

'To improve the
quality of research.'



An Examination of Online Sampling Techniques

Background

One of the key challenges of executing online research is that of sampling: where and how to acquire Internet sample, how to control the composition and consistency of samples and what is the optimal process for sampling on the Internet. Given these issues, along with the proliferation of badly designed surveys being forced upon web users and the heightened sensitivity to spam-like invitations, online panels are being created in hopes of providing better solutions to these sampling issues.

Objectives

The purpose of this paper is to examine Internet-based concept testing, going beyond the fundamental testing done previously to more comprehensively understand the Internet environment and the "new" respondent located therein.

Moreover, an investigation into Internet sampling will focus on the distinct differences in research findings between Internet ePanel¹ and Internet database²sampling.

Specifically, there are five distinct objectives to examine between Internet sampling methods:

1. To determine level of consistency between Internet and Mail Panel concept test results. Is consumer response to the concepts similar in terms of Purchase Interest, Uniqueness, Price-Value and Overall Liking when tested through Mail Panel versus over the Internet? And, most importantly, would the same business decisions be made regardless of methodology?
2. To understand the differences in demographics between Internet and Mail Panel samples. Has the dramatic increase in Internet penetration and usage seen recently in the U.S. resulted in samples that more closely mimic a "representative" sample of U.S. households on relevant demographics?

Do any noted differences impact reported key measures?

3. To examine the impact of applying weights to the Internet data (based upon specific demographic targets) on the concept scores. Do the scores exhibit substantial changes or not?
4. To determine the impact of length of field time on concept acceptance and sample demographics. Do early versus late responders exhibit different demographic profiles or response to the concepts?
5. To understand the impact of varying levels of concept finish on results. Specifically, do color photographs or black and white line drawings score differently online versus mail panel?

Research Design

A series of parallel tests were designed to include a broad range of concepts both in terms of performance and across a variety of manufacturers and categories. Specifically, the concepts ranged from "winners" to "losers" and all levels of performance in between. Further, concepts from a variety of major consumer packaged goods categories, including food, household products, health and beauty aids and OTC pharmaceuticals, were tested across two separate research-on-research projects.

Mail Panel versus Internet ePanel Parallel Testing: In total, 25 concepts were evaluated. Sample sizes were approximately 200-400 completes per cell. All Internet testing was conducted from March through May 2001.

Mail Panel versus Internet Database Parallel Testing: In total, 60 concepts were evaluated. Sample sizes were approximately 300 per cell. All Internet testing was fielded in late November and December 1999.

Mail Panel samples were balanced on the outgo to be representative of the U.S. household population in terms of geographic region, population density, income, household size, and age of household head.

Internet ePanel samples were also balanced on the outgo to be representative of U.S. Census population across same household demographics as Mail Panel samples.

A smaller number of demographic criteria were utilized in balancing the sample outgo used for Internet database research. The outgo was balanced with respect to geographic region, age and gender. As the Internet database is recruited and maintained at the individual level, the targets for balancing were adult individuals, rather than U.S. households.

Sample returns from any research design were not rebalanced or reweighted.

To ensure consistency between the two methods, concepts used in Mail Panel studies were scanned and shown as full screen images within the Internet interviews. In most cases, the concept was fully presented on screen without the need for scrolling. However, in some cases, respondents were required to scroll (up and down only) in order to read entire concept. To the extent possible, questionnaires, as well as concept clarity, were consistent across both mail and online methodologies.

Sample Response Rates

Cooperation and completion rates for panel-based research, online or offline, are sharply higher than Internet database sampling. Mail Panel cooperation rates averaged at 74% compared to 36% for Synovate ePanel and 6% for Internet database.

Improved contact and cooperation rates for panel-based research leads to greater design efficiencies: lower data collection costs, expedited field timing and higher quality data.

