



FORS | MARSH  
GROUP

# **Research on Consumer Tipping Behavior: Response to OMB Information Request**

**Prepared for Internal Revenue Service**

**Prepared by Fors Marsh Group LLC**

**July 2016**

The views, opinions, and/or findings contained in this report are those of Fors Marsh Group LLC and should not be construed as official government position, policy, or decision unless

so designated by other documentation. This document was prepared for authorized distribution only. It has not been approved for public release.

# Research on Consumer Tipping Behavior

## Response to OMB Information Request

Per OMB’s request for additional information regarding the IRS’s Research on Consumer Tipping Behavior project we have assembled the information below. Specifically, OMB requested that we provide additional information regarding:

- Quota sampling methods and variables
- Quality assurance processes
- Estimation procedures, including poststratification, regression, weighting, etc.
- Outreach/advertising/recruitment methods
- Any existing research on panel comparison to benchmarks

Each of these topics is addressed in a separate section below to provide further insight into the planned activities and highlights the degree of scientific rigor which is being brought to bear on this project by the IRS, Fors Marsh Group, and their subcontractors, specifically, Ipsos.

### Quota sampling methods and variables

#### Sample Balancing

Ipsos and each of its partners will select what is known as a “balanced return” sample, wherein the demographic distribution of “clicks” (meaning respondents who respond to a survey invitation by clicking the hyperlink and entering the survey) matches the demographic distribution of the overall U.S. population, as indicated in most recent results of the Census Bureau’s Current Population Survey (CPS).<sup>1</sup> Because different individuals and demographic groups respond at different rates, the different sampling rates are applied for these different groups. The demographic distribution of the contacted sample thus does not match the demographic distribution of the U.S. population.

Sample balancing (i.e. determining the proportion of sample to allocate to different demographic groups) will be done using four demographic variables: gender, age, region, and income. The links between each of these characteristics and tip rates have been the subject of past academic studies on tipping behavior<sup>2</sup>. These variables will be fully crossed, creating 96 sampling cells (see Appendix A). The levels (sample groups) within each of the variables are indicated in the table below.

Gender	Age	Region	Income
(1)Male	Age 18- 34	(1)NorthEast	Under \$20K

<sup>1</sup> To ensure sufficient sample records to complete the necessary number of interviews each month, multiple sample sources are needed. The sample for the IRS Consumer Tipping Study will be provided by Ipsos’ opt-in i-Say panel and four other opt-in panels, with the anticipated proportion of completed interviews provided by each source remaining constant each month (and following the proportions used in the pilot test). Each panel provider has prepared responses to the ESOMAR 28 questions for online samples and has been vetted by Ipsos’ online research department. These panel providers will email invitations to their panelists with a link that directs them to the Ipsos survey site after passing them through an intermediary site used by the panel provider to monitor whether they (A) respond and (B) complete the survey, so that their traditional panel incentive is paid. Panel partners will provide information on how many invitations are sent and will balance their samples using targets provided by Ipsos.

<sup>2</sup> See IRS report *Estimating Consumer Tipping Behavior: Review and Recommendations*, which is attached as a supporting document, for a review of past tipping studies.

(2)Female	Age 35- 54	(2)Midwest	\$20K - \$49,999
	Age 55+	(3)South	\$50K - \$99,999
		(4)West	\$100K+

Ipsos will select samples three times a week (Monday, Wednesday, and Friday). On Monday and Wednesday, the sample will be designed to produce a demographically balanced return sample equal to two days' total of completed interviews. On Friday, the sample will be designed to produce the balanced return sample equal to three days' total of completed interviews. The samples will then be divided into replicates (two replicates for the Monday and Wednesday samples, three replicates for the Friday samples), so that one replicate can be "released" (meaning survey invitations will be sent to those sampled individuals) each day. These invitations, which include invitation text, a link to the survey program, and a link to the panel provider's member policies (including confidentiality), will follow the standard email invitation formats used by Ipsos and each of its partners, so that sampled individuals will be familiar how to use them to access the survey. This approach will thus yield the targeted 322 daily completed interviews.

This approach of using sample replicates is employed to achieve greater efficiency when many sample balancing cells are employed by ensuring higher response rates in relatively sparse sampling cells.

