.010)
Total surplus		-0.008	0.001		0.002	0.000
		(0.001)	(0.001)		(0.004)	(0.003)
Baseline consumer expenditure ($bn / year):	$1.81			$122.61		

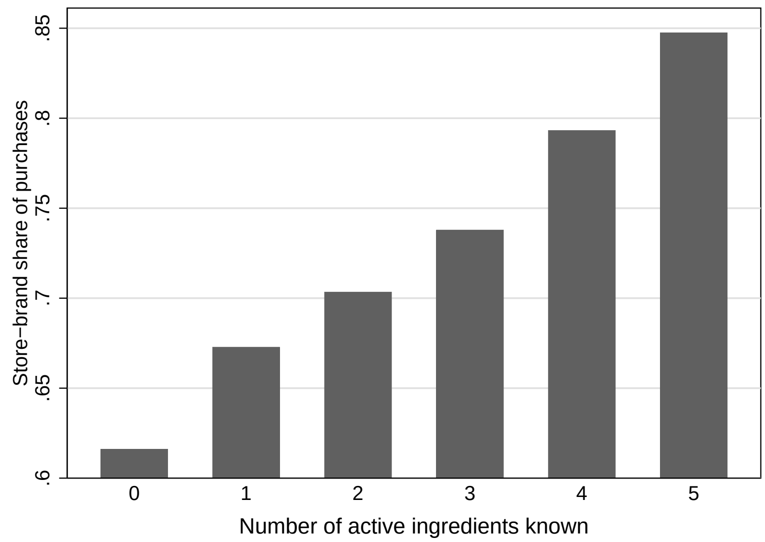
Notes: The two panels report results for pantry staples comparables and other food and drink comparables, respectively, with the number of comparables in parentheses. The "baseline" column reports average prices relative to estimated manufacturing costs and repeats summary information from table 1. Total expenditure are estimated 2008 totals for all grocery, drug, and mass merchandise stores in the US. Relative prices and national-brand shares are averaged over comparable product groups weighting by expenditure. The "informed consumers at baseline prices" counterfactual computes the effect of all households choosing according to true rather than perceived brand preference, holding prices constant at baseline levels. The "informed consumers at baseline prices" counterfactual further allows prices to adjust to reflect the change in consumer demand. Standard errors in parentheses are from 10 bootstrap replications in which we draw households at random with replacement and recompute all estimates. These standard errors thus account for correlation in sampling error across comparables. See section 6 for details of model specification and estimation.

Figure 1: Product knowledge, headache remedies



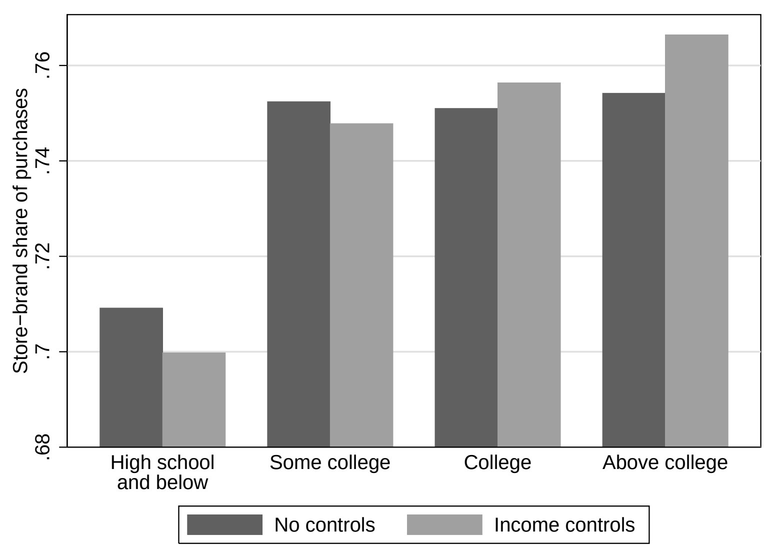
Notes: Figure shows the mean share of headache remedy active ingredients correctly identified by each group of respondents in the 2011 PanelViews survey, among those who answered all five questions.

