



opt to buy generic headache drugs more often. In fact, a doctor or pharmacist is 18 percentage points more likely than the typical shopper to buy a private-label headache remedy. The magnitude of the difference surprised one of the researchers, **Matthew Gentzkow**, Richard O. Ryan Professor of Economics and Neubauer Family Faculty Fellow at Chicago Booth. “The effects are really big across a lot of health-care categories,” Gentzkow says. The researchers estimate that if all US consumers were to start shopping like pharmacists, they could save a collective \$410 million a year on headache remedies.

Video: <https://www.youtube.com/watch?v=SS-Zu3PUyLs>

Around the world, billions of sales transactions every month, down to a can of Coca-Cola from a local store, are recorded in some way by Nielsen, the measurement and information firm that has been gathering data from retailers and consumers for 90 years. For most of its history, Nielsen shared those data primarily with its retail customers and manufacturing customers under strict agreements that protected customer confidentiality. Academic researchers gained access to some data by negotiating directly and often at length with Nielsen, or by partnering with a corporation and promising the data and results would be for internal use only. “We typically provided snippets of aggregate-level data [to academics],” says Frank Piotrowski, Nielsen’s senior vice president of measurement science and a 30-year company veteran. The situation was difficult for researchers—and for Nielsen, which wasn’t set up to easily release data to academics.

But in 2009, Nielsen executives were seriously thinking about how they could share data with researchers more systematically. The chairman of Nielsen’s Supervisory Board, James M. Kilts, who earned an MBA from Booth in 1974 and founded Booth’s **James M. Kilts Center for Marketing** in 1999, said he was interested in forging a partnership with the school. Working with Peter E. Rossi, cofounder of the Kilts Center and a former marketing professor at Chicago Booth, in 2010, Booth and Nielsen established a data warehouse under the umbrella of the Kilts Center as a central clearinghouse for the Nielsen information.

Now Nielsen is sharing three datasets through Booth, with a staggering amount of information. One dataset covers purchases by 40,000–60,000 households in the United States. Another contains sales results from 35,000 stores—grocery stores, drugstores, discount chains, and similar outlets—for the years 2006 through 2011. Those records span up to 3 million bar codes, and the data represent about 33% of the volume at mass merchandisers and about 55% of US retail volume from grocery stores and drugstores.

## The raw data

Nielsen has made three datasets available through Chicago Booth.

- **Consumer panel data:** These data come from households that agree to use a handheld, optical scanner to record the bar code of all purchases they make, whether at a grocery store, convenience store, discount chain, or other retail outlet. The consumer records each item's bar code, the quantity purchased, and whether he or she used a coupon to buy it. In many cases, the consumer also records the price paid if Nielsen doesn't already have it.

These purchases can be matched with detailed demographic information about a shopper, including his or her zip code, household income, birth date, age, education level, and occupation, which is then aggregated to deliver insights about various consumer segments. Individual purchases and information are kept confidential.

Researchers from any US-based academic institution can request consumer panel data about purchases made on more than 66 million shopping trips from 2004 to 2011. Although researchers can't see the name of the retailer selling each item, they know the type of store, its location, and whether the store is part of a chain.

- **Retail scanner data:** This fall Booth plans to release the Nielsen retail scanner data, which contains information from retailers about what they sell. The data include the average sales price for each item, on a weekly basis, in 35,000 stores across the United States. Researchers also can see how many units of competing products were available in each store. **Arthur Middlebrooks**, executive director of the **John M. Kilts Center for Marketing** and clinical professor of marketing at Chicago Booth, says it took months of negotiations to gain permission from retailers to use the data, but ultimately Booth was allowed to release the purchase information covering more than half of the grocery and drugstore outlets in the United States, and about one-third of the mass-merchandise locations.

- **Advertising and media data:** This dataset, which **Sanjay K. Dhar**, James H. Lorie Professor of Marketing at Chicago Booth, considers the crown jewel of Nielsen data, is available only to Booth researchers. These data show which consumer-products companies ran TV advertisements, in what markets the ads ran, and when they aired. By measuring the size of the TV market, Nielsen also can estimate how many people were reached by the ads.