The sample design assumes a 1-month re-use of sample (i.e. individuals who were sampled for the study in one month will be ineligible for another contact until the next month - something not as relevant for the one-month pilot but which will matter for the year-long main study). Variance estimation for statistics of interest (e.g., mean tipping rate) will account for any non-independence in reported tip rates for individual transactions that may occur from repeat respondents. This is accomplished through the clustering of standard errors based on geography as discussed in the poststratification section below. Those who have quit the survey will not be able to re-enter it at all.

## Quality assurance processes

### Data Collection and Sample Quality and Security Procedures

Ipsos employs a number of quality checks during the data collection process.

- Survey-level:
  - o Filtering of respondents based on participation history.
  - o Respondent screening based on demographic variables being captured for the survey (age, gender, zipcode, etc.).
- Engine-level:
  - o GeolP verification - validates survey country vs. respondent country determined based on IP.
  - o Language verification - validates survey language vs. respondent language.
  - o Device check - match between device used by respondent and the device setting of the survey.
  - o Algorithm to identify possibly unengaged respondents (straight-lining, speeding, providing invalid verbatim in open ended questions.)
  - o Concurrent session sniffout - filter respondents with more than one opened session, in the same browser, on the same survey.

- o Fraud Profile Flag 4 (FPF4) – machine time vs. time based on geo location mismatch.
- o Open and anonymous proxy checks.
- o VOID – analysis of web cookies, PanelistID/SupplierID (identifiers provided by sample sources), RelevantID (third party security service), SHA-1 hash function.

**Data Analysis Quality Assurance Procedures**

**Web Survey Quality Control.** The FMG Team will perform full testing of the programmed instrument to assure that skip logic, randomization, conditional data piping, question wording, and all other specifications for the survey instrument are met. FMG quality control process for our online surveys is thorough and includes checks to ensure there are no grammatical or formatting errors, question type is accurate (single punch vs. multi punch etc.), skip patterns function appropriately, and data restrictions for open-ended questions match requirements. The FMG Team also has data capture checks in place to examine the functionality of the programmed survey. As a standard quality control check, multiple FMG researchers will respond to the online survey and simultaneously record the answers on a paper copy of the survey; during these checks researchers will test all branching/paths of skip patterns in the questionnaire. Hard copies of the survey responses entered for each record will be compared to what was captured by our web-based technology. Should the data check reveal errors, we will make the necessary changes to the web-based technology and the checks will be conducted again until 100% accuracy is achieved. The FMG Team will record all discrepancies in a log file and make updates immediately.

**Survey Tracking.** We will establish and maintain a secure survey control system that will document the correspondence and track the status of all sample members. The heart of this system is a unique sample ID that is given to each sample member and used in place of name, address, or other personally identifiable information. All correspondence including any emails, phone calls, or other correspondence with a respondent will be logged and coded with a disposition based on the reason for the contact. This process ensures that all sample members are accounted for and given the proper disposition code in line with American Association for Public Opinion Research (AAPOR) and Council of American Survey Research Organizations (CASRO) guidelines. This will ultimately allow the FMG Team to appropriately calculate cooperation and response rates and track issues/problems with the survey effort. The tracking of disposition codes will allow for the creation of appropriate survey weights for eligible responders that are adjusted to account for sample members with unknown eligibility and eligible nonrespondents where appropriate.

**Data Verification and Cleaning.** Once data collection has been completed and all survey data entered, the datasets will be reviewed and thoroughly checked before any analyses are conducted. Records are inspected to determine whether any completed cases should be discarded. These data quality control checks are made to assure that the analysis file is clean. The Figure below details the minimum steps taken. Our full quality control checklist for online surveys is included in Appendix B.

