WHITE PAPER

Upgrading Market Research with TrueNorth®: A Case Study on Consumer Purchasing Behavior

NORC
at the UNIVERSITY of CHICAGO

Table of Contents

Introduction

Market research is a multibillion dollar industry that informs critical business decisions across the economy. This industry largely relies on nonprobability online panels for the data used to guide branding, advertising design, and audience strategy.

Nonprobability panels are appealing for multiple reasons. First, the cost of traditional probability-based methods has been, and continues to be, increasingly expensive. The cost per interview for a 10-minute survey on a nonprobability panel can be as low as $3, compared with considerably higher prices for a telephone or face-to-face probability survey. Second, market research is often interested in the attitudes and behaviors of low-incidence, hard-to-find populations. Nonprobability panels can offer a quick and easy way to find these target audiences on a large scale.

But relying on nonprobability panels comes with drawbacks. Research continues to show that despite declining response rates, probability-based surveys are more accurate than nonprobability approaches (Baker et al. 2010; Dutwin and Buskirk 2017, 2020; MacInnis et al. 2018). Indeed, in this paper, we find that nonprobability samples consistently overstate the percentage of people who purchase a product. These misleading estimates can have a major negative financial impact if not corrected. Moreover, a lack of standards among nonprobability panels for how to recruit respondents, as well as minimal transparency about how different panels do so, means it is hard to be an informed consumer and know which panels to choose. Results from nonprobability panels can be highly unreliable, providing different point estimates from one survey to the next (Kennedy et al. 2016).

Given these tradeoffs and concerns, we investigate in this paper how to take advantage of the best of both approaches—the scale and affordability of nonprobability samples with the accuracy and reliability of probability-based surveys. We are not the first to blend probability and nonprobability data, but we have set what we think is the industry standard with our highly advanced new calibration approach—TrueNorth® Calibration. TrueNorth reduces the bias of nonprobability samples at not only the topline level, but also deep within key demographic groups (Ganesh et al. 2017; Gupta et al. 2019; Yang et al. 2018). Moreover, the approach is tailored to the particular topic of each survey to reduce bias that might be specific to the given survey.

In this paper, we address the following research questions as they concern market research topics:

1. How do the results from nonprobability surveys really compare to high-quality probability surveys?

2. How do nonprobability surveys compare to each other, meaning, does the sample choice matter?

3. Can a calibration approach help reduce the bias in nonprobability surveys?

Methodology

To evaluate these research questions, we conducted a study using NORC's probability-based AmeriSpeak® Panel. Since NORC at the University of Chicago developed it in 2015, AmeriSpeak has produced more than 500 surveys, been cited by every major U.S. media outlet,[1] and become the primary survey partner of the nation's preeminent news service, The Associated Press.

AmeriSpeak's sampling captures a true picture of America, providing better representation than other probability-based panels for hard-to-reach populations, including low-income households, less-educated persons, young adults, rural households, persons who are less interested in the news, and social and political conservatives (Bilgen et al. 2019).

Here's why the approach yields more accurate research results:

- **The panel has an industry-leading recruitment response rate of 34 percent.** AmeriSpeak knocks on the doors of individuals who don't respond to initial outreach to ensure that the panel includes hard-to-reach populations. And people who were not randomly selected are prohibited from volunteering to join the panel.
- The panel's random sample of Americans **looks almost exactly like the U.S. Census**, within 1 to 2 percentage points on key demographics.
- AmeriSpeak has **low participant turnover and survey fatigue**, with an industry-leading retention rate of about 85 percent. To manage the panel, AmeriSpeak limits respondents to one survey invitation per week.
- Over 10 percent of Americans still do not use the internet, and such Americans are quite different from internet users on a range of important metrics. **AmeriSpeak is the only probability panel to systematically cover both those who do and do not use the internet**.

