ECON-C4100 - Capstone: Econometrics I

Lecture 8: Do pharmacists buy Bayer? Informed shoppers and the brand premium

Otto Toivanen

The paper

Bronnenberg, B. J., Dubé, J.-P., Gentzkow, M. & Shapiro, J. M. (2015). Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium *. *The Quarterly Journal of Economics*, *130*(4), 1669–1726

- We are not going through the entire paper, but selected main parts of it.
- Pay attention to
  1. how the authors motivate their research question
  2. the way data is used in the paper
  3. the way hypotheses are formed and discussed
  4. the discussion and specification of the regression model
  5. the discussion of the results
  6. the robustness tests

Research question

- Why do some buyers pay more for a branded product?
- Example: pain killers., 400mg.

Name of product	Pack size	Pharmaceutical company	Price e	Active substance
BURANA	10	ORION CORPORATION	3,98	ibuprofen
IBUTABS	10	ORION CORPORATION	3,75	ibuprofen
IBUXIN	10	RATIOPHARM	1,97	ibuprofen
IBUMAX	10	VITABALANS	1,97	ibuprofen

- Same active ingredient (ibuprofen), all sold for a long period, no news on any side effects on one vs. other.

Research question

- Brands may deliver superior quality or reliability or something else.
- Notice that the domestic producer ORION has two products available, at different prices.
- Consumers may overestimate the benefits of a brand (i.e., be confused or misled).
- Problem: How to separate the two hypotheses?
- Solution: Think with the help of a formal model.

Motivation of the research question: empirics #1

- (pp. 1670): *In our data (described in more detail later), we find that consumers would spend 44 billion less per year on consumer packaged goods (CPG) if they switched from a national brand to a store-brand alternative whenever possible. Prior work documents substantial brand price premia in a wide range of non-CPG categories, such as automobiles (Sullivan 1998), index funds (Hortaçsu and Syverson 2004), and online books (Smith and Brynjolfsson 2001).*

Motivation of the research question: existing literature #2

- (pp. 1670): *Economists have long debated the origins of brand premia. On the one hand, national brands may offer superior quality or reliability, or may deliver direct utility benefits (Becker and Murphy 1993). On the other hand, consumers may be willing to pay a premium for brands because they overestimate the benefits of the brand or are otherwise confused or misled. Determining which story holds is important for evaluating the efficiency of consumer goods markets and the welfare effects of advertising, and may be relevant to policy decisions in consumer protection and regulation*.

Model

- Think of a household which has to choose between a **national brand** and a **store brand**.
- The price difference is

$$\Delta p = p_n - p_s > 0$$

Model

- The household believes that the national brand delivers a higher utility

$$\Delta v = v_n - v_s > 0$$

- The true utility difference is

$$\Delta \tilde{v} = \tilde{v}_n - \tilde{v}_s > 0$$

- The latter can be larger or smaller than the former.

Model

- The household chooses the national brand if and only if

$$\Delta v - \Delta p \geq 0$$

or, put differently, iff

$$v_n - p_n \geq v_s - p_s$$

Informed and uninformed choice

- A choice is **informed** if a household buys the national brand **if and only if**

$$\Delta \tilde{v} > \Delta p$$

- A choice is **uninformed** if a household buys the national brand **if and only if**

$$\Delta v > \Delta p$$

Informed vs. uninformed choice

- An informed choice differs from an uniformed one only if and only if

$$\Delta \tilde{v} > \Delta p$$
$$\Delta v < \Delta p$$

  or the other way round.

- Notice: framework does not capture a possible placebo effect, i.e., that a national brand works better if the consumer knows hat he is consuming the national brand.

Informed vs. uninformed choice

- If

$$\Delta v = v_n - v_s > \Delta p = p_n - p_s > \tilde{v}_n - \tilde{v}_s = \Delta \tilde{v}$$

then under **informed** choice the consumer switches from the national brand to the store brand and gains

$$\Delta p - \Delta \tilde{v}$$

- We could derive a similar rule for an informed consumer switching from the national brand to the store brand.

Informed vs. uninformed choice

- Why would there be differences in true or perceived quality?

1. Perceptions of quality vs. true differences in quality.
2. Perceptions in vs. true differences risk of failure (not curing headache).
3. Attention to irrelevant factors.

How to identify welfare gains from informed choice?