SCOP MAIL PANEL VERSUS SYNOVATE INTERNET ePANEL

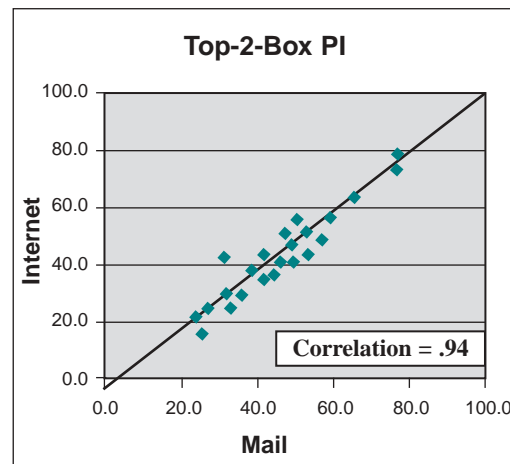
In May 2001, Synovate (then Market Facts) conducted internal validation work comparing concept testing data and demographics between two balanced household-based panels: Mail Panel and Internet ePanel. Household panels offer several benefits to researchers including stronger cooperation, customized sampling, nationally representative demographics and a wealth of pre-identified respondent and household background information.

Comparison of Key Measures

A comparison of key measures from Mail Panel and ePanel tested concepts indicated a strong correlation between the two for top-two-box Purchase Interest (.94). Correlations this high suggest that the two methodologies were indeed measuring the same thing. Thus, the conclusion can be drawn that online concept testing yields results comparable to those gathered via the Mail Panel and is a viable tool for executing concept evaluation.

Further examination of the parallel results shows an R-square of 88%. This indicates that one "explains" 88% of the variance in the other, or that they have 88% of their variance "shared," or in common. Again, strong evidence that Internet testing mimics the offline method.

TABLE 1: PURCHASE INTEREST CORRELATION



As shown above, Purchase Intent observations fell both above and below the diagonal drawn on the scatter plot. This illustrates that neither method (Internet or Mail Panel) is generating systematically higher or lower scores than the other. In other words, there is no consistent difference (either higher or lower) in the scoring of concepts by Internet versus Mail Panel respondents.

The high correlation observed for Purchase Intent was consistently observed across other key measures (e.g. Price-Value, Likability, Uniqueness and Purchase Frequency). Correlations this high indicate a strong degree of predictive validity between the two methods. Further, since consumer reaction was found to be very consistent across all key measures, concepts that were “winners” in one method were “winners” in the other as well, and the same for “losers” and “middle of the road” concepts. Both data sets would support similar business decisions, the ultimate test of methodological validity.

The following table summarizes the correlations for top-two-box, top-box, and means for key measures of concept acceptance.

TABLE 2: KEY MEASURES CORRELATION

	Correlations of Internet ePanel and Mail Panel Key Measures		
	Top-2-Box	Top-Box	Mean
Purchase Intent	.94	.94	.95
Likability	.92	.82	.92
Value	.91	.93	.93
Uniqueness	.96	.97	.97

While the high correlations indicate that the Internet ePanel and Mail Panel concept scores tend to move together quite closely on a concept by concept basis, the absolute levels of scores varied slightly.

The largest difference was noted for Purchase Interest and Perceived Uniqueness, where the scores varied by 2.2 and 1.9 points, respectively.

TABLE 4: SAMPLING COMPARISON

DEMOGRAPHICS	U.S. CENSUS	MAIL PANEL	INTERNET ePANEL
Children Present	36.0	32.4	29.2
Married	56.4	56.8	50.9
Single Member	25.0	25.2	27.4
College Graduate	22.4	27.1	33.4
<u>Household Income</u>			
Over \$100,000	10.3	8.8	10.8
Under \$17,500	19.0	20.4	15.6
<u>Age of Respondent</u>			
18-34 years	32.2	19.1	22.0
65 and older	16.4	20.9	10.6

TABLE 3: RAW DATA SCORE COMPARISON

Top-Two-Box Scores	Mail Panel	Internet ePanel
Purchase Intent	45.3	43.1
Liking	29.2	29.8
Value	41.4	41.5
Uniqueness	39.0	37.1
Frequency (Mean)	9.4	9.8

These findings support a previous hypothesis formed based on earlier parallel testing experiments: Higher income/education levels and more varied communications received by Internet respondents (versus the average offline respondent) contribute to a more stringent definition of uniqueness.

Comparison of Sample Demographics

A demographic profile of the samples indicated that they were relatively similar in terms of presence of children, percent married, household size, region and race.