In addition to 1,000 respondents from the AmeriSpeak Panel, we surveyed 800 respondents each from three nonprobability online panels. Two panels—Sample A and Sample B—are "traditional" nonprobability panels that recruit their participants via advertisements, pop-ups, and corporate partnerships, and then maintain their own panel. The third panel—Sample C—is a panel aggregator that collects responses from multiple nonprobability panels. Each panel was asked to balance survey respondents based on geography, race/ethnicity, and education level, as well as age by sex, according to the latest Census distributions.

The questionnaire asked respondents whether they purchase a range of items, including food products (nine items), beverages (eight items), retail-store beverages (15 items), energy bars or other health items (five items), and personal care products (five items). The survey was a median length of five minutes.

[1] AmeriSpeak research has been cited in major media outlets, including *Time Magazine*, *The New York Times*, *CNN*, *Forbes*, *Chicago Tribune*, and others. Our clients and sponsors include prestigious organizations such as AARP, Centers for Disease Control and Prevention, Consumer Reports, Legal Services Corporation, National Science Foundation, NASA, National Institute of Justice, Pew Research Center, Truth Initiative, and dozens of university-based researchers, among others.

Weighting Approaches

In addition to evaluating sample source, we analyze the impact of weighting on bias reduction. We compare a conventional weighting approach—weighting to key demographics—with our alternative, TrueNorth Calibration.

A standard approach for weighting nonprobability surveys is to first use basic demographic targets as quotas for who can complete the survey. Providers can then weight the survey to key demographics or other attitudinal or behavioral markers, such as whether the person is an early adopter of technology. For this analysis, all three samples have been weighted to February 2019 Current Population Survey distributions for age, sex, Census division, education, and race/ethnicity.

For the AmeriSpeak Panel results, we used our standard approach to weighting, which incorporates both the panelist's probability of selection into the panel as well as raking to the same set of demographic targets outlined above for the nonprobability panels.

To implement TrueNorth Calibration, we combine the probability sample from the AmeriSpeak Panel with the nonprobability sample, and then calibrate it using small area estimation. The process is as follows:

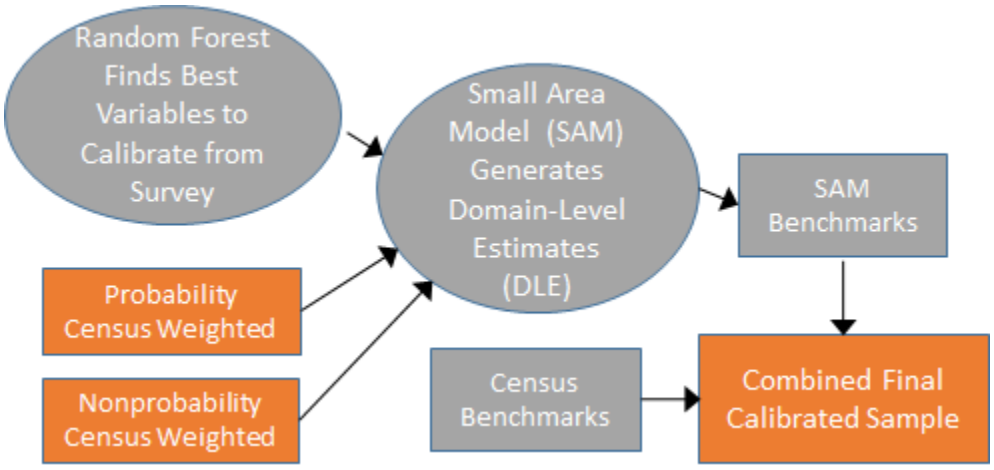
1. First, we identify two to four key analytical variables from the survey to target for reducing bias. These are identified using a machine learning approach called random forest modeling.

2. Second, we establish 20-40 domains in the data, where each domain is a specific, relevant subgroup. We typically use demographics as a starting point to create the domains, such as African-American males age 18 to 34 with a college degree. We create even deeper dimensionality by also defining domains with key questions that are specific to the particular study, such as whether respondents purchase a product or live in a particular market area.

3. Third, within each domain, we run a small area model (SAM) for the analytical variables identified in Step 1. These models generate benchmarks for calibration.