- Case #1: $\Delta\tilde{v} = 0$ i.e., no differences in true quality.
  - An informed HH would buy the store brand and pocket Δp.
  - Calculating the gain across households gives the aggregate consumer welfare improvement.
- Case #2: $\Delta\tilde{v}$ unknown.
  - Challenge: actual choices based on Δv, so how to identify $\Delta\tilde{v}$?
  - Approach: Suppose that for HH i, the perceived quality difference is $\Delta\hat{v}_i = \phi_i\Delta\tilde{v} + (1 - \phi_i)\Delta v$.
  - $\phi_i \in [0, 1]$ is a measure of how informed HH i is.

How to identify welfare gains from informed choice?

- Case #2: $\Delta\tilde{v}$ unknown.
  - The researchers observe whether HH i chooses the store brand ($y_i = 1$) or the national brand ($y_i = 0$).,
  - Recall

    $$\Delta\hat{v}_i = \phi_i \Delta\tilde{v} + (1 - \phi_i)\Delta v$$

  - Notice missing i subscripts on RHS for $\Delta\tilde{v}$ and Δv. This means we assume these are constant = same for every HH.
  - The variable ϕ_i an index for how informed HH i is.

How to measure whether HHs buy too often the national brand?

- Imagine you observed the choices of a large number of HHs, and their ϕ_i.
- The higher is ϕ_i, the more likely it is that the HH makes the right choice.
- If choice is independent of ϕ_i, then all HHs are making the right choice (i.e., $\Delta v, \Delta \tilde{v} > \Delta p$).

How to measure whether HHs buy too often the national brand?

- More generally, we can find the threshold value of ϕ_i denoted $\ddot{\phi}$ such that HH i only buys the national brand if

$$\Delta \hat{v}_i > \Delta p$$
$$\Delta \hat{v}_i = \phi_i \Delta \tilde{v} + (1 - \phi_i)\Delta v > \ddot{\phi}\Delta + (1 - \ddot{\phi})\Delta v = \Delta p$$

- i.e., if the HH puts "enough" weight on Δv.
- Households with $\phi_i < \ddot{\phi}$ buy the national brand, but would benefit from buying the store brand.

How to measure whether HHs buy too often the national brand?

- How can we learn whether consumers are buying too much or too little of the national brand?
- We should study what happens to $\mathbb{E}[y|.]$ as ϕ varies.
- If $\partial\mathbb{E}[y|.]/\partial\phi > (<)0$, then we know that consumers are buying too much (too little) of the national brand.

How to measure $\Delta\tilde{v}$

- Why would we like to know $\Delta\tilde{v}$?
- Because then we could measure the welfare (gain).
- How to measure it? Imagine again that we observe ϕ_i.
- Then pick those HHs with $\phi_i = 1$ i.e., fully informed HHs. They always make the right choice.
- Imagine then that you observe variation in Δp.
- There will be markets
  1. where every informed consumer buys the store brand ($\Delta\tilde{v} > \Delta p$),
  2. markets where every informed consumer buys the national brand ($\Delta\tilde{v} < \Delta p$),
  3. and markets where they split because they are indifferent ($\Delta\tilde{v} = \Delta p$).

Towards the empirics

- So, the researchers need data that maps into:
  1. Difference in true value of national and store brand
  2. Price differences between the national and store brand
  3. How informed a HH is.

Empirically measuring ϕ_i

- How to know if a HH is informed?
- The authors ran a survey, asking HHs about their knowledge of particular goods (response rate 62.2%, with some selection into the sample).
- Variables that they observe that affect $\phi_i(\boldsymbol{K})$:
  1. Knowledge of active ingredients
  2. Completed schooling
  3. College major.
  4. Occupation.

Empirically measuring Δp

- The authors do not directly observe prices.
- Therefore, they try to control for the variation in the price difference through $\boldsymbol{Z}$:
  1. Market variables.
  2. Chain variables
  3. Timing (quarter) variables

How to ensure HHs have the same preferences?

- If HHs have different views on what good is "good", they may make rationally different choices.

  $\rightarrow$ authors study goods that are homogenous in observable attributes.

- What about income? Authors have a both direct and indirect measures of it.

Data

- Nielsen Homescan Panel 2004 2011.
- Nielsen is a company that collects price data, packages it and sells the information back to the firms, providing them information on how their pricing differs from their competitors.
- Note the potential competition policy worry involved in this.
- + two custom made surveys (2008 and 2011)
- Data at store and HH level.
- 77Mio shopping trips by 125K HHs.