Figure 2: Store-brand purchases and knowledge, headache remedies



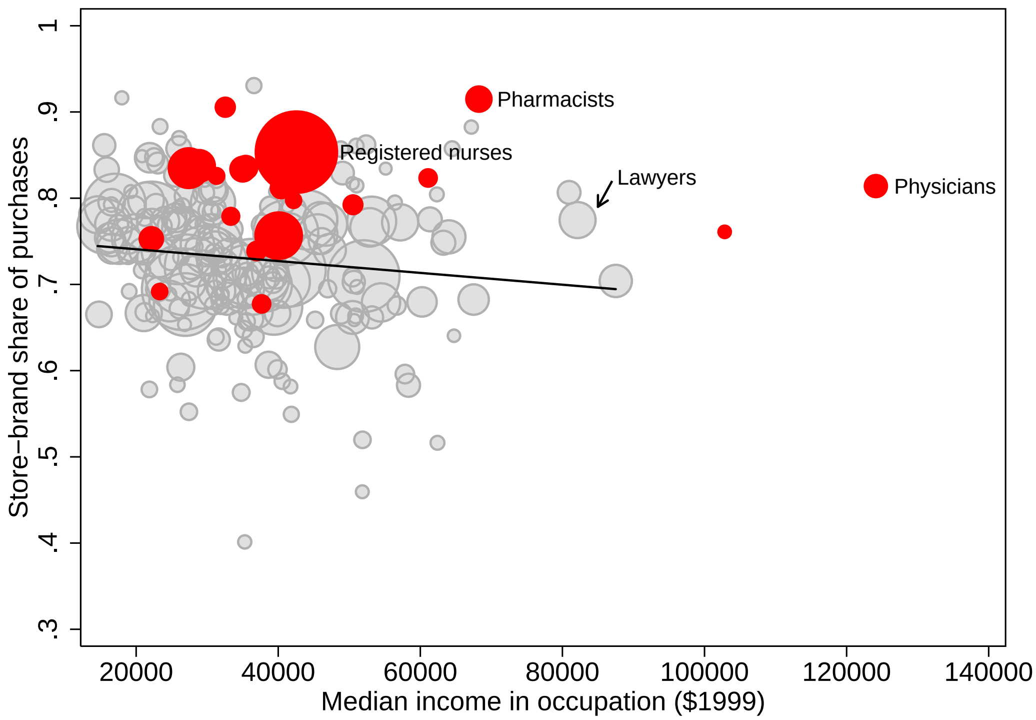
Notes: Horizontal axis shows the number of headache remedy active ingredients correctly identified in the 2011 PanelViews survey. The bars show the store-brand share of headache remedies for households in each category, weighted by equivalent volume (number of pills). Sample is restricted to panelists who answered all five active ingredient questions.

Figure 3: Store-brand purchases and education, headache remedies



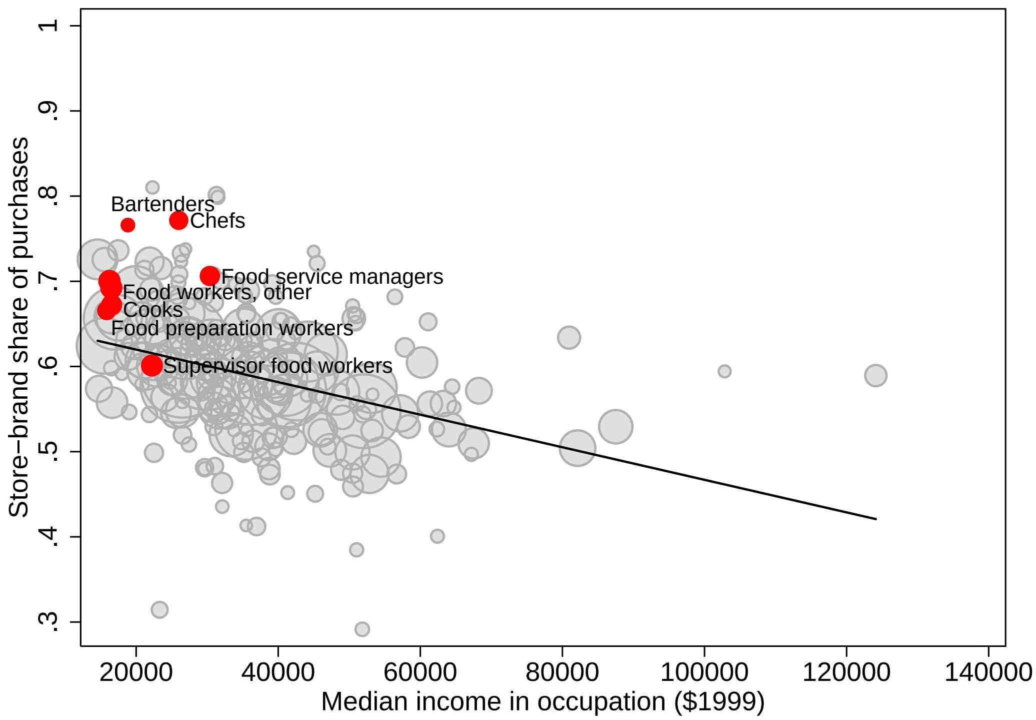
Notes: Bars labeled "no controls" show the store-brand share of headache remedy purchases for households in each education category, weighted by equivalent volume (number of pills). Bars labeled "income controls" show the predicted store-brand share in each education category from a regression on indicators for education categories and 16 household income categories, with the predicted values computed at the means of the covariates.