Data cleaning steps taken prior to analysis	
1) Receive datasets	9) Check skip patterns
2) Print format library (file information)	10) Check recodes
3) Run frequencies (weighted &	11) Check calculated variables

unweighted)	
4) Check variable names	12) Check coding of 'other, specify'
5) Check variable labels	13) Address problems
6) Check value labels	14) Make changes to formats
7) Check weights (against known pop totals)	15) Secondary review of final dataset
8) Check unweighted sampling	16) Recheck all resultant values

## Estimation procedures (poststratification, regression, weighting)

The IRS intends to use the consumer tipping data from the proposed survey in a number of ways. One of those ways will be to develop subnational, industry-specific tipping rates. This section provides a discussion of how Fors Marsh Group will develop those rates from the survey data. Other methods may be used for analyses not described here.

**Multilevel Regression and Poststratification (MRP).** One means of obtaining both nationally and sub nationally representative estimates of tipping and stiffing rates is MRP (Gelman & Little, 1997<sup>3</sup>; see Buttice and Highton, 2013<sup>4</sup> and Toshkov, 2015<sup>5</sup>, for recent reviews and critiques). Model based poststratification strategies have been employed to generate estimates that conform to administrative data using non-representative samples<sup>6</sup>. MRP has attained popularity by social scientists who wish to obtain geographically disaggregated estimates of a quantity of interest. Awareness of variation in tipping rates faced by establishments in different parts of the country will be of potential use for the IRS in so far as it provides a general understanding of patterns of tipping behavior and it might help detect differences in compliance.

Analyzing consumer tipping data for a particular industry using MRP would first involve estimating models of the number of transactions undertaken by consumers as well as their tipping behavior that take the form:

$$\widehat{E}_{ik} = \beta_E X_{ik} + \alpha_E G_k + C_E$$

$$\widehat{S}_{tik} = \beta_S X_{ik} + \alpha_S G_k + C_S$$

$$\widehat{T}_{tik} = \beta_T X_{ik} + \alpha_T G_k + C_T$$

<sup>3</sup> Gelman, A., & Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology*, 23(2): 127-135.

<sup>4</sup> Buttice, M. K., & Highton, B. (2013). How does multilevel regression and poststratification perform with conventional national surveys?. *Political Analysis*, 21(4), 449-467.

<sup>5</sup> Toshkov, D. (2015). Exploring the Performance of Multilevel Modeling and Poststratification with Eurobarometer Data. *Political Analysis*, mpv009.

<sup>6</sup> Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31(3), 980-991.

Goel, S., Obeng, A., & Rothschild, D. Non-Representative Surveys: Fast, Cheap, and Mostly Accurate. Working Paper

Where  $\widehat{E}_{ijk}$  is the expected total number of transactions engaged in by respondent  $i$  in location  $k$ ;  $\widehat{S}_{tik}$  is the expected probability that respondent's transaction  $t$  was tipped; and  $\widehat{T}_{tik}$  is an expected tip rate for transaction  $t$  calculated by dividing a reported dollar amount in tips by transaction bill size;  $X$  is a set of observable respondent-level demographic variables such as race, socioeconomic status, etc., that are likely to be correlated with both tipping behavior and the number of transactions; and  $G$  is a set of location-specific factors such as whether the location is part of a rural or urban region that capture variability in the number of transactions and tipping behavior by sector that is not explained by differences in  $X$  between locations. Note that while the location  $k$  is the most narrowly defined geographic area for which data is available, predictions can be generated for aggregated levels of geography  $g$ . Finally,  $C$  is a constant.

After estimating model parameters  $\beta$ ,  $\alpha$ , and  $C$ , predictions of  $\widehat{E}, \widehat{S}, \wedge \widehat{T}$  are generated for strata defined by all  $N$  combinations of values of  $X$  and  $G$  covariates<sup>7</sup>. Poststratification is then used to generate a transaction average tipping rate for a given location:

$$\widehat{T}_g = \sum_s^N \frac{\widehat{E}_s \widehat{S}_s P_s}{\sum_s^N \widehat{E}_s \widehat{S}_s P_s} \widehat{T}_s$$

Where  $P$  is the population of a given demographic/geographic stratum  $s$  in a given location  $g$ , taken from ACS/census data. In the empirical exercise based on the pilot data below, county level geographic factors are used to model individuals' number of transactions and tipping behavior. Predictions are generated for commuting zones, which are more likely to encompass the customer base of a given establishment. Commuting zones have been used in recent, prominent studies to define the geographic extent of environmental determinants of social outcomes.<sup>8</sup> Commuting zones may proxy for the typical geographic extent of respondents' daily travels, and thus the establishments they are likely to visit.