Data

- For each purchase, the authors observe
  1. Product code
  2. Price
  3. Chain id
  4. Pack size

Data

- For each HH, the authors observe
  1. Education of HH head
  2. HH income (categorical)
  3. # adults
  4. Race
  5. Age
  6. HH composition
  7. Home ownership
  8. Location

Data

- Is the Nielsen data representative?
- Not exactly, though probably not that far off either.
- The survey data is not (smaller, higher income, more educated and white HHs more likely to respond).
- Internal and external validity: Latter if and only if the whole population reacts similarly to information.

Familiarity with headache remedy brands

- Ensuring comparison of apples to apples
  - Brand
  - Item size
- Only include products with > 500 purchases, and products where national brand more expensive than store brand.

  $\rightarrow$ 420 comparables.
- This **sample selection on observables** (price difference) means that they can expect a monotonic relation between $\mathbb{E}[y|.]$ and ϕ.

Estimation equation

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

$y_i = 1$ HH i buys **store** brand, 0 if national brand.

$\boldsymbol{K}$ = vector of information variables.

$\boldsymbol{X}$ = vector of HH characteristics.

$\boldsymbol{Z}$ = choice environment characteristics.

- Equation (1) is a **linear probability model**.

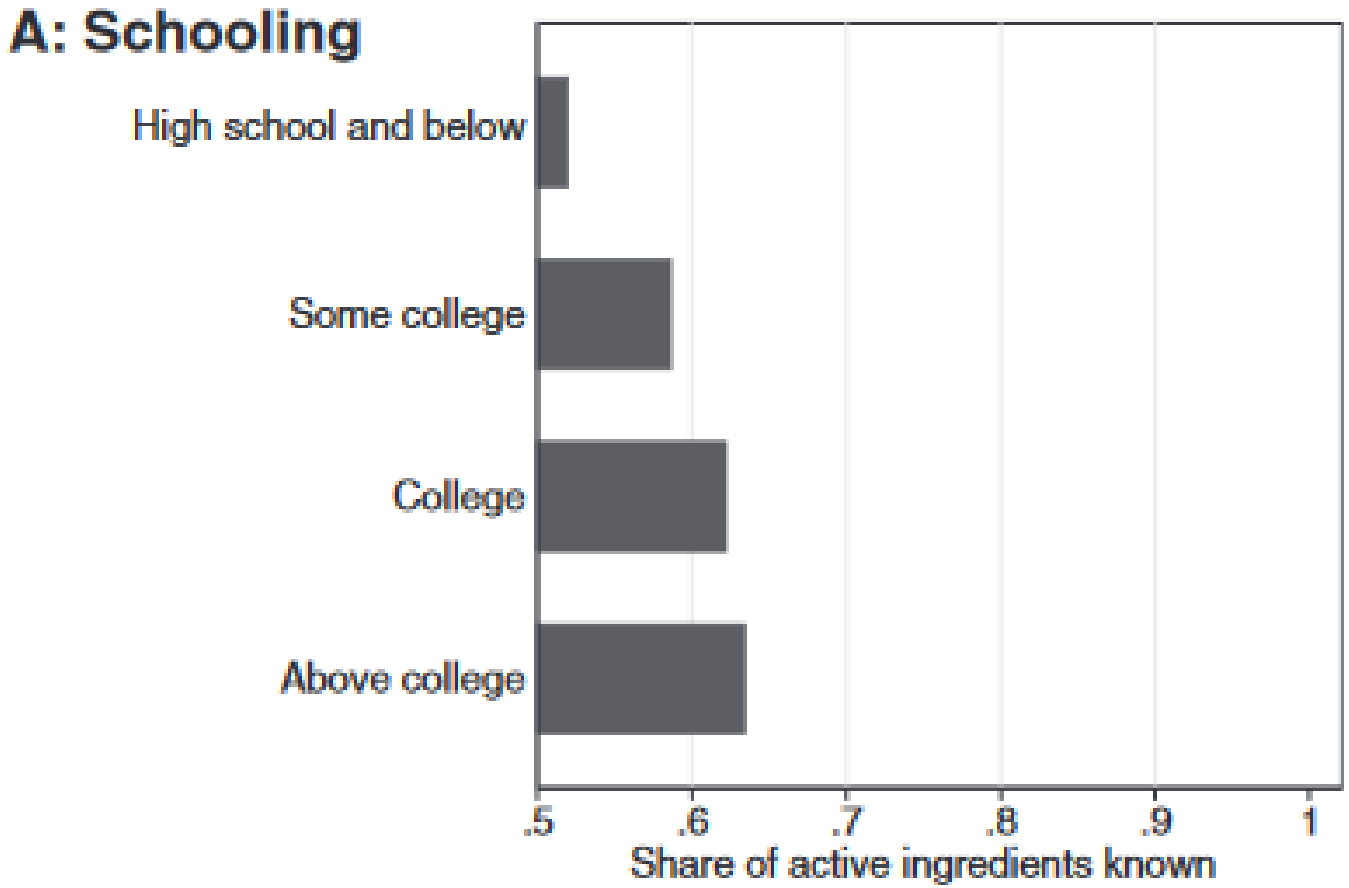
Painkillers

TABLE I

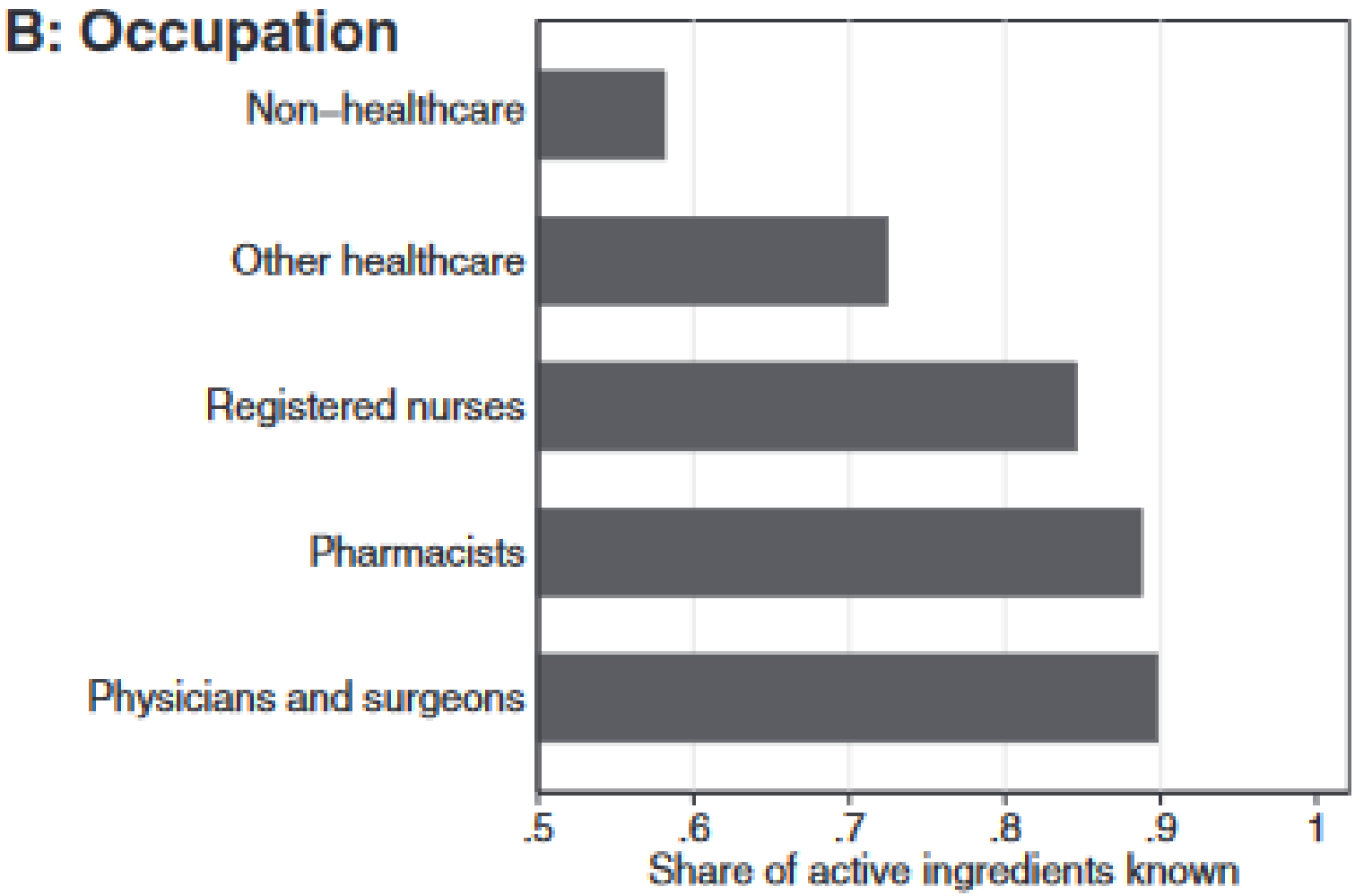
SUMMARY STATISTICS

	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, all (82)	10.87	0.58	0.47	0.54
Other health products, regression sample (44)	8.94	0.57	0.46	0.55