Figure 4: Store-brand purchases and occupation, headache remedies



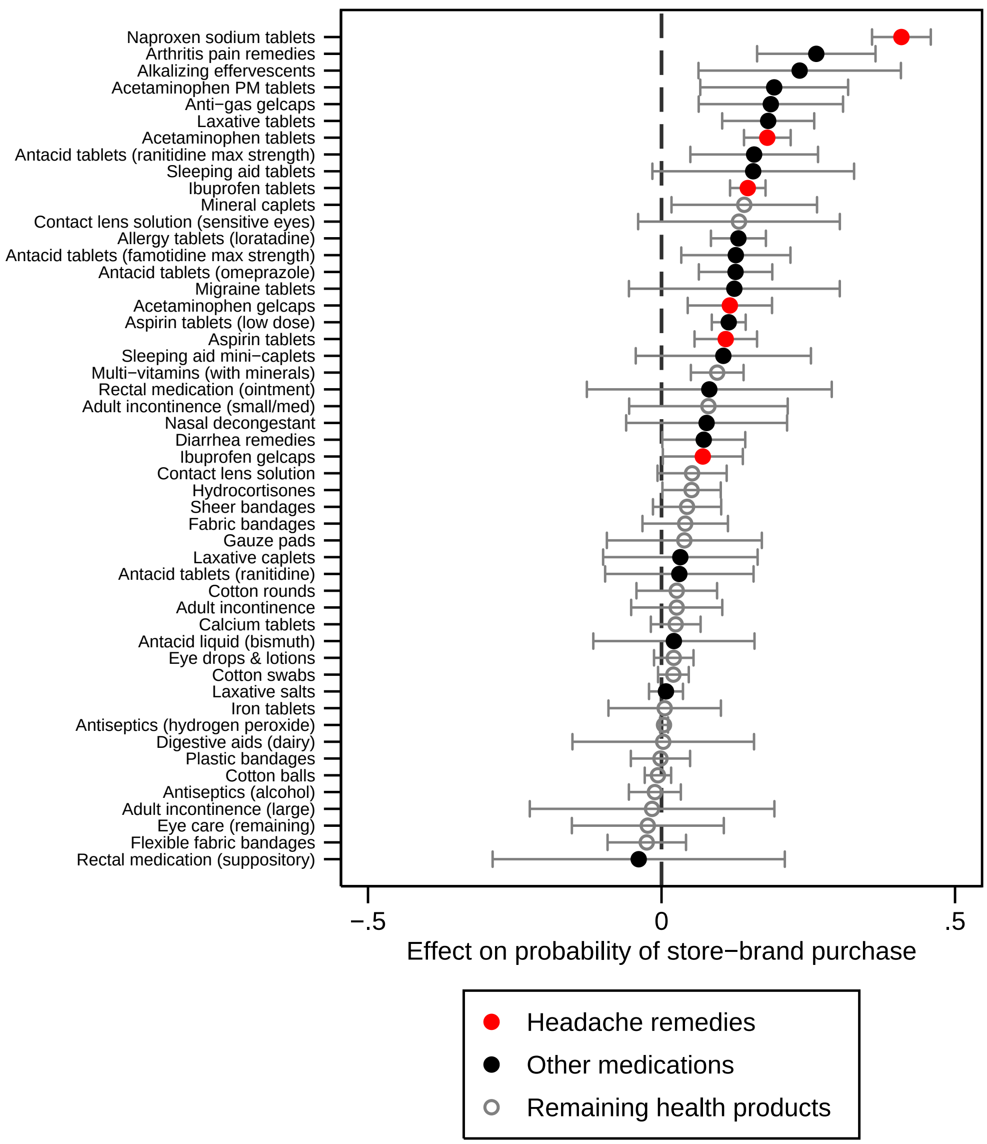
Notes: Figure shows store-brand share of headache remedy purchases by occupation (y-axis) and median earnings for full-time full-year workers in 1999 by occupation (x-axis), weighted by equivalent volume (number of pills). Filled (colored) circles represent healthcare occupations. Occupation weights are given by the number of households whose primary shopper has the given occupation in our sample (occupations with fewer than 25 such households are excluded from the figure). The area of each circle is proportional to the occupation weights, with different scale for healthcare and non-healthcare occupations. The line is the prediction from an OLS regression of store-brand share of purchase volume on median earnings excluding healthcare occupations and weighting each occupation by the occupation weights.