The primary differences in the samples centered on education, and to a lesser extent, income and age. The Internet samples had higher proportions of college graduates, but lower percentages of consumers 65+ and lower income consumers. Not surprisingly, these differences reflect the general differences in the Internet population versus the general U.S. population.

Although there were differences between the Internet and Mail Panel samples on some demographics, it should be remembered that these differences were insufficient, or at least not correlated strongly enough with concept appeal, to alter the consistency between Mail Panel and Internet ePanel concept scores.

Impact of Reducing Field Length

Internet field length remains a frequent and fervent topic of inquiry. While the Internet affords us the possibility of faster turnaround than ever before, the question still remains: can short field times provide reliable, consistent results? In order to determine whether or not the field time on studies such as this could/should be shortened to a 24-hour turnaround, the responses of early responders were compared to total sample to determine what differences, if any, would exist.

The samples were divided into two groupings: those who completed survey within the first 24 hours of email delivery and then total field completes. Field period was restricted to seven days regardless of participation rate.

At first glance the responses of early responders were quite consistent to total sample with a near perfect correlation on purchase interest (.995). Further, an examination of absolute levels of acceptance across a variety of key measures showed little difference between early and total responders.

TABLE 5: 24-HOUR RESPONSE DATA

Response Time:	First Day Only	7-Day Field Period
Top-2 PI	43.4	43.1
Top-2 Liking	31.9	32.0
Top-2 Value	41.5	41.0
Top-2 Unique	36.4	36.2

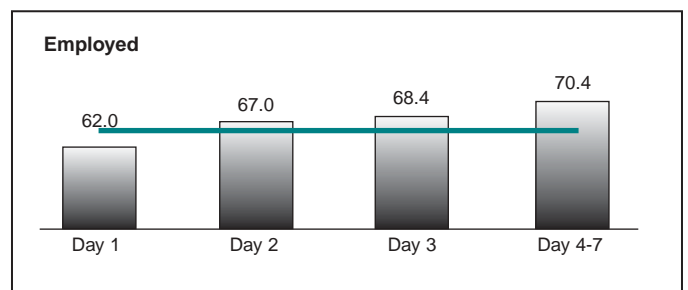
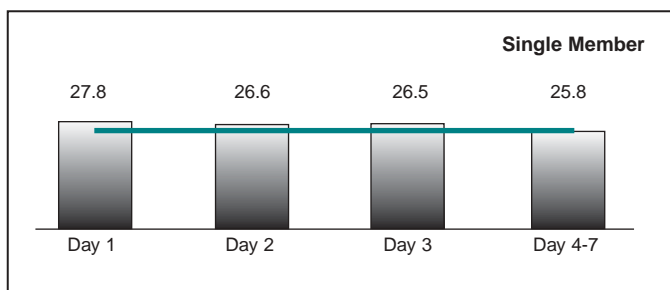
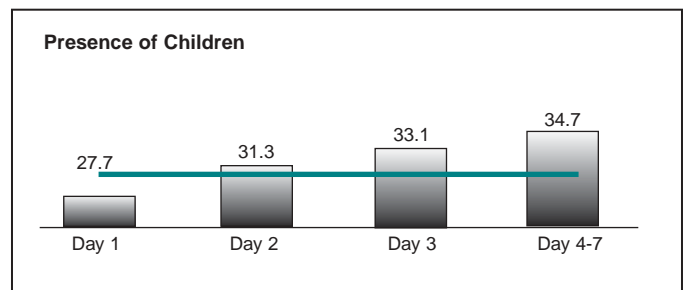
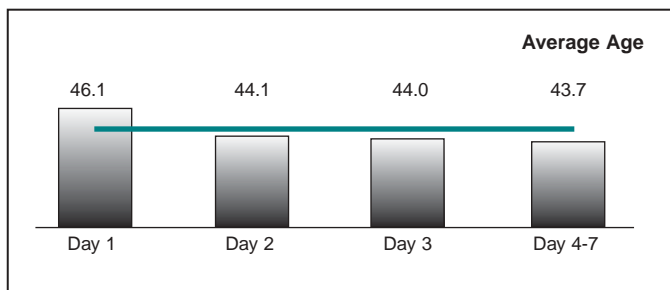
However, throughout the field period, respondents' demographic profiles did vary on several characteristics, as noted below. Early responders were less likely to have children present, more likely to be older and less likely to be employed than the later responders. In short, early responders tend to be those who are more likely either (1) to be more frequent Internet Users or (2) to have more time available for completing surveys on short notice. (Note: Horizontal bar across graphs below indicates U.S. Census %).