How informed are HHs?



How informed are HHs?



How informed are HHs?

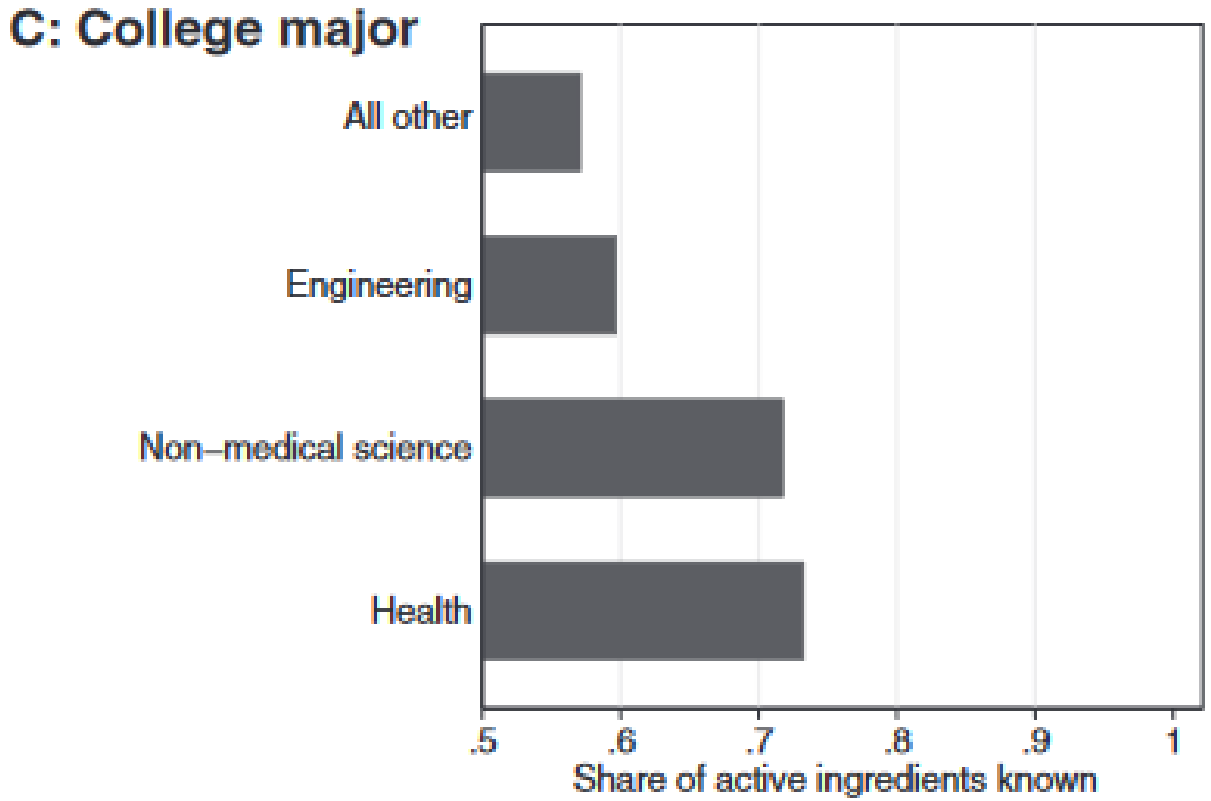


FIGURE I

Product Knowledge, Headache Remedies

Figure shows the mean share of headache remedy active ingredients correctly identified by each group of respondents in the 2011 PanelViews survey.