Figure 5: Store-brand purchases and occupation, pantry staples



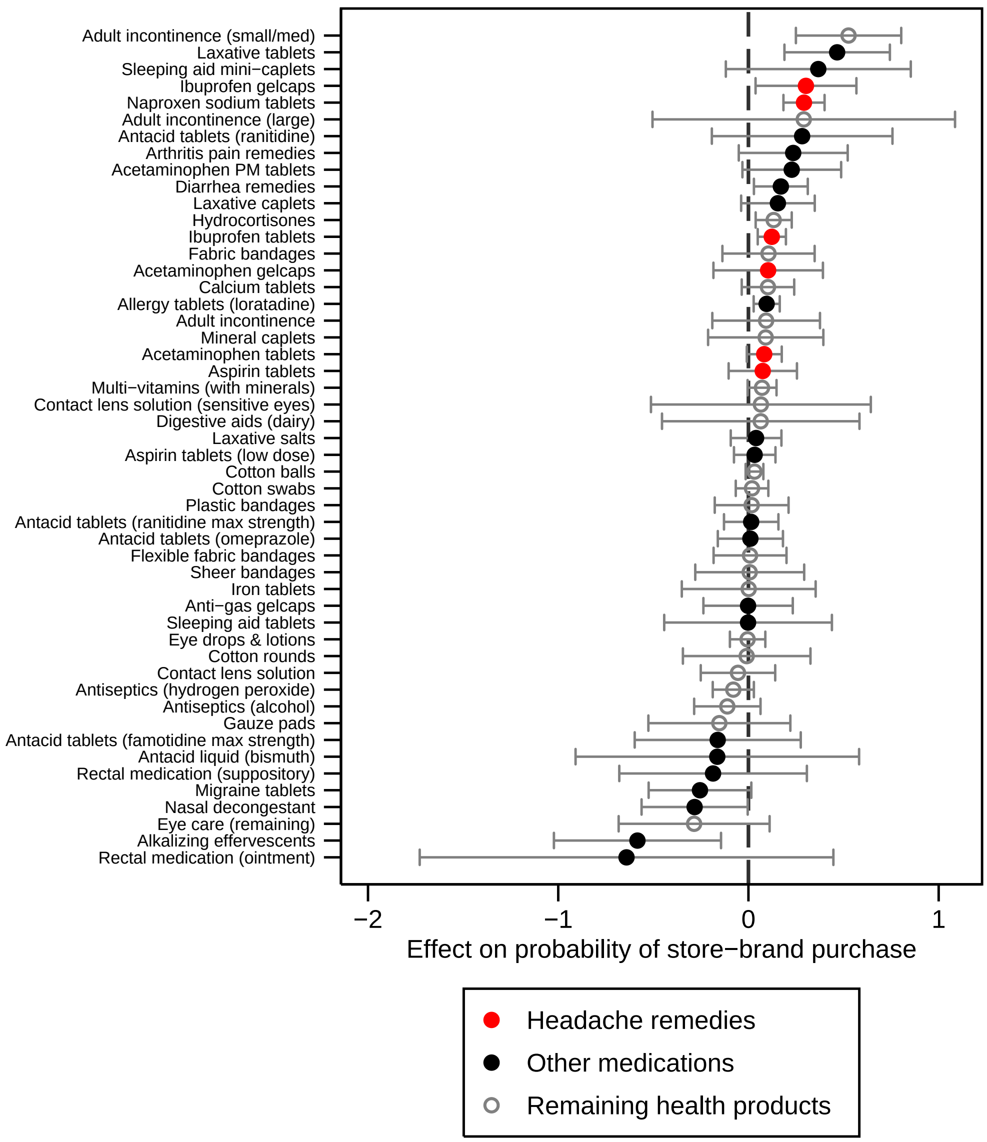
Notes: Figure shows store-brand share of pantry staple purchases by occupation (y-axis) and median earnings for full-time full-year workers in 1999 by occupation (x-axis), weighted by equivalent volume (pounds). Filled (colored) circles represent food preparer occupations. Occupation weights are given by the number of households whose primary shopper has the given occupation in our sample (occupations with fewer than 25 such households are excluded from the figure). The area of each circle is proportional to the occupation weights, with different scale for food preparer and non-food-preparer occupations. The line is the prediction from an OLS regression of store-brand share of purchase volume on median earnings excluding food preparer occupations and weighting each occupation by the occupation weights.

Figure 6: Active ingredient knowledge coefficients



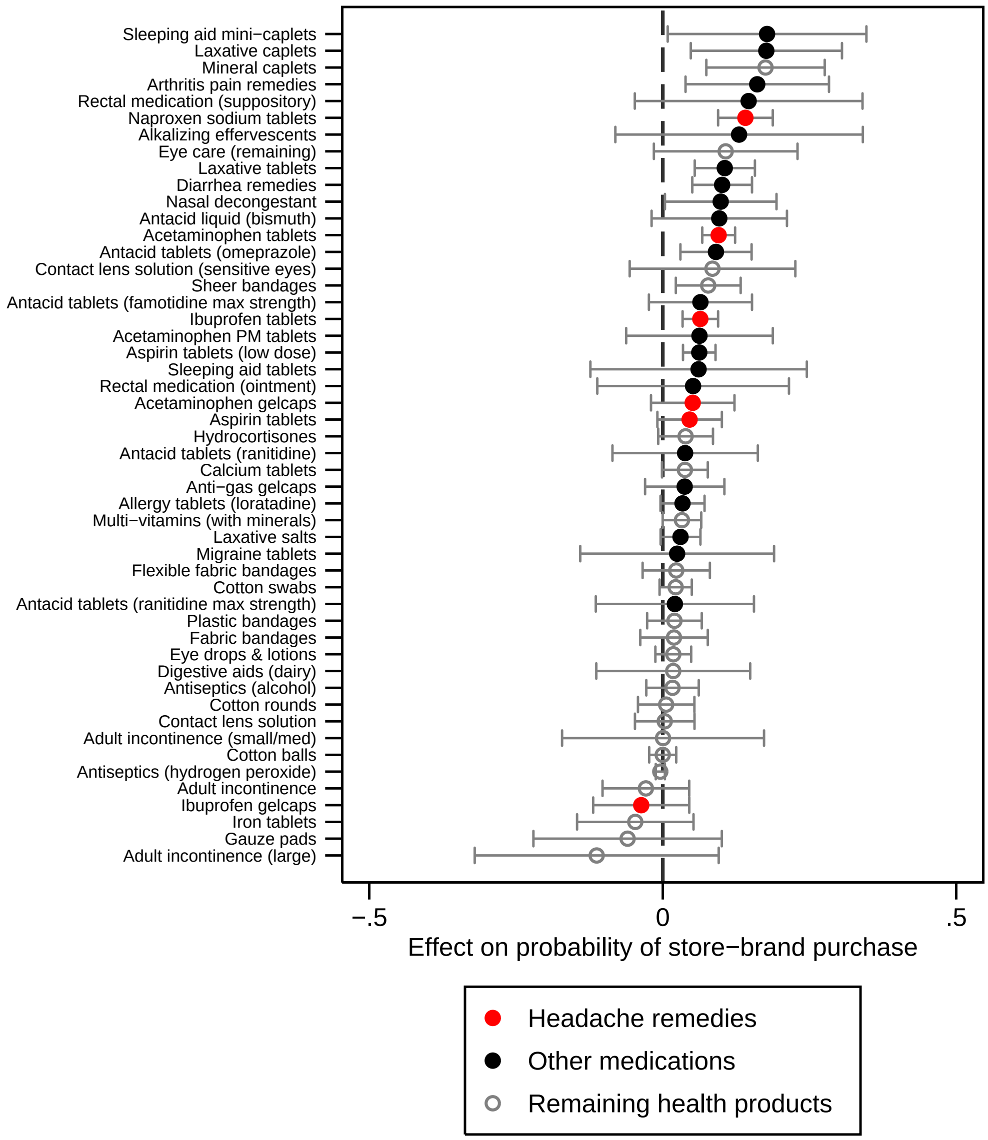
Notes: Figure plots coefficients and 95 percent confidence intervals on “share of active ingredients known” for each health-related comparable product group from a regression following the specification of table 2 column (3).