The information now available is a gold mine for researchers, marketers primarily, but also economists who see the potential to explore longstanding questions about consumer behavior. For years marketers focused

their research on the coffee market, largely because the only good survey data available were household shopping diaries collected by local newspapers. Those diaries only covered a handful of product categories, including coffee. The data now available make coffee data look like a drop in a cup.

And the data reveal a lot about a lot of things, not necessarily about consumers themselves but about corporate actions and public policy. Some of the findings could have major implications for retailers, manufacturers, and society at large. The insights could help shoppers find the best time to buy certain products, but they could also provide ammunition for supporters of soft-drink taxes, or critics of government benefit programs. “We’ve still barely scratched the surface,” says [Jean-Pierre Dubé](#), Sigmund E. Edelstone Professor of Marketing at Chicago Booth.

### **Educated customers buy generic**

Here’s a conundrum that has long bothered economists: why do people pay more to buy branded products when generic alternatives are available? How much of a consumer’s decision to pay more for a branded product is driven by knowledge about its benefits, and how much is due to incomplete information?

Some researchers are seeking to answer such questions through field experiments. Mariana Carrera of Case Western Reserve University and Sofia Villas-Boas at the University of California, Berkeley wondered why branded over-the-counter drugs still command up to 60% of market share. They got permission from a national supermarket chain to change the shelf tags for several branded and generic drugs at six locations. They posted new labels beneath the items’ price tags, some pointing out that the branded and generic drugs contain the same active ingredient or are therapeutically equivalent. Other labels showed the percentage saved for buying the generic drug. The third set of tags indicated how many other customers purchased the generic.

From the four-week trial, the researchers conclude that customers are already aware that generic substitutes exist and save them money, so posting that information didn’t change their buying habits. However, learning that other customers bought the generic drugs did shift purchase decisions—when the percentage of shoppers buying the generic exceeded 50%.

The Nielsen data provide another way to approach the same question. [Gentzkow](#); [Dubé](#); [Jesse M. Shapiro](#), professor of economics at Chicago Booth; and [Bart J. Bronnenberg](#) of Tilburg University in the Netherlands compare different kinds of shoppers, whom they classify as either sophisticated or typical consumers. In addition to looking at how pharmacists buy headache remedies, they look at how chefs buy pantry staples, hypothesizing that sophisticated shoppers may find it easier to cut through the informational clutter created by commercials and branding to focus on a product’s most important attributes.

Using the Nielsen data, the researchers are able to compare purchases of branded and private-label products at not just six stores but thousands of locations. They construct an experiment, in a sense, after purchases

have been made, using a set of criteria to classify those who are sophisticated, well-educated consumers versus those who are not. Then they use the data to compare the buying habits of those groups of people. The data allows them to compare shopping patterns at the same chain in the same week in the same city, just as Carrera and Villas-Boas did in their field experiment, but with far more stores.

The researchers also survey consumers and, through Nielsen, pose their own questions to tens of thousands of households. (This is not part of the Kilts-Nielsen datasets, but a separate initiative arranged by the researchers through Nielsen.) They ask consumers to name the active ingredient—such as aspirin, acetaminophen, or ibuprofen—in a list of branded headache medicines. Overall, college-educated respondents name the right ingredient 61% of the time. The survey respondents' career choices matter far more than their socioeconomic status in determining whether they can be considered typical shoppers or sophisticated ones. For example lawyers, who have had a lot of schooling, don't necessarily know much about the ingredients in headache medicine. But for those who majored in science or health, the correct response rate rises to 73%. The figures continue to climb for registered nurses, at 85%; for pharmacists, 89%; and for physicians, 90%.

Armed with this survey information, the Booth-led research team looks at how doctors and pharmacists shop for headache remedies, a product category in which a branded drug can cost three times as much as the generic offering. As the respondents' knowledge level increases, so does their willingness to buy generic headache medicines. Nearly 90% of pharmacists' purchases of headache remedies are private-label products, compared with 71% for the total population.

A similar pattern emerges among chefs shopping for pantry staples such as salt, sugar, and baking powder. Chefs are 13 percentage points more likely than the average customer to buy the less-expensive, private-label items, to opt for the store brand of flour, say, rather than the branded flour. If everyone in the United States shopped like a chef, the researchers estimate, spending on branded pantry staples would drop by 24%, leading to a fall of \$20 million per year in the total amount spent on pantry staples.