Thus, while the advent of the Internet allows for faster turnaround than ever before, it is only with great caution that one should execute very short field times. While concept scores were similar in this test, the demographic differences posted could impact scores on concepts aimed at particular target audiences, such as households with kids or older adults. Further, differences can be magnified depending on the day of week a study is fielded, as weekday versus weekend respondents do differ.

Impact of Level of Finish of the Concept

One of the first, and most often cited, Internet benefits leveraged for conducting better research is its graphics capabilities. Thus concept testing, typically dependent upon graphic stimulus, was a natural fit for this medium. However, given the wide range of concept quality typically tested (both in terms of composition and level of finish), it raises the question of whether certain types of stimuli perform better or worse on the Internet versus Mail Panel (where hard copy stimulus is used).

TABLE 6: DEMOGRAPHIC COMPOSITION ACROSS FIELD PERIOD



In this series of tests, as is common, some of the concepts presented as color photographs and others were black and white line drawings. We examined two types of stimuli to determine if there was any impact on results based on stimulus type.

Our findings showed that regardless of how the product was portrayed, similar results were observed via Internet relative to the scores achieved on the Mail Panel for the same concepts. Correlations between Internet and Mail Panel scores were high regardless of the form of stimulus (color picture or B&W line drawing). Therefore, we can conclude that concepts will perform similarly, in either interviewing environment regardless of stimuli presentation.

TABLE 7: CONCEPT FINISH COMPARISON

	Level of Finish on the Concept Board	
	Color Photo	Black & White Line Drawing
Top-2 Purchase Int	.90	.96
Top-2 Liking	.87	.94
Top-2 Value	.87	.94
Top-2 Unique	.94	.98

MAIL PANEL VERSUS INTERNET DATABASE

Currently, sixty-one percent of U.S. households are actively online indicating that the Internet is becoming a staple in the household.³ Research conducted online has reflected the intensive growth of access penetration yielding a multitude of service providers conducting traditional and non-traditional research over the Internet.

One such source for online research is access to email address databases – a collection of individuals who knowingly, or at times unknowingly, have been included in a collection of names receiving research surveys.

An email database differs from a pre-recruited household online panel in several ways, including:

- An email database is based on an individual email address with little or no known information about the household;

- Limited information is typically collected on database participants – perhaps simply age, gender and region;
- Participation and cooperation rates are much lower for database research and tend to average around 5%-15% overall;
- Re-contact for follow-up research may be more difficult since offline information (e.g. address) is not collected at recruitment.

Because of questionable opt-out practices and the explosion of Internet-led suppliers providing an email database for research, clients were experiencing poor data results. Some researchers began to oppose the use of online databases in all applications. But, the question remained as to whether an online database could be used to conduct quality research in a controlled, experienced environment.

During Fall/Winter 1999, a series of validation work was conducted comparing balanced Mail Panel concept testing to Internet database concept testing. A stratified sampling technique employing only Age, Gender and Region variables was used to select individual participants for Internet database samples. Mail Panel sample was completely balanced to represent U.S. Census households.

Through a study of 60 monadic concept evaluations, the following results were evidenced.

- ✓ Stratified Internet Samples correlated extremely well with Mail Panel samples across key survey measures including purchase interest, liking, price-value and uniqueness. This would indicate that complete demographic balancing is not required to measure reliable concept acceptance over the Internet.

TABLE 8: MAIL PANEL vs. DATABASE CORRELATION

	Top-2 Box	Top Box
Purchase Interest	.92	.89
Liking	.92	.90
Value	.95	.92
Uniqueness	.97	.96

- ✓ Internet database and Mail Panel samples were very similar across key demographic variables including presence of children, marital status, household size and race. While some differences were noted for age, gender and education, these differences were not strong enough, or important enough, to impact key survey measures.
- ✓ Level of concept finish does not impact strength of purchase interest correlation between Internet and Mail Panel results.