Figure 7: Pharmacist / physician occupation coefficients



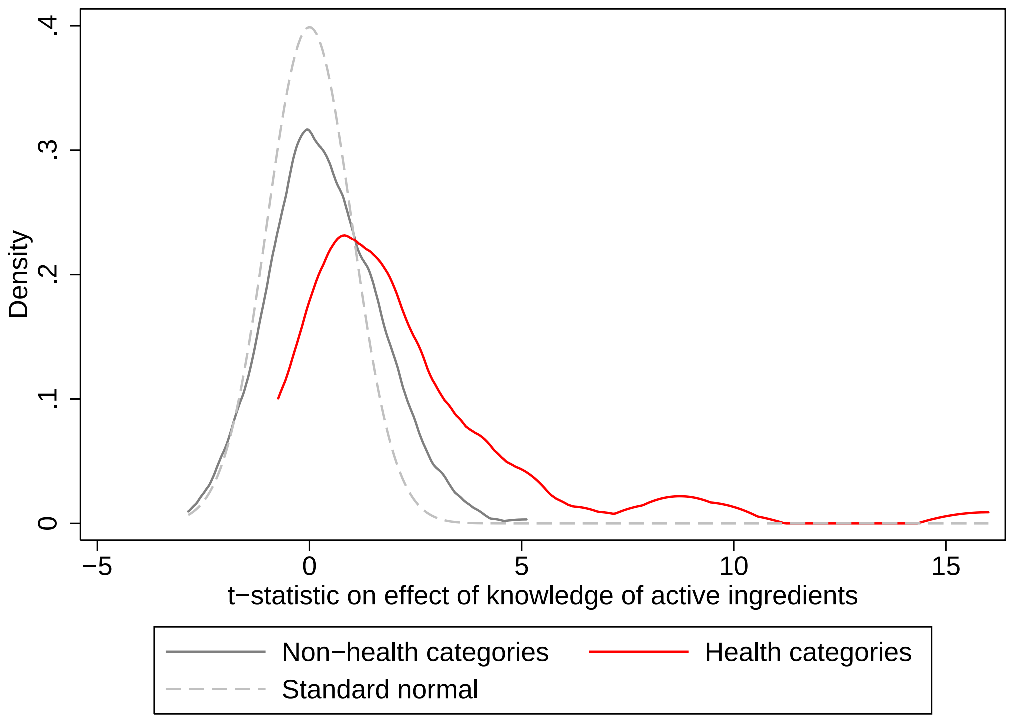
Notes: Figure plots coefficients and 95 percent confidence intervals on "pharmacist or physician" for each health-related comparable product group from a regression following the specification of table 3 column (3).

Figure 8: Other healthcare occupation coefficients



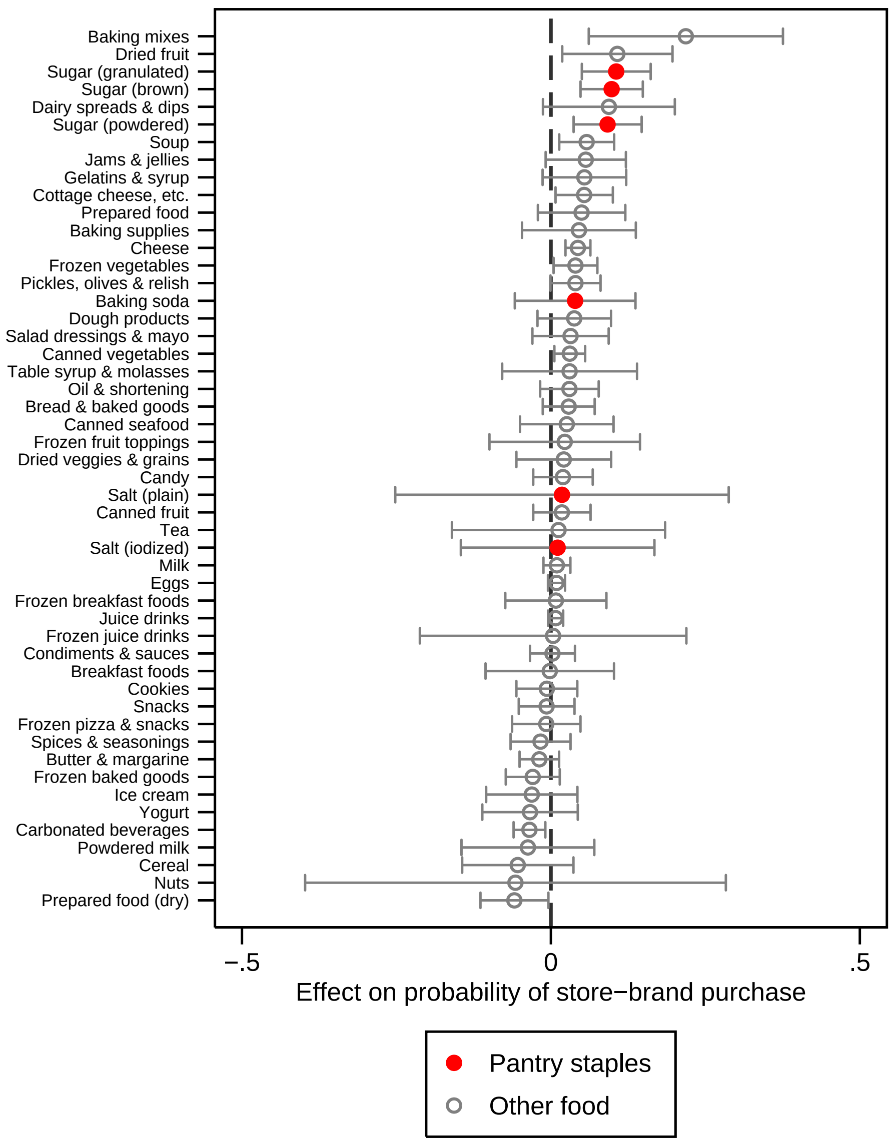
Notes: Figure plots coefficients and 95 percent confidence intervals on "other healthcare occupation" for each health-related comparable product group from a regression following the specification of table 3 column (3).