Information and purchase decision

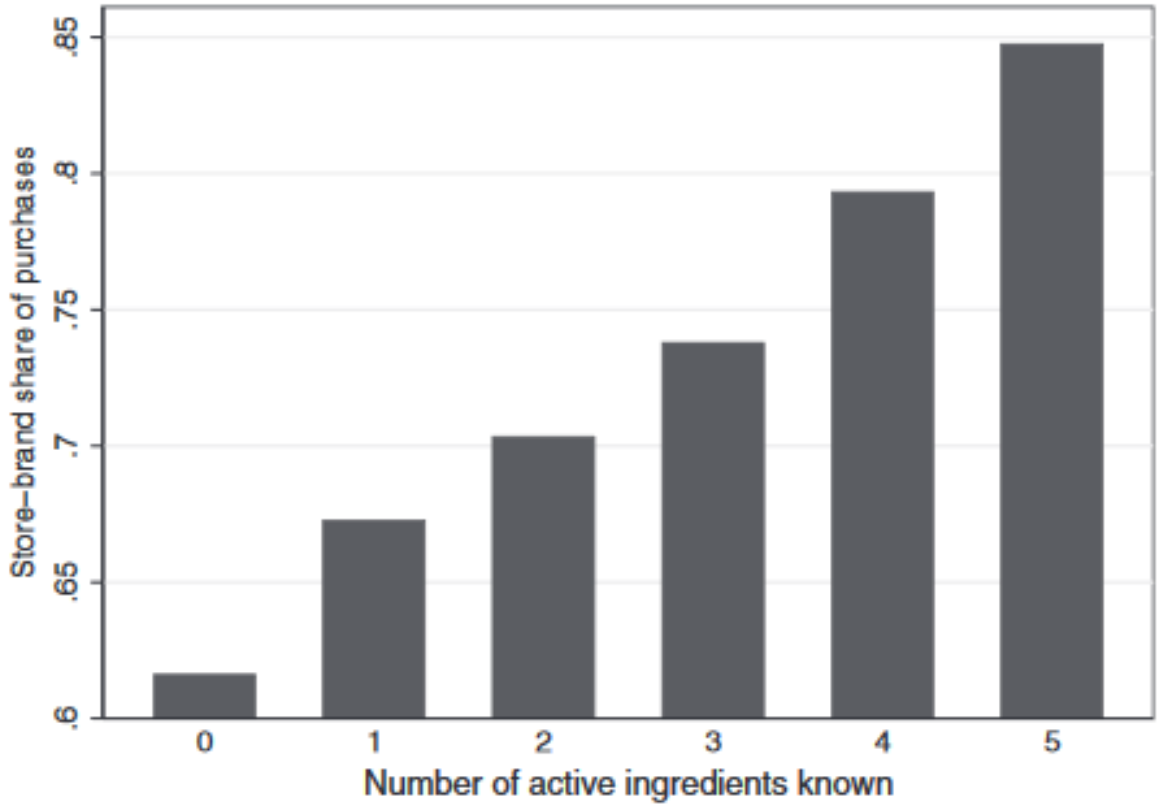


FIGURE II

Store-Brand Purchases and Knowledge, Headache Remedies

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.