The average tipping rate for a given location is thus estimated to be the average tipping rate across all strata, weighted by the estimated proportion of tips made in each stratum. This regression and poststratification procedure would be undertaken separately for each industry. The benefit of using a linear, additive model to produce predictions for individual strata rather than using a non-linear model (e.g. logit for the tipping rate, Poisson/negative binomial for the transaction counts) is that, if the linear model provides reasonably accurate estimates of tipping and tipping rates, producing representative tipping and tipping rates for a given geography does not require that the set of strata be defined by the full list of individual predictors  $X$  (i.e. numbers of white males, age 18-34 with only high school education, with household income between \$25-35k) which is not available in

<sup>7</sup> Note that when predictions for the number of transactions and the tipping rate fall below zero, the predicted amounts are set to zero; when predictions for the probability that a transaction is tipped fall below zero or exceed one, the predicted probabilities are capped at zero or one as appropriate.

<sup>8</sup> Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, 129(4), 1,553-1,623.

standard sources such as Census and ACS. Rather, only geographic means of X are necessary to produce the estimates.

### **Implementation of Poststratification Strategy using the Pilot Study Data.**

As part of the planning for the full fielding of the consumer tipping survey, FMG conducted a pilot study to assess the degree to which estimates of the national tipping rate generated from data collected from a probability (GfK) and non-probability (Ipsos) sample would systematically differ from both each other as well as estimates based on point of sale data. In order to examine to what extent the proposed regression-based poststratification methodology can replicate the national tipping rate generated from the vendor's poststratification weights and to assess the sensitivity of the resulting national and subnational estimates to differences in the choice of sample, the regression and poststratification method is applied separately to Ipsos and GfK pilot data on transactions at Full Service Restaurants. If the resulting estimates for the two samples are similar, that would provide strong evidence that the poststratification methodology accounts for systematic differences, if any, between the probability and non-probability sample.

For the regression stage, models of the respondent's numbers of transactions as well as transaction-level models of whether the transaction was tipped and the tipping rate for tipped transactions are estimated separately for the Ipsos and GfK respondents. Unweighted and vendor weighted descriptive statistics for the predictors and outcome variables for the GfK and Ipsos samples are presented in Tables 1 through 4 in Appendix C. While, due to the limited sample size and time period represented by the pilot data, the mean tipping rate is not poststratified on season or day of the week, the poststratification of the full fielding data will attempt to account for these potential temporal imbalances. In addition, while in the full fielding the vendor weights may be used in the regression stage in order to examine the robustness of the estimates, in this exercise the regressions are unweighted due to the lack of documentation for the GfK vendor weights. Given the lack of transparency there is the potential for a lack of comparability in the post-stratification procedures of the two vendors. Creating our own weights allows us to avoid this potential problem and provides additional clarity to the analyses.

For the poststratification stage, data from the 2013 5 year ACS is used to define the target population. Demographic/geographic strata are defined as age-gender-education-county. To these strata are appended additional county-level analogues to the individual-level predictors, including the racial/ethnic composition of the county (% of county population which is non-Hispanic white, non-Hispanic black, Hispanic, and other) and the fraction of the county's population who fall into households of a given income share. Geographic variables include the urban-rural status of the respondent's county, the fraction of the country's population which was born outside the United States, and the respondent's census division..

Resulting estimates for the national transaction mean tipping rates are presented in the table below. Below each point estimate in parentheses is a standard error, which is calculated through a cluster jackknife procedure where the regression and poststratification is replicated one time for each commuting zone in the sample and relevant statistics are calculated excluding one commuting zone at a time. The