4. Finally, the combined data is calibrated to these SAM benchmarks as well as to standard Census demographic benchmarks.

The graphic on the next page illustrates the process.



Results

To analyze the impact of sample source and calibration, we focused on two key topics in marketing research. First, we evaluate how our sample and calibration approach affect data on **consumer purchasing patterns**, in terms of what items consumers are buying. Second, we evaluate how the data choices affect the results for the **market profile** of consumers who have specified that they purchase a particular product. This second analysis is particularly crucial, as this type of information guides marketing campaigns, branding choices, and ad buys.

Consumer Purchasing Patterns

To evaluate consumer purchasing patterns, we analyze questions with a yes or no answer to whether the respondent buys a product regularly. The table on the next page shows the results for one of the items—Sierra Mist—asked about as an example. The table includes results for the probability-based AmeriSpeak Panel and for each nonprobability panel used. We include percentages for:

- the weighted AmeriSpeak probability sample
- each nonprobability sample separately, using a conventional demographic weighting approach
- each nonprobability sample combined with the AmeriSpeak sample, using a demographic weighting approach
- each nonprobability sample combined with the AmeriSpeak sample calibrated, using our TrueNorth approach

To analyze the accuracy of the different approaches, we use the absolute error between the sample of interest and the AmeriSpeak Panel results. This is defined as the absolute value of the difference between the sample being evaluated and the AmeriSpeak results. We also analyze the range across the nonprobability samples—meaning the difference between the maximum percentage of the three samples and the minimum percentage.

Results for Sierra Mist			
Sample	**Percentage**	**Absolute error: Nonprob - AS**	**Nonprob range: Max - Min**
AmeriSpeak	21%	--	--
Weighted nonprobability alone			
Sample A	27%	6	5 points
Sample B	32%	11	
Sample C	30%	9	
AmeriSpeak+Nonprobability: Demographic weighting			
Sample A	23%	2	3 points
Sample B	26%	5	
Sample C	25%	4	
AmeriSpeak+Nonprobability: TrueNorth Calibration			
Sample A	22%	1	2 points
Sample B	24%	2	
Sample C	22%	1	

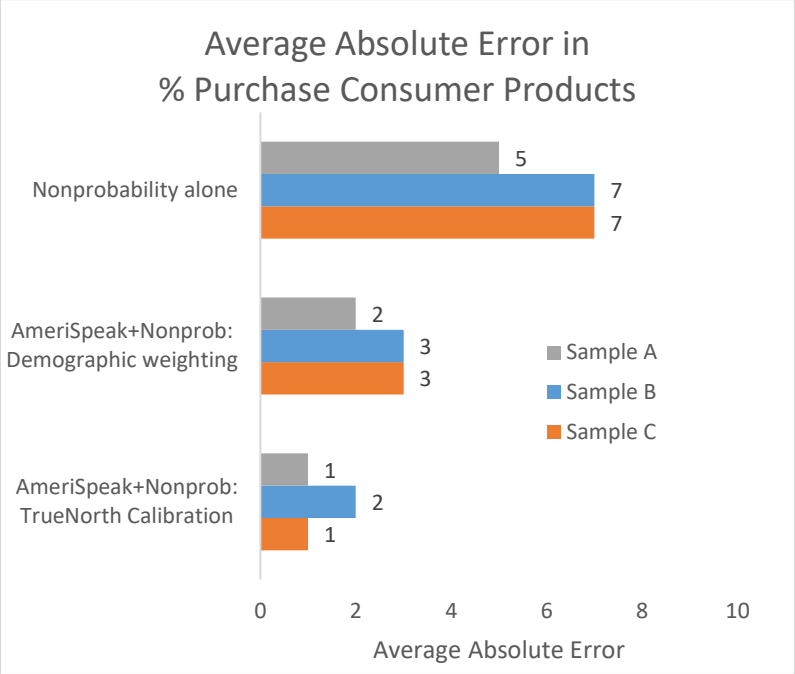
There are a few key patterns to notice from this table. First, the nonprobability samples *consistently overstate the percentage of people who purchase a product*, even though the samples are weighted to key demographics. For Sierra Mist, the absolute error ranges from 6 percentage points to 11 points.