These preliminary findings were completely supported by concept validation work conducted among Mail Panel and Internet ePanel research discussed earlier in this paper. Regardless of Internet source – household panel or individual database – offline and online concept testing results were fully validated with similar, matching demographic profiles noted.

In addition, side-by-side testing conducted among Internet database samples researched several other design factors validating the database stratified sampling approach.

Comparison of Packaged Goods Categories

Further analysis showed correlations between Internet database and Mail Panel scores were also generally high within each of the consumer-packaged goods categories tested. This indicates the two methodologies parallel each other both across and within a wide range of products.

TABLE 9: CPG CATEGORY CORRELATIONS

	Food	Household Products	Health & Beauty Aids
Top-2 PI	.93	.89	.88
Top-2 Liking	.95	.87	.89
Top-2 Value	.94	.88	.94
Top-2 Unique	.98	.93	.97

Impact of Increasing Number of Balancing Variables

In the sixty Internet database cells analyzed thus far, three sample balancing criteria were utilized (age, gender, region). A subset of nine cells was also tested using an increased number of balancing variables at time of sample selection, incorporating household income and marital status as well as age, gender and region. The purpose

of this experiment was to determine whether or not increasing the number of balancing variables produced results more consistent with Mail Panel scores (which include same five balancing variables).

Both Internet Database sampling approaches (3-Criteria and 5-Criteria) correlate equally strong with the corresponding Mail Panel results. The inclusion of the additional balancing criteria did not impact the consistency versus Mail Panel, thus concluding that fewer variables can yield Internet samples from which reliable concept evaluations can be conducted.

TABLE 10: ALTERNATE SAMPLING APPROACHES

Correlation	Mail Panel vs. Internet (3 Criteria)	Mail Panel vs. Internet (5 Criteria)	Correlation Between Internet Sampling
Top-2 PI	.83	.84	.98
Top Box PI	.78	.78	.91
Average PI	.89	.91	.96

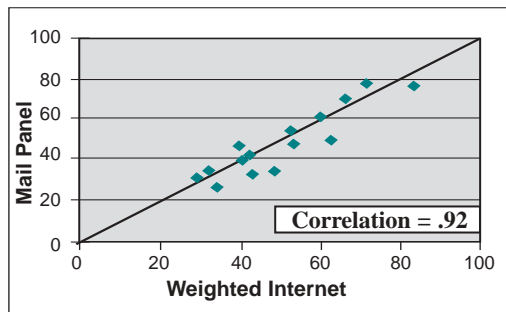
Impact of Weighting the Data

In order to understand the impact of weighting the data on results, a selected subset of sixteen Internet cells were weighted to be representative of adult primary grocery shoppers using six factors: age, income, household size, region, gender and education. The weighted results for the Internet were then compared with unweighted Internet concept scores and Mail Panel scores for the same sixteen concepts.

Although there were some significant differences for selected demographics (education, income and age) between the unweighted Internet samples and representative (weighted) sample, weighting the Internet scores produced little movement in the key measures. In fact, correlations between unweighted Internet and weighted Internet approached 1.0.

The correlations between Internet and Mail Panel results also did not improve when the Internet results were weighted. The lack of impact on results strongly suggests that the stratified sampling frame used for database research provides valid, comparable samples from which to draw business decisions.

TABLE 11: TOP-2 BOX PURCHASE INTEREST



However, weighting Internet database scores did substantially increase sampling error generating a weighting efficiency of 70% (without education). This means that a sample size of 100 would have the sampling error associated with a much smaller sample (i.e. a sample size of 70).

COMPARISON OF INTERNET SAMPLES ON VALIDATION RESULTS

Synovate sets the industry standard for providing clients with superior, business-smart research. One of our strongest priorities remains allowing clients the highest level of flexibility and quality available in conducting research whether online or offline.

To date, we may be the only Research Company who has validated not just online concept testing, but online concept testing across two different Internet sampling techniques – Internet ePanel and Internet Database.

The success or failure of any online research is dependent on the research knowledge and skill set of the supplier conducting the research. Development of online knowledge, skills and experience requires a great deal of investment – both in time and money. Many companies have cut back on these investments by offering the quickest and most convenient sample sources available without ensuring the quality and validity of this approach.