Overall, shopping with more knowledge could cause a \$1.1 billion drop in annual spending, as customers save money through buying private-label goods. Some of the first to shift their preferences might be Gentzkow's MBA students, who told him they often bought branded Tylenol and Advil before learning about the new research.

### **Brand preferences are hard to shake**

The revelation about pharmacists grew out of an earlier question. Forgetting generics, when it comes to picking one brand over another—an area where marketers are constantly trying to sway our emotions—what if our preferences are actually stubborn, and tied to where we grew up?

That's the question Gentzkow set out to answer with Dubé and Bronnenberg, building on previously published research Dubé and Bronnenberg completed with [Sanjay K. Dhar](#), James H. Lorie Professor of Marketing at Chicago Booth, which analyzed Nielsen data provided by a large consumer-products company covering 30 categories of goods, including soft drinks and ground coffee.

As the researchers had hopscotched through the data from one US city to the next, they found broad, persistent, and inexplicable differences in market share. Take the instant-coffee brands Folgers and Maxwell House. Folgers dominates the West Coast, while Maxwell House leads the Atlantic Seaboard. When it comes to beers, Budweiser ranks highest in most US cities, but second in Chicago, behind Miller. Similar patterns hold in dozens of categories. "We thought these products would have had the same market share—because they're the same thing," Dubé says.

Baffled by the results, Dhar, Dubé, and Bronnenberg eventually combed companies' archives, searching for the date products launched in a particular city. By linking these dates to market share data that depict what happens after the brands launch in that city, simply looking at which brand launched first explains more than 80% of the geographic variance in several categories. In other words, the data show that a brand's "early-mover advantage," a notion Dubé had long been skeptical about, can persist for generations.

Dubé, Bronnenberg, and Gentzkow extend this research from the overall preferences of entire cities to the habits of individual shoppers. In a survey commissioned through and administered by Nielsen, they ask more than 38,000 households for each adult's current home location, country and state of birth, gender, birth date, education, and employment status, among other pieces of information.

Then the researchers match the survey results with the same group's grocery purchases, as recorded by the Nielsen consumer panel data, collected from shoppers themselves. The conclusion: 40% of the geographic variation in product market share can be explained by where customers were living when they formed their early shopping habits, and the effects can take years to dissipate. If shoppers are made aware of these unconscious preferences, they might become more rational in their purchases. For example, they might be more willing to use a coupon to try a new brand.

### **It's better to be good than cheap**

Coke and Pepsi dominate the soft-drink aisles of grocery stores, while RC Cola often languishes on a dusty shelf. It's easy to see which brands and companies lead their respective product categories, but it has long been difficult to explain why. Early-mover advantage has something to do with it, but so too do the strategies that follow a product's launch. Researchers at Columbia and Princeton are using the Nielsen data to investigate how firms establish their market share, looking at the prices and sales of products from thousands of retailers to get a bigger picture of how those companies do business.

The goal is to resolve a longstanding debate in business-strategy circles about why some firms grow huge while others don't. Economists have floated plenty of theories. Larger firms could be more efficient at holding down costs and selling their goods at cheaper prices, or they might have better advertising or higher quality products. Some might have established monopolies that allow them to charge big markups.

The three researchers—Columbia's David E. Weinstein, Columbia PhD student Colin Hottman, and Princeton's Stephen Redding—are combing through sales data from hundreds of thousands of bar codes in categories from food to electronics to appliances. The number of individual items they're studying is comparable to the inventory of a Kmart or Walmart megastore. "I had to buy a server to handle it," says Weinstein.

The researchers' approach to the question of why some firms grow dominant hinges on a method of measuring how consumers perceive product quality. The basic idea is that if two items sell for the same price, their brands' market shares should reflect the items' perceived quality. The brand of the higher-quality item would have a higher market share, and that of the lower-quality item would capture less share. If the products being compared sell for different prices, the researchers adjust their measurements accordingly. If one item is 10% cheaper than a similar item, the researchers predict its brand will have a slightly larger market share than its more-expensive competition.

After that weighting is complete, what's left over is the difference in market share that can't be explained by the formula, and that difference shows how the quality of a product impacts its sales. Quality, by this definition, is in the eye of the consumer.