Second, there can be *considerable variability in the results across the nonprobability samples*. The difference between the minimum and the maximum percentage across the three nonprobability samples with conventional demographic weighting for Sierra Mist is 5 percentage points.

Third, combining the probability and nonprobability samples using conventional demographic weighting reduces the bias in the percentage of people who purchase a product. The average absolute error drops from a maximum of 11 percentage points for the separate nonprobability samples to a maximum of 5 points for the combined probability and nonprobability samples with demographic weighting.

Finally, combining the probability sample with the nonprobability sample using our TrueNorth Calibration reduces the bias even more. **For the TrueNorth approach, the maximum absolute error for the combined calibrated samples is 2 percentage points.** TrueNorth also tends to reduce the variability among the nonprobability samples. The variation in the percentage across the three samples for Sierra Mist is now 2 points instead of 5 points.

These patterns hold for the other items we asked about. The graph compares the average absolute error for each nonprobability sample in the percentage of respondents who report buying a product for all 42 items asked about.



TrueNorth Calibration reduces the absolute error for all 42 products for both Sample B and Sample C and for 35 of the 42 products for nonprobability Sample A.

The TrueNorth approach also reduces the variability across the three nonprobability samples. The average difference for all 42 items between the highest and lowest percentages among the three nonprobability samples weighted to demographic targets is 6 percentage points; among the TrueNorth results, the average drops to 2 points.

Another way to look at this is to consider the number of instances where the differences between the sample of interest and the AmeriSpeak Panel are *statistically significant*. For the conventionally weighted nonprobability samples on their own, the results are significantly different from the AmeriSpeak results for 91 point estimates out of 126 total (42 items across three samples). When the nonprobability samples are combined with the AmeriSpeak Panel and weighted using demographic targets only, the number of point estimates with statistically significant differences declines to 59. TrueNorth Calibration further reduces the bias in the nonprobability samples: the number of point estimates with significant differences drops to just 20 out of 126.

Market Profiles

Perhaps more important than knowing the percentage of people who purchase a particular product, is knowing *who* purchases the product. Even if nonprobability samples can sometimes closely approximate the topline result of a probability sample, the results for demographic subgroups can be wildly different (Kennedy et al. 2016). Being able to improve the demographic analysis from nonprobability samples would be a significant contribution to market research.

To evaluate the demographic profiles produced by probability and nonprobability samples, we analyze the demographics of those who say they purchase a product. Meaning, we evaluate what percentage of those who say they buy a product are, for example, young people. Knowing the demographics of your market guides your advertising design, spending, and even branding.

The table below includes the percentage of those who reported purchasing nail care products who are 18-29 years old and those who are 55 years or older. Again, we begin by walking through a single item as an example and then provide a summary of the results across all 42 items included on the survey. As in the previous section, we show the results for the weighted AmeriSpeak probability sample, for the nonprobability samples using a demographic weighting approach, for the combined probability and nonprobability samples weighted to demographics, and finally, the results for TrueNorth Calibration.