Figure 9: Active ingredient knowledge coefficients, health vs. non-health products



Notes: Figure plots the distribution of t-statistics on "share of active ingredients known" for all health-related and non-health-related comparable products groups from a regression following the specification of table 2 column (3). Distribution is estimated using an Epanechnikov kernel with optimal bandwidth. The standard normal density is plotted with dashed lines.

Figure 10: Chef coefficients



Notes: Figure plots coefficients and 95 percent confidence intervals on "chef" for each food and drink category. Coefficients are estimated separately for each comparable in a regression following the specification of table 5 column (3). Coefficients for pantry staples are plotted individually by comparable. We aggregate coefficients for all other comparables to the Nielsen product group level, reporting the precision-weighted mean of the estimated coefficients and constructing the confidence intervals based on the harmonic mean of the estimated variances.

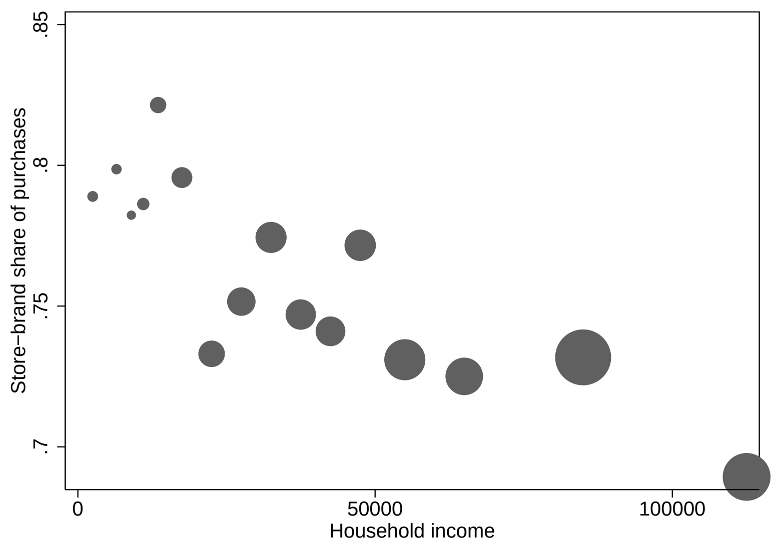
Appendix Table 1: Knowledge and headache remedy purchases, robustness

Dependent variable: Purchase is a store brand

		Headache remedies			Pantry staples
		Share of active ingredients coefficient	College education coefficient	Pharmacist / physician coefficient	Chef coefficient
(1)	Baseline	0.1898 (0.0108)	0.0351 (0.0061)	0.1529 (0.0295)	0.1175 (0.0189)
(2)	Control for market-chain-week	0.2038 (0.0142)	0.0316 (0.0076)	0.1888 (0.0379)	0.1118 (0.0220)
(3)	Control for market-chain-store-quarter	0.2067 (0.0174)	0.0325 (0.0095)	0.1137 (0.0530)	0.1101 (0.0259)
(4)	Control for market-chain-store-week	0.2305 (0.0294)	0.0290 (0.0146)	0.1904 (0.0849)	0.0995 (0.0463)
(5)	Control for average annual purchase volume	0.1828 (0.0109)	0.0371 (0.0062)	0.1438 (0.0293)	0.1066 (0.0189)
(6)	Control for average annual grocery spending	0.1924 (0.0108)	0.0319 (0.0061)	0.1534 (0.0285)	0.1195 (0.0191)
(7)	Control for median occupational income	0.1905 (0.0108)	0.0350 (0.0063)	0.1528 (0.0319)	0.1147 (0.0189)
(8)	Condition sample on item size availability	0.1786 (0.0122)	0.0404 (0.0066)	0.1375 (0.0384)	0.0998 (0.0215)
(9)	Condition sample on item size availability and control for item size	0.1691 (0.0118)	0.0366 (0.0064)	0.1349 (0.0376)	0.0974 (0.0204)
(10)	Weight observations by Nielsen projection factor	0.1879 (0.0137)	0.0532 (0.0085)	0.1180 (0.0334)	0.1092 (0.0242)
(11)	Impute characteristics of actual shopper	0.1969 (0.0110)	0.0405 (0.0068)	0.1578 (0.0332)	0.1256 (0.0224)
(12)	Logit controlling for market and quarter	0.2119 (0.0101)	0.0327 (0.0062)	0.2240 (0.0461)	0.1290 (0.0210)

