

# ”Just-in-Time” Surveys - A Case Study in Large- Scale Real-Time Web Sampling and Interviewing

Beth Rothschild, Douglas Willson, and Adine Frank

National Analysts Inc.  
1835 Market Street, Philadelphia, PA, 19103

**Key Words:** Internet Survey; Case Study.

## 1. Introduction

A common marketing practice is *segmentation* - where the market is divided into separate relatively homogenous groups, and tailored marketing efforts are directed at each group individually. Segmentation provides economies of scale to marketers, allowing them to differentiate marketing across customers while focusing scarce resources on relatively large groups rather than individuals. Accordingly, marketing research is often directed towards collecting information to support the development of a segmentation.

The paper describes a case study concerning a large beverage manufacturer that desired to segment retail outlets where their products were sold. Historically, the company’s segmentation strategy was based on type of retail outlet (*i.e.* supermarket, convenience store, *etc....*). This orientation focused marketing activities almost exclusively on *where* beverages were purchased. Additional demand-side information concerning *what* was purchased was available only at the aggregate store or chain-level, and only for some stores. While simple, this strategy provided no differentiation by clientele (*e.g.* type of consumer - who was purchasing), purchase motivators (*e.g.* type of occasion - *when* and *why* the beverage was purchased), and specifically *what* was purchased by these individual. Linking these detailed consumer-based insights associated with specific purchase trips with the store-level characteristics would provide a significant improvement on the firm’s current segmentation framework.

This research strategy presented many challenges. The beverages in question could be purchased in many different types of outlets, including supermarkets, convenience stores, drug stores, warehouse clubs, *etc....*, for immediate or subsequent consumption. To be useful, the research needed to consider

a number of different beverages within the product category. These broad research objectives contrasted sharply with the detailed occasion-based focus that was required to develop actionable marketing strategies reflecting true consumer insights.

From a survey research perspective, it was doubtful that consumers could recall and reliably report on all purchase occasions, since the products in question are typically purchased relatively frequently. Perhaps more importantly, effective targeting required mapping the information concerning store characteristics for specific purchase trips that were collected in the survey to an external database of store-locations. In particular, respondents were required to report the specific store name and location where they were shopping for each occasion - yet it is well known that survey responses to these types of questions are often error-prone.

The paper begins in Section 2 with a discussion of the research challenges associated with the study. Section 3 outlines our responses to these challenges, involving several Internet surveys of consumers. While current research suggests that coverage biases for Internet surveys may be significant in some situations, the Internet framework had important advantages for reducing nonsampling errors from other sources. In addition, our customer’s segmentation objectives and specific features of their market combined to create many sampling challenges; these challenges were most easily addressed using a web survey methodology that allowed for real-time sampling of purchase occasions, and conditioned on individual self-reported purchase behaviors. Section 4 concludes and suggests avenues for future research.

## 2. Research Challenges

### 2.1 Sampling Challenges

The sample design for this study had two components - a probability sample of individual consumers, and a sample of specific purchase occasions for se-

lected individuals. An Internet panel was used as the frame for the sample of consumers. While it is well known that there may be potential coverage biases with Internet sample sources, several factors suggested that an Internet survey might actually produce smaller total error for this survey, in comparison with alternative modes, when other potential error sources were considered. From an overall coverage perspective, the beverage category of interest was quite broadly consumed in the general population, and previous research in the category (using non-Internet sample sources) suggested that differences between covered and non-covered individuals (when using those without Internet access as a proxy for the noncovered population) were not substantial, after controlling for demographics such as income, age, gender, and race/ethnicity. The Internet also afforded many significant opportunities for reducing nonsampling sources of error, and several of these advantages were specific to the Internet and not available through other electronic data collection modes. The study sponsor also required that data be collected quickly, and the Internet provided the ability to collect large numbers of survey interviews quickly.

Sampling of individuals and purchase occasions was challenging for many reasons. There were 13 store types and over 400 product categories, generating over 6000 specific types of purchase occasions. In addition to the large number of different types of occasions, it was well-known that the distribution of occasions was seriously skewed by store type (*i.e.* convenience stores had many more trips than other types of stores), and product category. Furthermore, while some characteristics of the marginal distribution of trips across stores and product sub-categories were known, characteristics of the joint distribution and the relationship to consumer demographics were essentially unknown. Given the ambitious survey objectives, timeline, and the absence of solid information concerning many important design parameters, designing an effective sample appeared both challenging and risky.