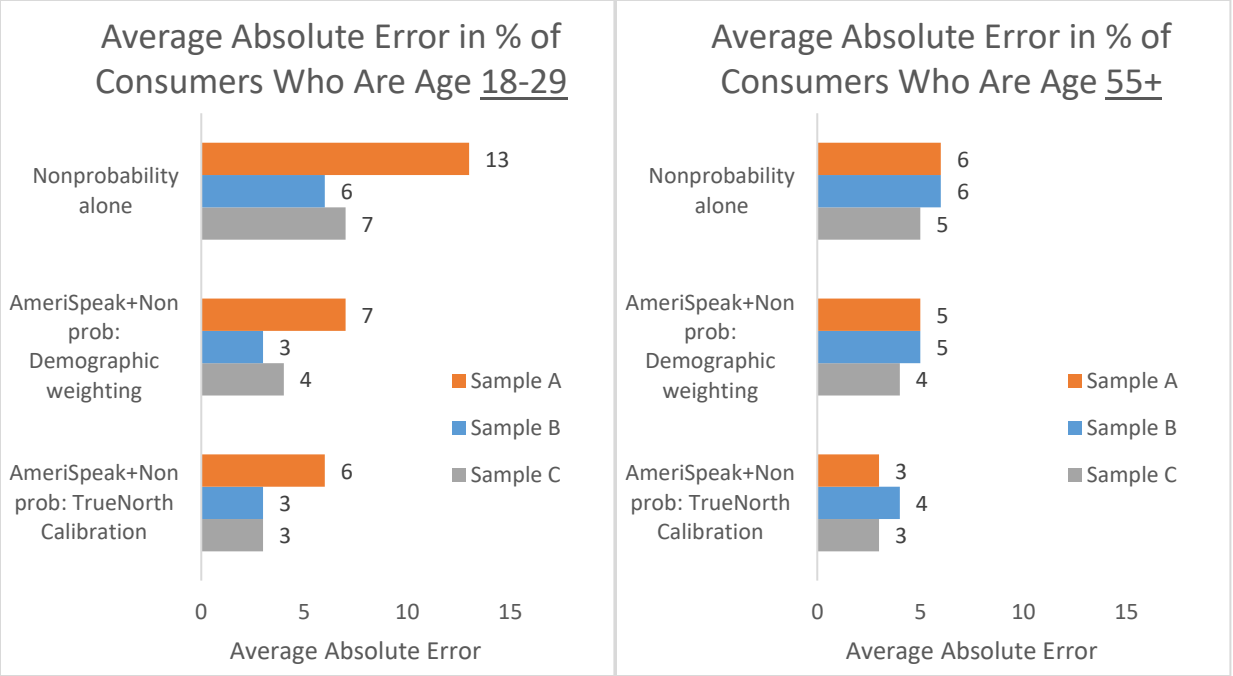
Results for Nail Care Products						
	% of purchasers who are 18-29			% of purchasers who are 55 or older		
Sample	Percentage	Abs. error: Nonprob - AS	Nonprob range: Max - Min	Percentage	Abs. error: Nonprob - AS	Nonprob range: Max - Min
AmeriSpeak	17%	--	--	31%	--	--
Weighted nonprobability alone						
Sample A	30%	13	5 points	25%	7	4 points
Sample B	26%	9		26%	5	
Sample C	25%	8		29%	3	
AmeriSpeak+Nonprobability: Demographic weighting						
Sample A	23%	6	2 points	25%	6	2 points
Sample B	21%	4		26%	5	
Sample C	21%	4		27%	4	
AmeriSpeak+Nonprobability: TrueNorth Calibration						
Sample A	22%	5	3 points	29%	2	2 points
Sample B	19%	2		29%	2	
Sample C	19%	3		31%	0	

Nonprobability samples can produce an even more divergent picture when it comes to demographic profiles. The absolute error relative to the probability sample for the percentage of nail care product purchasers who are 18-29 years old ranges from 8 points to 13 points. For those 55 or older, it ranges from 2 to 6 points.

The nonprobability samples also exhibit considerable variation among them when it comes to demographics. For 18-29 year olds, the percentage of purchasers can be 30 percent with conventional demographic weighting for Sample A or 25 percent for Sample C.

When we apply TrueNorth Calibration, we both reduce the absolute error between the nonprobability and probability results, and reduce the variation across the nonprobability samples, as was the case at the topline level. For nail care products, the absolute error for the TrueNorth approach ranges from 2 to 5 points for 18-29 year olds—a third of the error for nonprobability samples alone. And the variation across the three nonprobability samples declines from a range of 5 points to 2 points.

The graphs on the next page summarize the average absolute error for the two age groups across all 42 items asked about on the survey.