For example, Coke and RC Cola are similar products, but Coke is more successful because people prefer it. Why they prefer it, whether they're influenced by advertising or other factors, is a secondary matter for the researchers. All they consider, besides price, is the perceived quality.

The early results are upending a long-held belief that the most successful companies are the ones that focus on reining in expenses or establishing economies of scale. "The key determinant of firm success is quality, much more than cost," Weinstein says. "The most successful firms produce goods that consumers think are high quality. It's not just about lower prices."

In the past, researchers could observe sales at the overall company level, but there was very little information about each product a firm was producing. Now economists can understand how each company operates at a minute level of detail. "It's a bit like going from studying atoms to suddenly being able to study subatomic particles, and seeing what you can learn from observing electrons or protons," says Weinstein.

## Small data, big challenges

As storage becomes ever cheaper and software programs to analyze data grow faster and more sophisticated, universities and businesses are tackling projects that would have been unthinkable a decade ago. The Kilts-Nielsen data does not meet the definition of “big data,” generally used to describe datasets so big that they’re difficult to work with using commonly available tools; however, it is the first initiative to give Chicago Booth’s technology department something approaching a “big data” problem.

The Nielsen dataset Booth released last year, covering consumer information gathered in people’s homes, fills only about 20 gigabytes, or less storage than a standard memory card for a digital camera. But the retail scanner dataset, to be released this fall, represents sales of 2.6 million bar codes from about 35,000 stores, covering more than 90 retail chains, from 2006 through 2011. The data fill 14 terabytes, or 14,000 gigabytes—the equivalent of 3,200 DVDs.

Under the terms of its agreement with Nielsen, Booth has to strip out the name of the retailer for each product sale. Booth also organizes the data to preserve the original characteristics of each product, and any changes over time. Retailers don’t care much if certain factors, such as the number of ounces in a carton of orange juice, change over time. They simply want to know what’s been selling in recent weeks. But researchers do care if product sizes change because their studies can cover months or years, and such differences can affect customer behavior.

It turns out that traditional backup technologies fall short for 14 terabytes of information. Booth’s Information Technology department typically uses a system that backs up data on magnetic tapes, but it can take weeks to store the retail scanner data on them, so the department had to learn to configure part of the same system being used for storage to handle backup and testing.

To successfully transfer Nielsen’s data to customers, **Darren Young**, a systems and security architect for Booth, borrowed ideas from colleagues who work with particle physicists and genomics researchers. Analysts at the University of Chicago’s Computation Institute recommended Globus Online, a large data-transfer product they developed, originally designed for moving enormous sets of particle physics and genomics data.

Weinstein has published other papers based on Nielsen data, including a widely noted study that uses sales of 10 million–20 million products and shows that overall grocery prices are lower in larger cities. The work validated the theories about international trade and economic geography that earned Princeton professor and *New York Times* columnist Paul Krugman the Nobel Memorial Prize in Economic Sciences.

Weinstein hopes to use data from individual bar codes of products sold at large numbers of stores to develop more-accurate measures of US economic inflation, which affects cost-of-living adjustments for workers and Social Security recipients, as well as Federal Reserve policies linked to mortgage and other interest rates. When calculating the Consumer Price Index, the Bureau of Labor Statistics (BLS) tries to create a representative basket of goods. The main difficulty with this approach is that the agency measures only prices, not quantities. As prices of items rise and fall relative to one another, consumers make



substitutions, and it isn't always clear whether the BLS is accurately reflecting these choices. Using the scanner data could suggest refinements to the index based on what people actually buy in a given month.

In a similar vein, Francisco Palomino and Robert Dittmar, both at the University of Michigan's Ross School of Business, and Ozge Sahin at Johns Hopkins Carey Business School are looking at companies' overall success by using the Nielsen data to measure "innovation risk" in approximately 10 industries. By noting when new bar codes appear in the Nielsen sales results, the researchers can see when a manufacturer introduces a new product. If that product has poor sales, a company may be wasting resources that could be better allocated elsewhere, Palomino says. The researchers' next step will be to track whether this innovation risk correlates to companies' stock prices.