## 2.2 Questionnaire Design and Data Collection Challenges

Given that purchase patterns and frequencies varied significantly across individuals, it was clear that sampling of occasions needed to condition on detailed individual purchase histories collected in the survey. However, there was some doubt that respondents could provide enough reliable information to support the detailed sampling requirements.

Respondents' ability to recall specific purchase oc-

casions was conjectured to be poor. Given the large number of stores, products, and brands, it was also likely that respondents would have difficulty providing accurate classifications of store types and product sub-categories. Detailed purchase characteristics including brands purchased, product sizes and prices, would also be difficult to recall. Finally, the detailed store name and location required for targeting purposes was also difficult to recall. Overall, it was clearly important to strike a balance between the volume and level of detail of information concerning occasions that was collected from respondents, and reliability and accuracy of this information.

## 3. Methodology

Sampling and survey challenges were addressed in two phases. In the first phase, a pilot survey was designed to provide information for developing the occasion-based sampling methodology. The pilot sample included interviews with approximately 1000 respondents, and allowed us to gauge rates of eligibility, purchase incidences by sub-category, and the relationship between these variables and demographic factors. The second phase involved the main survey of approximately 3400 respondents and over 6700 purchase occasions. These survey responses were used to segment occasions and assign specific stores to store segments. The following sections discuss our approaches for addressing sampling and data collection challenges.

### 3.1 Sample Design

Information from the pilot test was used to design the main survey sample, a multi-stage probability sample of purchase occasions. In the first stage, individual consumers were sampled from an Internet panel using a simple stratified design. Stratification by gender, age, and household income ensured that subpopulations of interest were adequately represented in the final survey sample. In the initial part of the survey questionnaire, individuals reported summary information concerning the frequency of recent purchase occasions in broad categories. In the second stage of the sample design, clusters of recent purchase occasions were sampled by month and broad product sub-category, and respondents were then asked to report on summary characteristics of sampled occasions within selected clusters. In the final stage, a probability proportional to size (PPS) sample of two specific purchase occasions was selected from the set of occasions identified in stage two. At this stage, respondents were asked to report

detailed information concerning these occasions, including specific products, brands and amounts purchased, purchase motivators, and store characteristics.

Although the first stage purchase history information collected for individuals was relatively coarse, targeted probability sampling at first and subsequent stages allowed us to collect successively more detailed information about specific purchase occasions while minimizing the reporting burden for respondents. Ultimately using the information from each stage for real-time occasion-based sampling required that the information be collected quickly and accurately, so that specific occasions could be immediately classified into one of the 6000+ occasion-based cells. Questionnaire design features that supported this activity are discussed in Section 3.2 below.

Information from the pilot test played a crucial role in designing the sample. By construction, the pilot test investigated a larger number of trips for each individual than was thought necessary for the main survey. Analysis of nonresponse/break-off rates allowed us to evaluate the relationship between nonresponse rates and questionnaire burden. The pilot test also provided information concerning seasonal purchase patterns that was subsequently incorporated in selection probabilities for second stage clusters. Finally, the pilot tests provided detailed purchase information by store-type, product sub-category, and brand that allowed us to tailor stage three selection probabilities to meet survey objectives. Sampling rates for stage three were developed through simulation and optimization, using the pilot survey sample information to evaluate the effectiveness of alternative designs.

An important advantage of electronic data collection, and Internet surveys in particular, is the ability to change probabilities of selection during data collection to align the realized sample with survey objectives. For this survey, effective management of sampling rates required reporting, in real-time, on over six thousand occasion-based sampling cells. In addition, specific programming and procedures were also developed to modify sampling rates "on-the fly". Given the pace of data collection and the large number of interviews, the NAQuest<sup>TM</sup> reporting structure played an important role in supporting mid-stream adjustments of sampling rates and screening quotas. Overall, this approach to sampling reduced the required full survey sample size from initial estimates by a factor of six, while meeting precision targets on marginal analytic domains of interest such as store types and beverage sub-categories.

### 3.2 Questionnaire and Data Collection

Internet surveys share many advantages of traditional electronic surveys, allowing for complicated logical branching and supporting a variety of logical and numerical edits. The Internet also allows for true "forms" data collection, where the data fields are organized in a 2 dimensional response grid and respondents provide multiple inputs, where appropriate, on each screen. The questionnaire structure that was employed in the survey used a sequence of response grids to guide respondents through the sample selection process as quickly as possible, while supporting the collection of accurate and reliable data form respondents.