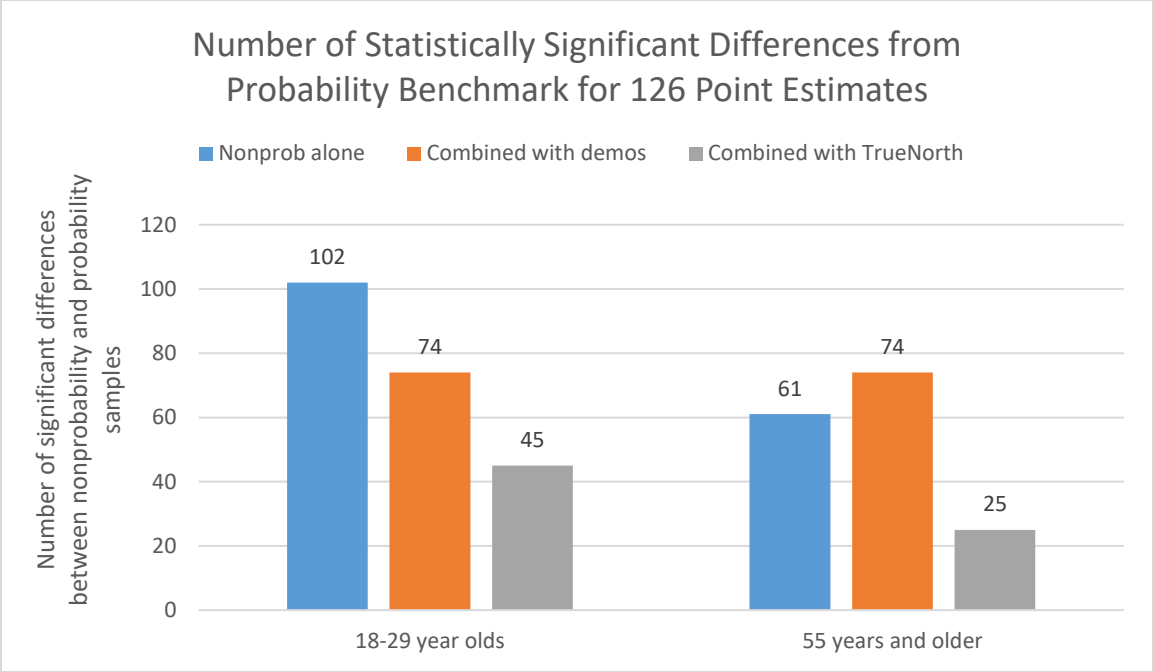


The patterns described above for nail care products hold for other items asked about. On average, the absolute error for the percentage of young people among those who purchase a product is between 6 and 13 percentage points depending on the nonprobability sample. This reduces to 3 to 6 points with the TrueNorth approach. For this group, TrueNorth cut the error in half. And for those age 55 or older, the error goes from 5 to 6 points down to 3 to 4 points.

The variability among nonprobability samples declines as well—from an average of a 9 point difference between the maximum and minimum percentages in the three samples for 18-29 year olds to an average of 4 points with TrueNorth Calibration.

Choosing the “best” nonprobability sample provider also may not be possible. On average, Sample B is most accurate (lowest error) when it comes to purchasers who are age 18-29, but less accurate than Sample C for those who are 55 or older. There is no clear guidance from these results as to which sample is the right one, or at least the one with the consistently lowest error.

These averages mask another important pattern—across the 126 point estimates, 102 items had statistically significant differences between the nonprobability sample and the AmeriSpeak probability sample for the percentage of purchasers who are age 18-29, and 61 items had significant differences for purchasers age 55 and older. With TrueNorth Calibration, the bias is substantially reduced—to 45 and 25 items, respectively, with significant differences. In sum, for nearly two out of three cases where the nonprobability results were stastically different from the probability results, TrueNorth was able to bring the results in line with the AmeriSpeak standard.