### **Subsidies build brand loyalty**

Because they track buying patterns over time, the Nielsen data allow researchers a close-up view of how household purchases respond to changes in income or government benefits. Booth's Dhar, with Romana Khan of Özyeğin University in Turkey, and Ting Zhu of the University of British Columbia's Sauder School of Business, is examining the purchasing habits of families that enroll in WIC, a US federal nutrition program for women and young children.

The WIC program—formally called the Special Supplemental Nutrition Program for Women, Infants, and Children—promotes nutrition and healthy eating habits for low-income pregnant women, new mothers, and their children. The program was established in 1972, after doctors noted that pregnant women were coming to their offices with health problems caused by a lack of food. Administered by the US Department of Agriculture, WIC served about 9 million participants each month in 2012, including 53% of the country's infants.

In most cases, participants receive monthly vouchers, or electronic benefits cards to purchase foods from a restricted list designed to supplement their diets with specific nutrients. Their choices—which are determined by WIC offices in each state—generally include specific brands of cereal, whole-wheat bread, fruit, vegetable juice, eggs, milk, cheese, peanut butter, and canned beans. Food-package costs average \$45 a month, for an annual taxpayer cost of \$5 billion. The WIC program generally reimburses each retailer the full price of a participants' purchase.

While WIC is generally a short-term, two-year subsidy to improve the health of mothers and children, the researchers wonder whether its impact on behavior may persist beyond participation in the program, creating new habits and brand preferences. Once households lose their WIC benefits, their choice of what to buy is no longer limited to a list of WIC-approved items and brands. But does WIC get participants hooked on certain products?

To find out, the researchers use the Nielsen consumer panel dataset, in which participants self-report their purchases. That dataset also identifies individuals who have received WIC benefits, which lets the researchers track the buying habits of people before, during, and after they participate in the WIC program. The researchers start by looking specifically at breakfast cereals, a category that appears to benefit from WIC, as the program accounts for 7% of the \$9 billion cereal category, the researchers estimate.

Dhar, Khan, and Zhu are in early stages of their work, but what they have found confirms what previous, less data-intensive surveys have shown: the WIC program may have a lasting effect on consumer choice, changing behavior in ways that outlast participation in the program. After WIC, households spent more on cereal annually, a roughly 15% increase relative to what they spent before. Many households who bought WIC-approved cereals continued to buy the same cereal brands even after leaving the WIC program.

## The original data tracker

When Arthur C. Nielsen Sr. founded Nielsen in 1923, the US retail landscape was changing dramatically. Department stores were becoming dominant, destroying the near-monopoly that small, local dry-goods stores had enjoyed for much of the 19th century. Catalog retailers such as Sears, Roebuck & Co. and Montgomery Ward were revolutionizing the buying habits of small-town and rural families, who had never had such a range of options before.

Nielsen had developed his business conducting qualitative surveys for companies such as General Electric and DuPont, but he wanted to measure more precisely how firms were faring against their competitors. At the time, manufacturers knew how much merchandise they were shipping to stores, but didn't have a clear understanding of what customers were actually buying.

In the 1930s, Nielsen convinced grocery stores and drugstores to allow his auditors to collect store invoices, showing what flowed into the stores from suppliers and wholesalers. Then, at regular intervals, the auditors measured inventory in the stores to calculate how much had been sold. This retail measurement index, as Nielsen called it, established the concept of market share, which became a critical tool for companies to evaluate their performance.

Nielsen also found ways to measure new technologies. In 1942, the company created a radio rating service, using survey data to show which stations people were listening to. Then, in 1950, Nielsen began attaching recording devices to televisions in a sample of 1,200 US households. The devices used photographic film to register what TV channels the families were watching, and the families mailed the film cartridges to Nielsen, allowing Nielsen to calculate the market share of particular programs.

Arthur C. Nielsen died in 1980, when the data tracking he pioneered was still in its infancy.



Photo: Nielsen

Involvement in WIC has a particularly strong effect on participants whose income is in the bottom quintile of national income. The researchers find that those participants—who generally have poor nutrition, suffer high rates of obesity, and have bad health—are most likely to change their purchasing behavior during and after receiving WIC benefits.