The survey for this project was programmed in NAQuest<sup>TM</sup> National Analyst's proprietary questionnaire development and interviewing software. Figure 1 displays an example of the response grid that was used in Stage One of the occasion sample. For this question, respondents were asked to identify their number of purchase occasions in recent months for several broad categories of purchases. Using this information, a cluster of purchase occasions was sampled for each individual.

Figure 1

Q1-3b For each month below, please indicate on how many shopping trips (taken within the continental United States), if any, you purchased Product 1 only; Product 2 only; Product 3 only; Product 1 and Product 2; Product 1 and Product 3; Product 3 and Product 2; and Product 1, Product 2 and Product 3.

Note: Please count each store where you purchased as a separate shopping trip. Think about your various types of shopping trips (e.g. parties, gifts, holidays, etc.) and please estimate the number of shopping trips of each type that you made.

Please write in whole numbers. You do not need to type 0.

	Time Period				
	October 2002	September 2002	August 2002	December 2001	November 2001
Product 1 only	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Product 2 only	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Product 3 only	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Product 1 and Product 2	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Product 1 and Product 3	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Product 3 and Product 2	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Product 1, Product 2, and Product 3	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

More detailed information concerning this cluster of purchase occasions was obtained in Stage Two of the occasion sample. Figure 2 displays the response grid where this information was collected. Respondents were asked to identify each trip with a name that would be used in subsequent reporting if the trip was selected for detailed analysis. Figure 2 also shows the use of drop boxes for store types to minimize reporting burden and response errors.

Figure 2

Q14-1 Please think about the last 5 shopping trips that you took to purchase **Product 1** in September.

For each trip, in Column A, please give each shopping trip a different name that will help you identify it (e.g., football party trip, birthday party trip, stock-up trip, etc.)

In Column B, please select at which type of retail store you made your purchases.

In Column C, please select in which state you made your purchases for that shopping trip.

	Column A Trip Name (Type in)	Column B Type of Retail Store (Select one)	Column C Location of Retail Store (Select one)
Trip # 1	<input type="text"/>	Please select store type	Select a State
Trip # 2	<input type="text"/>	Please select store type	Select a State
Trip # 3	<input type="text"/>	Please select store type	Select a State
Trip # 4	<input type="text"/>	Convenience Store	Select a State
Trip # 5	<input type="text"/>	Drug Store	Select a State
		Gas Station (with mini-mart)	Select a State
		Limited Assortment/Gourmet	Select a State
		Mass Merchandiser	Select a State
		Small Grocery Store	Select a State
		Supercenter/Hypermarket Store	Select a State
		Supermarket	Select a State
		Wholesale Club/Warehouse Store	Select a State
		Other Store	Select a State

In the final stage of the occasion sample, two specific occasions were selected for detailed reporting. Additional information concerning store names was collected using pre-coded lists that varied by state and store type. One again, this approach was taken to reduce burden and reporting error. In this particular beverage category, there are hundreds of thousands of merchant locations; for the survey, over 90 percent of sampled occasions reported store name using the pre-coded lists. An example is included in Figure 3.

Figure 3

Q2.4 What is the name of the **Supermarket** in which you purchased **Beverages** in September 2002? Please select from the list below or if you do not see the name, please click "Other" and type in the name.

Please select one or type in a name.

Name of Supermarket  
(Select one)

Aldi Food Store                       Super S Foods  
 B Kwik Food Store                     Sutton Place Gourmet  
 Fas Chek Food Mart                     Trader Joes Market  
 Greers Food Store                       Vallarta Supermarket  
 Gristedes Supermarket                 Whole Foods Market  
 La Michoacana                          Wild Oats Market  
 Niemann Mini Mart                     Other (Specify)

                                      

Figure 4 displays the response grid for collecting information concerning purchase quantities and prices for specific package sizes. Pre-coded brand lists for product sub-categories were also employed to collect accurate brand information.

Figure 4

Q3-10e1\_2 For each brand purchased at **La Michoacana** on the first shopping trip in September 2002, please enter the total quantity purchased and the approximate price **Per Unit** of each size listed in Columns A and B.