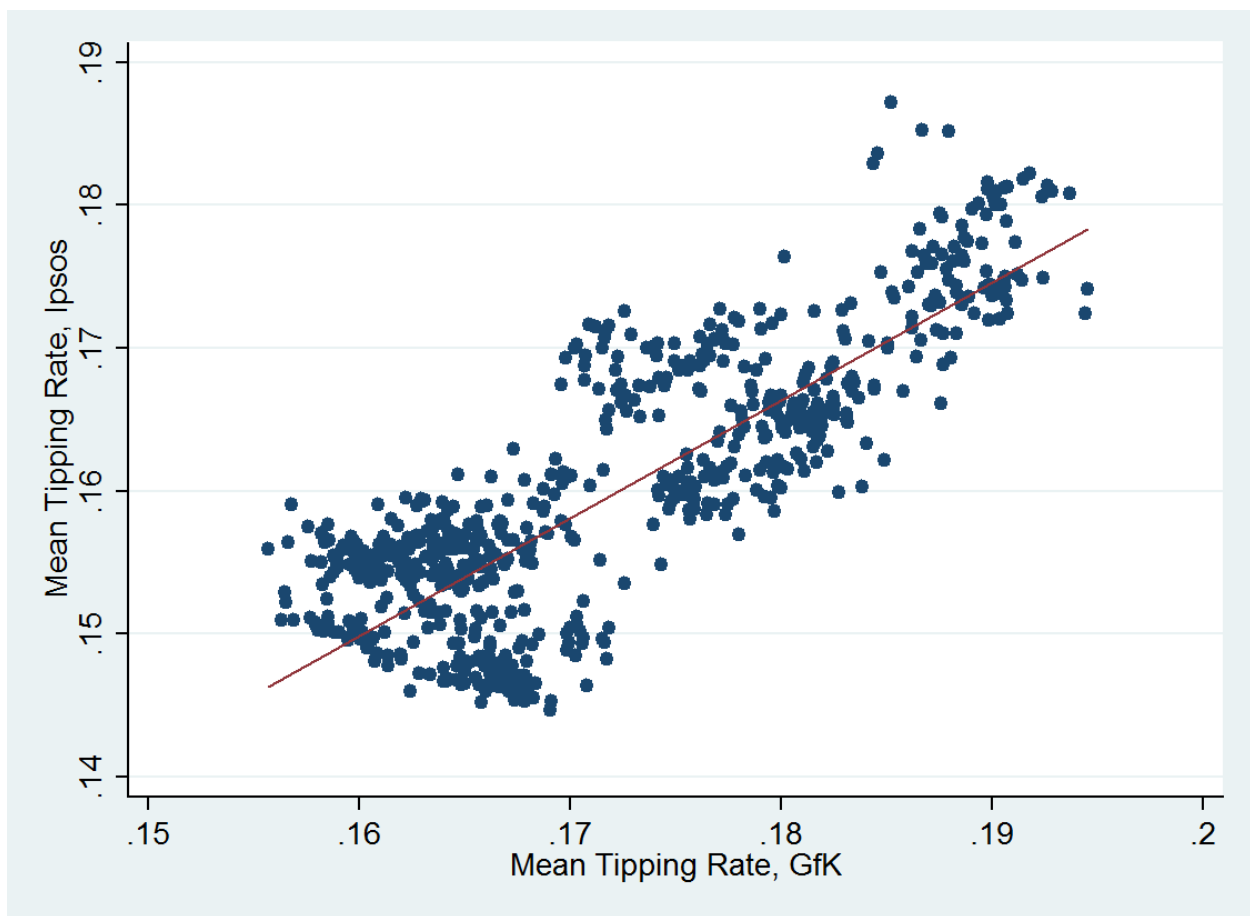


resulting point estimates for the national tipping rate are approximately 18% and 17% for the GfK and Ipsos samples, respectively, and are similar to the tipping rates obtained using the vendors' poststratification weights. The standard error for the difference between the estimates is statistically insignificantly different from zero.

Transaction Mean Tipping Rates by Sample			
	GfK	Ipsos	Difference (Ipsos - GfK)
<b>Tipping Rate</b>	18.2	17.2	-1.0
	(0.7)	(0.9)	(1.1)

Examining the geographic variation in tipping rates implied by the models, the correlation between the GfK and Ipsos commuting zone average tipping rates is .82. For each commuting zone, the difference in the tipping rate is calculated along with its standard error. The difference in estimated mean tipping rate is not statistically significantly different from zero for any commuting zone.