But the data suggest the program also has a big impact on food manufacturers, showing why firms lobby intensively to have their products included on each state's list of WIC-approved foods. General Mills once revealed that in 1998, WIC accounted for 3% of the company's annual sales, say the researchers, though companies generally don't release their sales related to the WIC program. This early research suggests that WIC, beyond being a short-term subsidy for food companies, also affords these companies an opportunity to build brand loyalty and expand a market.

As Dhar, Khan, and Zhu dig further, they are asking related questions. Do WIC participants get better nutrition, achieving the program's goals? Does a subsidy for nutritious foods result in increased spending on nonnutritious foods? Do the healthier habits WIC supports persist, and for how long? That work is ongoing.

### **Government vouchers lead to higher prices**

Data about WIC participants may yield other insights. For example, are WIC program subsidies distorting the free market, affecting the prices for many consumers, even those not participating in the program?

For decades, shoppers who enrolled in the WIC program had to submit a paper voucher to get their groceries. The government reimbursed stores on a per-item basis, at an agreed-upon price or price range. That meant stores had an incentive to charge as high in that range as possible because the government had agreed to essentially pay that price. Stores weren't supposed to raise prices just to bilk the government, but there were few checks.

Similar cost-inflation pressures have existed in government health-insurance programs such as Medicaid, where doctors order extra procedures that don't cost the patient anything at the time of treatment. Doctors earn more, and taxpayers pick up the tab. Since WIC participants are unaffected by the cost of items, they don't care if a store raises prices.

Katherine Meckel, an economics PhD student at Columbia University, is studying how stores responded when the government cracked down on how much retailers could charge for WIC items. She's looking specifically at Texas, where the program's regulations became stricter between 2004 and 2009. In that state, stores are required to accept electronic benefits cards instead of vouchers. The electronic cards track purchases more precisely, so it's hard for retailers to substitute items or record inaccurate prices. Stores also have to agree in advance on what they would charge for certain brands of bread, milk, and other items on the WIC list. All states have to implement this new system by 2020, and so far six have done so. Texas was an early adopter.

Meckel is analyzing purchase data from the 6,000 Texas households participating in the Nielsen consumer panel dataset, which contains purchases made by the households at various retail stores. She's matching

that information with government data to identify 2,000 stores that accept WIC benefits. The dataset covers about 300,000 purchases made by roughly 6,000 households.

With the dataset, Meckel can observe whether a panelist is part of the WIC program, and she tracks product prices before and after the tighter surveillance of WIC took effect. While her results are preliminary, one early discovery she finds surprising: many consumers who were not WIC participants were paying higher prices because they shopped at stores that participated in the WIC program. An individual store set the price for an item and got reimbursed by the government. The store had an incentive to charge a high price to people shopping with WIC benefits in order to collect the maximum reimbursement possible. But as a result, the store charged all of its customers, even people shopping without WIC benefits, that same high price.

That said, she finds that the new regulations may help contain costs but also drive retailers away from the program, in part because they can be more easily punished for breaking rules. If a small store runs out of a WIC-approved brand of bread, it will get fined if it sells the customer a different loaf.

Indeed, many of the retailers dropping out of the WIC program are small stores in impoverished areas, and many are in cities. While tighter regulation saves the government money, it also makes it harder for WIC participants to use the program. Meckel says officials might consider relaxing rules for stores that have trouble complying because of their size, in order to preserve access for low-income families.

### **A soda tax could save billions**

Other behavior patterns in the Nielsen data could be used to promote better nutrition for a wider group of people. The issue: soft drinks. Public-health experts have long argued a link between soda and high levels of obesity, but the debate over how to discourage soda consumption has raged for a decade.

One perennially unpopular idea is to tax soft drinks, as consumers can buy these drinks tax-free in almost one third of US states. In New York, a proposal that would have imposed a 1¢-per-ounce soda tax failed to make it to the ballot in 2010. Soon after, in 2012, New York Mayor Michael Bloomberg took a different approach and proposed making it illegal for his city's restaurants, movie theaters, and stadiums to sell sugary drinks bigger than 16 ounces. The controversial ban was quickly tied up in court.

A new study provides more support for the notion of a soft-drink tax. University of Chicago economics PhD student Avigail Kifer uses Nielsen data to analyze the soda purchasing habits in 40 US counties, tracking millions of purchases. She matches those results with data from the US federal government's Centers for Disease Control and Prevention on each county's mean body-mass index, or BMI, a measure of body fat. She then compares the mean BMI of each county to the soda-buying habits of its residents, as

measured in the Nielsen consumer panel dataset, which asks households to record every item purchased. Through that data, she tracks consumption as well as the specific price paid for each bottle of soda.