In Column C, please check the box if any of the items were purchased on sale. If any items were bought as gifts please check the box in Column D. Finally, in Column E, enter the quantity, if any, that you bought as a gift.

The Total Column will calculate your total spending for you automatically so that you can judge if the approximate cost per unit you are entering seems accurate based on your total spending. If the total seems too high or too low, please adjust the numbers in Column A and B accordingly.

Please write in whole numbers for Columns A and E. For Column B, please enter the price as dollars and cents such as \$14.99 or \$4.50.

	Column A Quantity	Column B Approximate Cost Per Unit	Column C Gift Quantity	Total Spent
<b>Product 1</b>				
- 12 oz or 16 oz single can/bottle	<input type="text"/>	\$ <input type="text"/>	<input type="text"/>	\$ <input type="text"/>
- 40 oz single can/bottle	<input type="text"/>	\$ <input type="text"/>	<input type="text"/>	\$ <input type="text"/>
- 4-pack(s) or 6-pack(s) of cans/bottles	<input type="text"/>	\$ <input type="text"/>	<input type="text"/>	\$ <input type="text"/>
- 12-pack(s) of cans/bottles	<input type="text"/>	\$ <input type="text"/>	<input type="text"/>	\$ <input type="text"/>
- case(s)	<input type="text"/>	\$ <input type="text"/>	<input type="text"/>	\$ <input type="text"/>
<b>Product 2</b>				
- 12 oz or 16 oz single can/bottle	<input type="text"/>	\$ <input type="text"/>	<input type="text"/>	\$ <input type="text"/>
- 40 oz single can/bottle	<input type="text"/>	\$ <input type="text"/>	<input type="text"/>	\$ <input type="text"/>
- 4-pack(s) or 6-pack(s) of cans/bottles	<input type="text"/>	\$ <input type="text"/>	<input type="text"/>	\$ <input type="text"/>
- 12-pack(s) of cans/bottles	<input type="text"/>	\$ <input type="text"/>	<input type="text"/>	\$ <input type="text"/>
- case(s)	<input type="text"/>	\$ <input type="text"/>	<input type="text"/>	\$ <input type="text"/>

                                      

Finally, an important feature of Internet data collection that is not available with other modes is the ability to use a hyper-link to visit another site during the interview to obtain important information. To support the segmentation objectives for this study, respondents needed to provide an accurate store name and address so that the purchase occasion could be mapped to a specific store in a national retail establishment database. In the survey, respondents were provided with a link to the Yellow Pages and asked to look up a verify the name, address, and telephone number of the store associated with each sampled occasion. Respondents verified the name, address, and telephone number for over 56 percent of the trips in the survey database using the hyper-link. As a result, a base match rate of slightly over 56 percent was obtained with a simple (deterministic) match merge on telephone number. Additional probabilistic matching procedures eventually raised the overall match rate to over 86 percent, an extremely high rate when compared with conventional matching rates using survey self-reports from consumers. An example of the location questions and hyper-link are contained in Figure 5.

Figure 5

02.6a What is the telephone area code of Albertsons Store? If you are not sure, please look it up by using the following link, <http://yellowpages.msn.com> or by using a phone book. If you are unable to find the area code, please leave it blank.

*Please write in an area code.*

Area Code

02.6b What is the telephone number of Albertsons Store? If you are not sure, please look it up by using the following link, <http://yellowpages.msn.com> or by using a phone book. If you are unable to find the phone number, please leave it blank.

*Please write in a phone number with a dash.*

Phone Number

02.7 Which of the following is the correct answer?  
Please select one.

Stop



#### 4. Conclusions

This case study illustrates that the Internet is ideally suited for large-scale real-time sampling and interviewing. In addition, a number of specific conclusions are supported by the study:

- Pilot tests provide important opportunities for reducing uncertainty concerning sampling and questionnaire specifications. Recontacting pilot survey respondents is also extremely cost effective within the web environment.
- When properly configured, complicated real-time sampling protocols can be implemented, monitored, and adapted on the web. Fast, real-time reporting capabilities are required in this environment when using an Internet panel sample source because of the speed at which responses accumulate.
- In situations where self-reported survey responses must be matched to external databases with store names and addresses, match rates can be significantly improved through the use of a hyper-link to the Yellow Pages
- Measurement errors can also be significantly reduced through the use of drop down menus with pre-coded store and brand names