Discussion

There are multiple takeaways from this investigation. First, nonprobability samples can provide very different results from what researchers would get with a high-quality probability sample. These differences are significant for topline percentages, and perhaps more importantly, for demographic analysis.

Second, using TrueNorth Calibration helps tremendously to reduce bias at both the topline level and for key demographics. With TrueNorth, market researchers can be assured that they have the right market profile to guide key decisions about ad buys, marketing designs, and branding.

Third, nonprobability samples exhibit a great deal of variability. This means that the nonprobability panel chosen can have substantial impact on the conclusions researchers draw. And without transparency on nonprobability panels' methods or clear guidance on which panel has the lowest error, it is not possible to know *a priori* which panel to choose. A key advantage of TrueNorth Calibration is that it reduces the variability in results across different nonprobability samples, making this sample choice less consequential.

Moving away from using nonprobability panels to an approach that relies on the quality of a probability sample and the innovative calibration approach, TrueNorth has multiple advantages for market researchers. In particular, with TrueNorth Calibration, market researchers can improve the conclusions they draw about the demographic profile of their target market. They can also improve the reliability of their results, both across nonprobability sample providers and across samples from the same provider. **By getting your research right, you can be confident in your business decisions.**

Appendix A: Questionnaire

S_FOOD.
Have you purchased, or not purchased, any of these food products for your personal use in the last 12 months?

Some of these are not common food products, so please select yes only if you are absolutely sure you purchased the product.

Please indicate Yes or No.

GRID ITEMS, RANDOMIZE:

1. Almond oil
2. Hemp oil
3. Saffron spice
4. Shrimp paste
5. Truffles
6. Turmeric spice
7. Vanilla extract
8. Wasabi root
9. Szechuan peppercorns

RESPONSE OPTIONS:

1. Yes
2. No

S_PERSONAL.
Please indicate which, if any, of these items you have purchased for your personal use in the last 12 months.

Please indicate Yes or No.

GRID ITEMS, RANDOMIZE:

1. Nail care products (like hardener and polish)
2. Hair care products (like shampoo, hair care kits, conditioner)
3. Foot care products (like creams and orthotics)
4. Skin care products (like lotions and cleansers)
5. Oral care products (toothpaste, mouthwash, toothbrush)

RESPONSE OPTIONS:

1. Yes
2. No

S_POWDERS.
Please tell us which products you purchased or did not purchase in the last 12 months.

Please indicate Yes or No.

GRID ITEMS, RANDOMIZE:
1. Energy bars
2. Energy gels
3. Sports drinks
4. Protein bars
5. Protein supplements (powders)

RESPONSE OPTIONS:
1. Yes
2. No

S_BEVER1.
Please indicate which, if any, of these bottled or canned beverages you purchased in stores for your personal use in the last 12 months.

"Stores" include retail stores like a convenience, grocery, or discount store (such as a Target or Walmart). Please do not count any purchases you might have made at a restaurant.

Please indicate Yes or No.

GRID ITEMS, RANDOMIZE:
1. Beer or wine
2. Energy drinks or shots
3. Coffee and tea
4. Cocktail mixes
5. Juice and cider
6. Protein shakes
7. Carbonated soft drinks / Pop
8. Water

RESPONSE OPTIONS:
1. Yes
2. No

S_BEVER2.
Please indicate which, if any, of these beverages you have purchased for your personal use in the last 12 months from a retail store like a convenience, grocery, or discount store.

Please do not count any purchases you might have made at a restaurant.

Please select all that apply. Include both the diet and regular versions in your answer.

GRID ITEMS, RANDOMIZE:

1. 7 Up
2. Coca Cola
3. Cola (store brand)
4. Dr. Pepper
5. Fanta
6. Flavor-wise
7. Ginger ale (any brand)
8. Mello Yello
9. Mountain Dew
10. Orange sodas (any brand)
11. Pepsi
12. Pibb Xtra
13. Root beer (any brand)
14. Sierra Mist
15. Sprite

RESPONSE OPTIONS:

1. Yes
2. No