Kifer finds that residents of the high-BMI counties in her sample were, on average, 16 pounds heavier than residents in the low-BMI counties. The high-BMI-county residents also exhibited stronger preferences for soda, both diet and regular. Soda cost less in those counties, and the residents drank more of it. This relationship, says Kifer, “further confirms the role of sugar-sweetened carbonated beverages in the obesity epidemic.” People in high-BMI counties consumed modestly smaller quantities of other sugary drinks, and a roughly equal number of low-calorie drinks.

The demand for both diet and regular soft drinks turns out to be fairly elastic—when the price goes up, people buy less of them. Moreover when soda prices go up, consumers don’t often substitute sugary drinks such as V8 Splash, lemonade, or orange juice, in part because they’re usually more expensive. Many switch to water.

Kifer finds that 3.5% of the weight difference between residents in high- and low-BMI counties may be attributable to differences in how much regular soda they drink. And she concludes that an excise tax of 3¢ per ounce would, by encouraging people to buy something other than soft drinks, eliminate the soda-specific weight gain. Such a tax would lead to a 5.8% drop in a county’s mean BMI over a 20-year period.

That could save a lot of money currently being put towards health care. An estimated \$190 billion has been spent treating obesity-related diseases, according to research by Cornell’s John Cawley and Lehigh University’s Chad Meyerhoefer. One study, commissioned last year by the Trust for America’s Health and the Robert Wood Johnson Foundation, and carried out by the National Heart Forum, predicts that a 5% decline in mean BMI by 2030 will allow the state of California to save \$81.7 billion and spare 796,430 Californians from developing type 2 diabetes. A drop of 5.8% in BMI, which Kifer predicts, could save a good chunk of change.

### **Given information, people eat healthier**

If a consumer drinks a 30-ounce sugary soda, she may do so knowing full well that it’s unhealthy. But in many cases it can be hard to identify healthy foods. Nutritional data presented on a product label can be difficult to decipher, and nutritional claims made in commercials or on a product’s packaging can be misleading. A person buying a food because it claims to be low-calorie or low-fat may be unknowingly opting for a less-healthy option within a product category. Low-fat peanut butter, for example, may be higher in sodium than regular peanut butter.

To make it easier to identify healthy foods, in 2008 a group of nutritional and medical experts, some affiliated with the Yale University School of Medicine, created a new nutritional scoring system, NuVal. NuVal employees go to grocery stores and scan the labels of tens of thousands of products to record their

nutritional information. Then NuVal's algorithm spits out scores that range from one to 100 and are easy to compare. Keebler Townhouse Bistro Multigrain Crackers, for example, get a NuVal score of three. Fresh broccoli florets get a perfect score of 100.

Several large chains, slightly more than two dozen so far, are posting NuVal scores on shelf price tags. The people behind NuVal hope that the scores, which measure relative nutrition rather than calorie counts, will help people, particularly those who might otherwise get swept up in fad-based diets.

The question is whether easy-to-read nutritional information will change shoppers' purchasing habits. After all, everyone should already know that broccoli is healthy, and yet many choose to eat something else. But Dubé; [Günter J. Hitsch](#), professor of marketing at Chicago Booth; and Booth PhD student Jin Zhang have early evidence that the scores may work as intended.

The researchers stress that their work is in its early stages, but they have started by looking at data for purchases of refrigerated yogurt. They use two datasets. The one purchased from NuVal shows that the mean score for refrigerated yogurt products is 48, and the standard deviation is 28. There's a 91-point difference between the least- and most-healthy yogurt, a large enough difference to get a shopper's attention. To find out what consumers do with this knowledge, the researchers use their second dataset, the Nielsen consumer panel, observing purchases before NuVal scores became available as well as after. They observe the behavior of households that shop in stores posting NuVal scores, and that of households that shop elsewhere.