Information and purchase decision

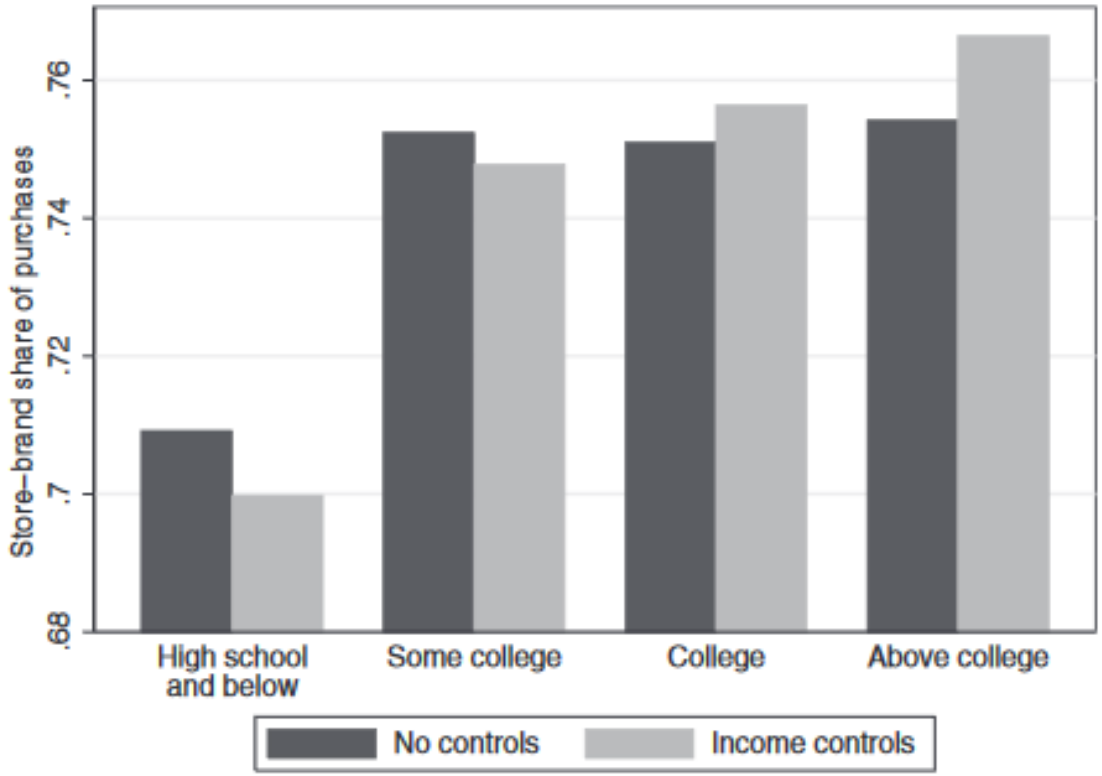


FIGURE III

Store-Brand Purchases and Education, Headache Remedies

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 19 household income categories, with the predicted values computed at the means of the covariates.