**Figure 1 - GfK versus Ipsos Mean Tipping Rates by Commuting Zone**



**Conclusion.** The results of this exercise do not provide strong evidence for systematic differences in national or subnational tipping rates for the populations

represented by the GfK or Ipsos samples once regression-based poststratification adjustments are applied to both samples. Given that small area estimates regression based poststratification methods have been found to be more reliable as the estimation sample increases<sup>9</sup>, it is possible that the discrepancy in the point estimates may simply be a function of sampling induced measurement error.

### **Outreach, advertising, and recruitment method**

Specific information regarding a panel’s recruitment sources is typically tightly held for several reasons. In addition to the economic considerations that disclosure of specifics could potentially harm the financial well-being of the panel by divulging information that is unique to the panel and as such a competitive advantage, there are also potential concerns regarding the privacy of the panel participants. As such, ESOMAR<sup>10</sup>, the World Association for Opinion and Market Research, has provided guidelines for the types of information that online panels should provide to potential users (which can be found [here](#) in questions 2-6). This information is provided below for the sources to be used on this project.

<b>Recruitment Sources Used in the Project</b>	
Ipsos iSay	<p>Our panels are not just lists or databases of individuals, but actively-managed research Access Panels:</p> <ul style="list-style-type: none"> <li>• Individuals who have volunteered to take part in market research surveys</li> <li>• Created and managed for long-term use and access</li> <li>• Extensively profiled to efficiently target respondents</li> </ul> <p>The vast majority of our panelists are referred to us through various online suppliers. We only use high quality recruitment sources to entice people who are eager to take surveys. We strategically focus on developing processes that reflect the newest internet practices as may currently be found through social networks. Email lists, banners, website and text ads, co-registration, and search engine marketing are also used.</p>
Lightspeed GMI	<p>This is an actively-managed panel composed of people who made a conscious decision to participate in online surveys through a double opt-in registration process.</p> <p>Several methodologies are used to recruit panelists, including opt-in email, co-registration, e-newsletter campaigns, and traditional banner placements, as well as both internal and external affiliate networks. Social media is included through our recruiting partners.</p>
MarketCube	<p>MarketCube owns and operates the Univox Community - an actively-managed panel with an individual-level compensation model. They also have access to a vast network of social media and publisher respondents that can be utilized to supplement internal assets.</p> <p>Additionally, MarketCube has developed close relationships with a variety of panel companies with whom they can partner on difficult-to-reach subpopulations. These strategic partnerships allow them to leverage relevant lists, databases, and networks to</p>

<sup>9</sup> Buttice, M. K., & Highton, B. (2013). How does multilevel regression and poststratification perform with conventional national surveys?. *Political Analysis*, 21(4), 449-467.

<sup>10</sup> <https://www.esomar.org/about-esomar.php>

	fulfill specific client requirements.
ROI Rocket	This large ad network has provided over 30MM panelists to date and offers access to over 5MM active respondents at any given time. They have experience in utilizing their sample for online communities, custom panels, in-depth interviews, longitudinal research studies, etc.
SSI	This is an actively-managed panel incorporating participants from partnership sources managed by SSI, recruited via banners, invitations and messaging. Prospects go through rigorous quality controls before being included in SSI panels.

## Existing research on panel comparison to benchmarks

Ipsos has maintained an opt-in panel since the early 1950's. Ipsos was an early adopter and leading vendor of research using opt-in panels since that time. Opt-in panels have panel members that were recruited by a variety of means. In the early years, people were approached using mailing lists, telephone lists, word of mouth, among a few methods. Panel members were mostly contacted by mail and sometimes by telephone. The accuracy of results from the Ipsos panel was periodically the focus of research. Some of the research considered key demographics and attitudes for telephone samples from the panel versus random digit dial (Groeneman, 1994), mail samples versus results from the General Social Survey (Putnam, 2000) and mail samples versus BRFSS (Pollard, 2002). In all cases, the Ipsos non-probability panel results matched well against their benchmark surveys.