The researchers consider their conclusions preliminary, in part because NuVal data were only introduced in January 2009, and in part because they are refining how they define a control group. But Zhang says the researchers can claim, conservatively, some evidence that a consumer presented with NuVal yogurt scores will pick a healthier option on the shelf. Once that is further confirmed, they have more questions. Will a consumer who's seen NuVal scores remember those when they go to other stores that don't display the scores? Will they remember and act on what they've learned? For now, that remains hidden in the data.

## What about privacy?

Few shoppers want academics trailing them in the grocery store any more than they want government minders listening to their phone calls. But privacy advocates haven't yet made a fuss about the Nielsen data, and for good reason. The Nielsen data are both incredibly precise and anonymous to personal information. Researchers can compare the behaviors of shoppers in a single store, in a single week, and they can know exactly what is on shelves (provided consumers purchased the product that week). They can see sales results from individual stores, by week, by item, and by individual household. Yet they don't have the retailer name or full zip code of a store to identify it, much less shoppers individually. On top of that, the data are subject to a one-year lag.

### How to get a good deal on champagne

The Nielsen datasets are still new to Booth. “We’ve been in development mode for about three years, getting things put together and cleaned, organized, and validated,” says [Arthur Middlebrooks](#), executive director of the Kilts Center and clinical professor of marketing at Chicago Booth. The scanner data are being made available to non-Booth researchers this fall, while the media data will still be available only to Booth researchers. The center’s goal is to release an updated dataset to researchers annually. In January 2014, Middlebrooks says, the Kilts Center hopes to release Nielsen data collected from and about consumers in 2012.

Dhar predicts a wave of new research will be published in the next year, as some of the earliest projects launched when the data became available near completion. Of the more than 70 research projects under way, one could suggest the best time to buy certain products. Juliette Caminade and Denrick Bayot, University of Chicago PhD students in economics, are using the Nielsen retail scanner data to examine the prices of goods for which demand shoots up during particular holidays. They focus on sparkling wine, as 25% of all champagne sales occur during the last two weeks of December.

You might expect retailers to take advantage of the demand to raise prices, Caminade says. But if you buy during off-peak times, trying to outwit the stores, it turns out you’re actually wasting money. Prices of sparkling wine fall about 10%—across all brands and products—when consumers want it most, according to the data. The researchers are primarily interested in why this effect occurs, so they’re looking at Nielsen’s consumer panel information to tease out whether the demographics of people who buy sparkling wine during the rest of the year are different from those who only buy it during the holidays. Preliminary results show that year-round champagne purchasers are more affluent, making them less price-sensitive than the holiday drinkers. But the questions continue. Next time you pop a cork, you may contribute to the answer.



## Have a question?

Nielsen's retail scanner and consumer panel datasets are available through Chicago Booth's James M. Kilts Center for Marketing to approved academic researchers, which includes US-based tenured and tenured-track faculty, plus PhD students under their advisement.

To obtain a subscription to the data, interested researchers need to first subscribe and register. Subscriptions are available on individual and institutional bases. The Kilts Center does not make a profit by providing Nielsen data to academic researchers, but it does charge modest subscription fees to recover some of the costs of processing, storing, cleaning, and distributing the data.

See [research.ChicagoBooth.edu/nielsen](http://research.ChicagoBooth.edu/nielsen) for more information.

## Works cited

Bart J. Bronnenberg, Jean-Pierre Dubé, and Matthew Gentzkow, "The Evolution of Brand Preferences: Evidence from Consumer Migration," *American Economic Review*, 2012.

Bart J. Bronnenberg, Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M. Shapiro, "Do Pharmacists Buy Bayer? Sophisticated Shoppers and the Brand Premium," Working paper, June 2013.

Mariana Carrera and Sofia Villas-Boas, "Generic aversion and observational learning in the over-the-counter drug market," Preliminary draft, April 2013.

Avigail Kifer, "A Hierarchical Bayesian Approach to Soda Demand Estimation: Explaining Cross-County Differences in BMI," Working paper draft, June 2013.

Katherine Meckel, "Cost Containment vs. Participant Access in the WIC Program," Working paper draft, June 2013.

David E. Weinstein, Colin Hottman, and Stephen Redding, "The Sources of Firm Heterogeneity," Preliminary draft, June 2013.

Artículo on-line:

<http://www.chicagobooth.edu/capideas/magazine/fall-2013/secrets-of-shopping>