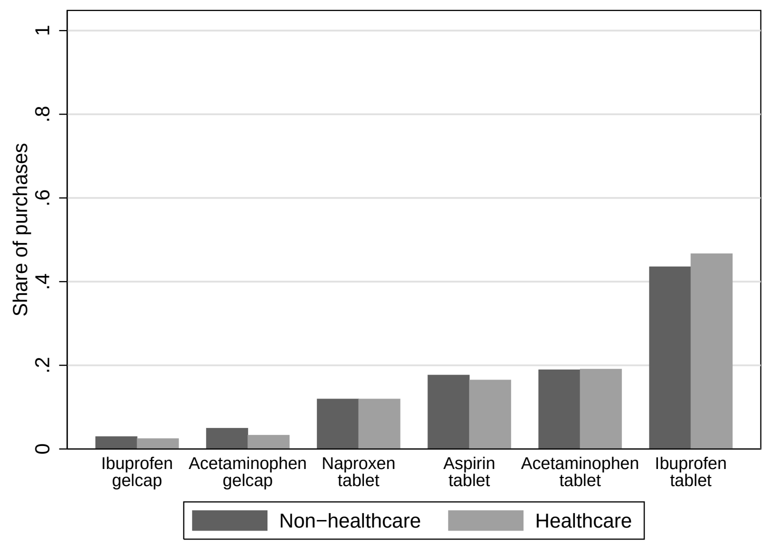
Notes: Each row gives (i) the coefficient on "share of active ingredients known" from a specification analogous to table 2 column (3); (ii) the coefficient on "college education" from a specification analogous to table 3 column (3); (iii) the coefficient on "pharmacist or physician" from a specification analogous to table 3 column (3); and (iv) the coefficient on "chef" from a specification analogous to table 5 column (3). Row (1) repeats the results from our main specifications. Row (2) is the same as the baseline but replaces market-chain-quarter fixed effects with market-chain-week fixed effects. Row (3) is the same as the baseline but replaces market-chain-quarter fixed effects with market-chain-store-quarter fixed effects. Row (4) is the same as the baseline but replaces market-chain-quarter fixed effects with market-chain-store-week fixed effects. Row (5) is the same as the baseline but adds a control for the average annual volume of headache remedies (columns 1-3) and pantry staples (column 4) purchased by the household. Row (6) is the same as the baseline but adds a control for the household's average annual grocery spending. Row (7) is the same as the baseline but adds a control for the median income of the occupation of the primary shopper. Row (8) is the same as the baseline but restricts attention to transactions such that at least one comparable national-brand purchase and at least one comparable store-brand purchase are observed in the Homescan data in the same retail chain, quarter, and item size as the given transaction. Row (9) is the same as row (8) but replaces comparable product group fixed effects with comparable product group-item size fixed effects. Row (10) is the same as the baseline but weights observations by the Nielsen projection factor. Row (11) is the same as the baseline but imputes characteristics of the actual shopper by assuming that the primary shopper is the actual shopper when there is no secondary shopper and that the primary shopper is the actual shopper 74 percent of the time when there is a secondary shopper; see the online appendix for details. Row (12) is the same as the baseline but replaces the linear probability model with a logit model and the market-chain-quarter fixed effects with market and quarter fixed effects; observations are not weighted and reported coefficients are average marginal effects.

Appendix Figure 1: Store-brand purchases and household income, headache remedies



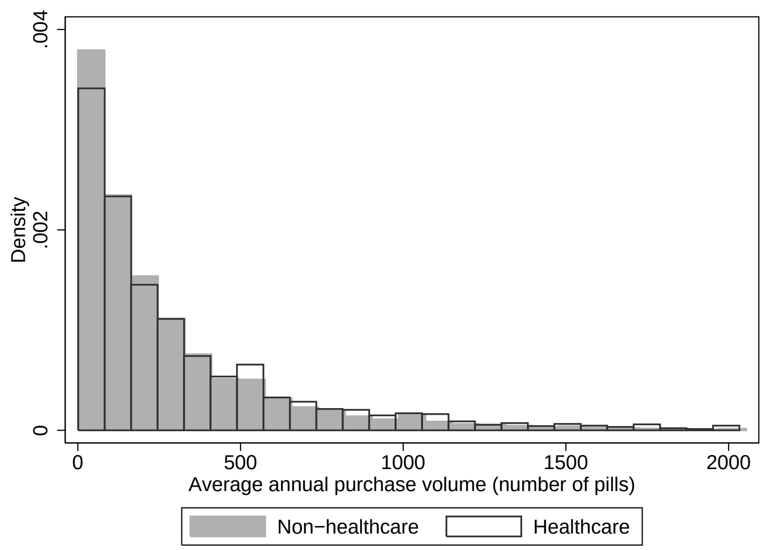
Notes: Figure shows the store-brand share of headache remedy purchases for households in each income category, weighted by equivalent volume (number of pills). Household income is imputed at the midpoint of the range for each category, with the top category imputed at 120,000. The area of each circle is proportional to the number of households in the income category in our sample.