Information and purchase decision

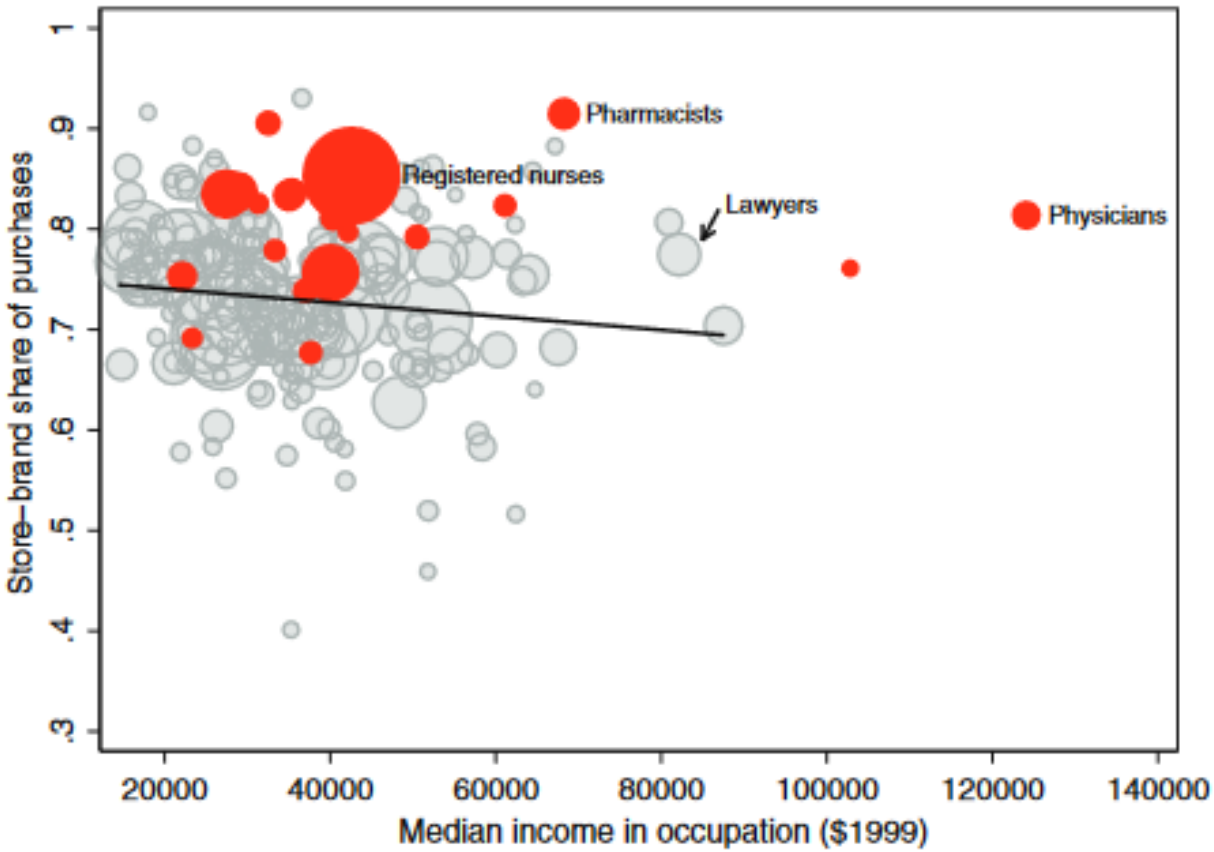


FIGURE IV

Store-Brand Purchases and Occupation, Headache Remedies

TABLE II

KNOWLEDGE AND HEADACHE REMEDY PURCHASES

Primary shopper characteristics	(1)	(2)	(3)	(4)	(5)
College education	0.0094	0.0206	0.0212	0.0255	0.0214
	(0.0072)	(0.0074)	(0.0075)	(0.0073)	(0.0068)
Share of active ingredients known	0.1792	0.1805	0.1805	0.1898	0.1463
	(0.0111)	(0.0112)	(0.0111)	(0.0108)	(0.0105)
Log(household income)		−0.0284			
		(0.0063)			
Believe store brands are "just as safe"					0.2058
					(0.0070)
Demographic controls?	X	X	X	X	X
Market and quarter fixed effects?	X	X	X		
Income category fixed effects?			X	X	X
Market-chain-quarter fixed effects?				X	X
Sample	Second survey wave	Second survey wave	Second survey wave	Second survey wave	Second survey wave
Mean of dependent variable	0.7392	0.7392	0.7392	0.7392	0.7392
R^2	0.1331	0.1351	0.1365	0.3561	0.3934
Number of households	26,530	26,530	26,530	26,530	26,530
Number of purchase occasions	195,268	195,268	195,268	195,268	195,268

TABLE III

OCCUPATION AND HEADACHE REMEDY PURCHASES

Primary shopper characteristics	(1)	(2)	(3)	(4)	(5)	(6)
College education	0.0171	0.0288	0.0351		0.0431	0.0133
	(0.0061)	(0.0064)	(0.0061)		(0.0100)	(0.0123)
Pharmacist or physician	0.1527	0.1683	0.1529	0.1667	0.1445	0.0304
	(0.0296)	(0.0294)	(0.0295)	(0.0380)	(0.0493)	(0.0379)
Other health care occupation	0.0792	0.0834	0.0790	0.0624	0.0489	0.0198
	(0.0099)	(0.0098)	(0.0102)	(0.0172)	(0.0224)	(0.0160)
Health major				0.0096		
				(0.0165)		
Non–health science major				0.0507		
				(0.0245)		
Demographic controls?	X	X	X	X	X	X
Market and quarter fixed effects?	X	X				
Income category fixed effects?		X	X	X	X	X
Market-chain-quarter fixed effects?			X	X	X	X
Sample	All	All	All	College major reported	Not currently employed	Second survey wave
Primary shopper survey response:						
Know all active ingredients						X
Store brands are "just as safe"						X
Mean of dependent variable	0.7424	0.7424	0.7424	0.7536	0.7390	0.8732
R^2	0.1166	0.1195	0.3037	0.4401	0.4330	0.6049
Number of households	39,555	39,555	39,555	14,190	13,479	4,274
Number of purchase occasions	279,499	279,499	279,499	92,020	103,624	33,373

NOTE: Column (6) subsample those who answered active ingredient questions correctly + believe store brands to be "as safe".

Brief look at pantry staples: Probability of buying store brand

- Columns 1 & 2 are pantry staples, column 3 headache remedy.

TABLE V

EVIDENCE ON DOMAIN SPECIFICITY

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 health care 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	29,561	44,502	39,555
Number of purchase occasions	404,372	588,484	279,499

Brief look at pantry staples: Probability of buying store brand

- The coefficients for *Pharmacist or physician* and *Other health care occupations* in Table V rule out effects greater than 5.2 and 2.2 percentage points.
- **Note**: This is an important way of studying the economic significance of estimated coefficients. How large/small effects can we rule out (based on statistical significance)?

Summary

- More informed shoppers buy more store and fewer national brands.
- This effect varies by category: It is large in health categories, and small in food and drink categories.
- Think of implications to e.g. (regulation of) packaging, advertising, ...