Synovate has recruited its own Internet ePanel adhering to the most stringent guidelines followed throughout the industry. Investments paid into development and maintenance of this Internet ePanel, while resulting in higher research costs to clients, does yield research advantages in:

- Data consistency
- Data reliability
- Demographic representation

Following is a comparison of the benefits and advantages offered through use of a balanced household ePanel approach compared to Internet Database sampling.

Data Reliability and Consistency

Internet ePanel concept testing research yielded higher quality data, compared to Internet Database research, both in terms of data consistency (correlation to mail panel results) and reliability (reduced difference between raw data results).

Internet ePanel research evidenced a stronger Purchase Interest relationship to Mail Panel research, than Internet Database research. Higher predictability and more accurate data measurement ensure greater data validity compared to database sampling.

TABLE 12: INTERNET SAMPLING PURCHASE INTEREST COMPARISON

PURCHASE INTEREST CORRELATION TO MAIL PANEL RESEARCH		
	Internet ePanel	Internet Database
Top-2-Box	.94	.92
Top Box	.94	.89
Average PI	.95	.91

In addition, comparison of absolute level of difference across all key concept measures indicates that the Internet ePanel research yields more reliable and comparable raw data to Mail Panel than the Internet database. With lower variance between key measures, researchers can feel more secure in data results being used during the key business decision making process and for comparison of online data results back to long-term mail panel research norms and databases.

TABLE 13: INTERNET SAMPLING RAW DATA

ABSOLUTE DIFFERENCES IN RAW DATA RESULTS (MEDIAN) BETWEEN INTERNET AND MAIL PANEL RESEARCH		
	Internet ePanel	Internet Database
Purchase Interest	2.9%	4.0%
Overall Liking	3.5%	3.6%
Price-Value	5.3%	4.0%
Uniqueness	4.4%	5.0%

Demographic Representation

Internet ePanel research also indicates better demographic representivity compared to Internet database sampling. While differences in demographics did not impact key measures for this concept testing validation, they may impact research targeting specific population segment such as families, young adults or seniors.

TABLE 14: INTERNET SAMPLING DEMOGRAPHICS

	U.S. Census	Internet ePanel	Internet Database
Single Member	25%	27%	18%
College Graduate	22%	33%	38%
Income >\$100k	10%	11%	7%
Income <\$17.5k	19%	16%	10%
18-34 years	32%	22%	22%
65 and older	16%	12%	11%
Children Present	36%	29%	35%

Summary and Discussion

This research has discussed the validity of conducting concept testing online through extensive collaborative, side-by-side research with proven mail panel research.

Initial results are extremely positive in that regardless of online sample source – balanced household ePanel or individual email database – there exists high levels of correlation and validity to offline research. Across all key concept measures, including overall purchase interest, mail panel and online research results are highly predictive and reliable.

Key to this research was the final comparison of benefits offered by Internet ePanel research. As recruitment and maintenance of any household research panel require extensive corporate investment, the hypothesis that panel-based research would provide higher quality data was proven with this research. Investment costs are directly related to recruitment, registration and maintenance of a balanced, representative panel as well as the experience and knowledge of seasoned research veterans to ensure that highest quality procedures are in place - - - procedures not necessarily followed by database vendors.

While both Internet ePanel and Internet database research does yield the same validated key measures, and comparable business decision-making results to mail panel research, several key benefits leveraged by ePanel research includes data reliability and consistency as well as improved demographic representation of U.S. household population. Data reliability and demographic representation are key factors determining the quality of any research and a requirement for following standards set in place through over 50 years of research experience at Synovate (formerly Market Facts)

Research Notes

- ¹ Internet ePanel defined as pre-identified sample source specifically recruited and registered via high quality double opt-in process for research purposes. Panel member knowingly and willingly has agreed to participate in online research surveys and has provided extensive individual and household background information for balancing criteria.
- ² Internet database defined as large-scale registration of individual email addresses for which little prior agreement or cooperation is cited for selection purposes. Database provides limited individual information and no household information.

³ Source: June 2001 survey by Dataquest