Appendix Figure 2: Physical attribute choice and occupation, headache remedies



Notes: Share of purchases is computed from a set of linear probability models of the likelihood of purchasing the given product. Bars labeled "healthcare" show the predicted probability from the given model for purchases made by households whose primary shopper is in a healthcare occupation. Bars labeled "non-healthcare" show the predicted probability for the same purchases under the counterfactual in which the household's primary shopper is not in a healthcare occupation. Each linear probability model's unit of observation is a purchase occasion. Observations are weighted by equivalent volume (number of pills). All specifications include an indicator for college completion, income controls, demographic controls, and market-chain-quarter fixed effects. Income controls are dummies for 16 household income categories. Demographic controls are dummies for categories of race, age, household composition, and housing ownership. Predicted probabilities set the market-chain-quarter fixed effect so that the mean predicted probability is equal to the empirical share. See the online appendix for a supporting table with additional details.

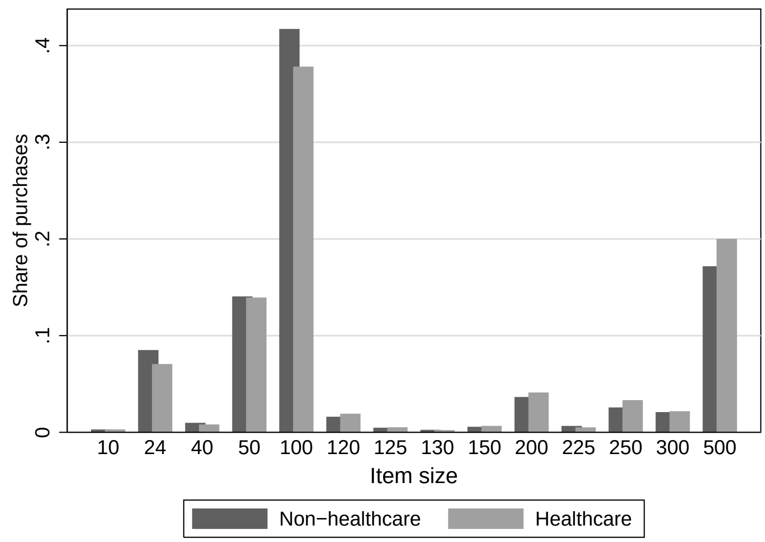
Appendix Figure 3: Average annual purchase volume and occupation, headache remedies



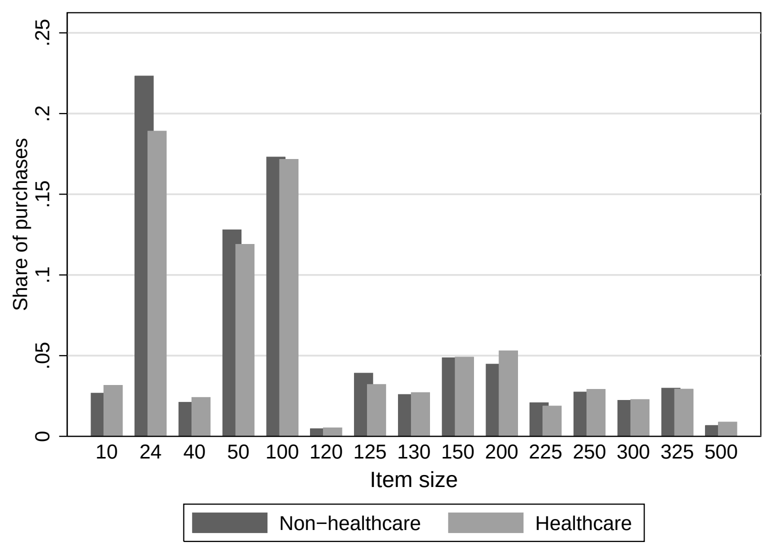
Notes: Bars labeled "healthcare" show the distribution of average annual purchase volume of headache remedies for households whose primary shopper is in a healthcare occupation. Bars labeled "non-healthcare" show the distribution of average annual purchase volume of headache remedies for households whose primary shopper is not in a healthcare occupation. Households in the top percentile of the overall average annual purchase volume distribution are excluded from the figure.

Appendix Figure 4: Item size and occupation, headache remedies

Panel A: Store-brand



Panel B: National-brand



Notes: In panel A, bars labeled "healthcare" show the shares of store-brand headache remedy purchases for a given item size, as fractions of total store-brand headache remedy purchases made by households whose primary shopper is in a healthcare occupation. Bars labeled "non-healthcare" show the same for households whose primary shopper is not in a healthcare occupation. In panel B, bars labeled "healthcare" show the shares of national-brand headache remedy purchases for a given item size, as fractions of total national-brand headache remedy purchases made by households whose primary shopper is in a healthcare occupation. Bars labeled "non-healthcare" show the same for households whose primary shopper is not in a healthcare occupation. Only the top 15 item sizes in terms of total number of purchases across both store-brand and national-brand headache remedies are included in the figure.