The use of Ipsos's panel has moved away from mail and phone to online research. Research has been released comparing Ipsos results with benchmark survey results. One paper compared the results for Ipsos online panel to telephone reference samples for four different studies looking at Demographics, behaviors and attitudes, political values, voting intention and State-level polling. In this study, the online non-probability panel results were very similar to the benchmark samples (Young, Vidmar, Clark and El-Dash, 2012). Later on that year, Ipsos online panel was used in pre-election polls to follow the voter preference trends. Ipsos poll results were observed to be one of the most accurate of the online polls relative to the actual election results (Silver, 2012).

The quality of samples from Ipsos's opt-in panel continues to be a focus of research. Ipsos continues to focus on the breadth and depth of samples from its panel, blending of samples from multiple panels and incorporating river sampling (Young, Vidmar, Clark and El-Dash, 2012; Lewis and Choi, 2015).

## References

- Groeneman, S. (1994), "Multi-Purpose Household Panels and General Samples: How Similar and How Different?" Annual Meeting of the American Association for Public Opinion Research, Danvers, MA.
- Lewis, Z. H., and Choi, M. (2015) "Using Non-Probability Sampling Techniques to Track Seasonal Flu Activity", Annual Meeting of the American Association for Public Opinion Research, Hollywood, FL.
- Pollard, William (2002), "Use of Consumer Panel Survey Data for Public Health Communication Planning: An Evaluation of Survey Results," Proceeding of the Survey Research Methods Section.  
<https://www.amstat.org/sections/srms/Proceedings/y2002/Files/JSM2002-000768.pdf>
- Putnam, R.D. (2000). "The Collapse and Revival of American Community. Appendix I: Measuring Social Change," *Bowling Alone*, New York: Simon and Schuster.
- Roshwalb, A., El-Dash, N., and Young, C. A. (2013). "Towards the Use of Bayesian Credibility Intervals in Online Survey Results." Ipsos Public Affairs White Paper,  
[http://www.ipsos-na.com/dl/pdf/knowledge-ideas/publicaffairs/IpsosPA\\_POV\\_BayesianCredibilityIntervals.pdf](http://www.ipsos-na.com/dl/pdf/knowledge-ideas/publicaffairs/IpsosPA_POV_BayesianCredibilityIntervals.pdf)
- Silver, N. (2012) "Which Polls Fared Best (and Worst) in the 2012 Presidential Race," New York Times FiveThirtyEight Blog,  
[http://fivethirtyeight.blogs.nytimes.com/2012/11/10/which-polls-fared-best-and-worst-in-the-2012-presidential-race/?\\_r=0](http://fivethirtyeight.blogs.nytimes.com/2012/11/10/which-polls-fared-best-and-worst-in-the-2012-presidential-race/?_r=0)
- Young, C.A., Vidmar, J.P., Clark, J., and El-Dash, N. (2012). "Our Brave New World: Blended Online Samples and the Performance of Nonprobability Approaches," Ipsos Public Affairs White Paper, [http://www.ipsos.com/public-affairs/sites/www.ipsos.com.public-affairs/files/12-10-12\\_PA\\_Brave\\_New\\_World\\_WP.pdf](http://www.ipsos.com/public-affairs/sites/www.ipsos.com.public-affairs/files/12-10-12_PA_Brave_New_World_WP.pdf)



## Appendix A: Sample Balancing

### Nested Age\*Gender\*Region\*Income Balancing

Sampling Cell	Balancing %
Male 18-34 Northeast Under \$20K	0.333
Male 18-34 Northeast \$20K-49.9K	0.675
Male 18-34 Northeast \$50K-99.9K	1.095
Male 18-34 Northeast \$100K+	0.605
Male 18-34 Midwest Under \$20K	0.442
Male 18-34 Midwest \$20K-49.9K	0.898
Male 18-34 Midwest \$50K-99.9K	1.219
Male 18-34 Midwest \$100K+	0.562
Male 18-34 South Under \$20K	0.79
Male 18-34 South \$20K-49.9K	1.604
Male 18-34 South \$50K-99.9K	1.985
Male 18-34 South \$100K+	1.004
Male 18-34 West Under \$20K	0.491
Male 18-34 West \$20K-49.9K	1.043
Male 18-34 West \$50K-99.9K	1.33
Male 18-34 West \$100K+	0.655
Male 35-54 Northeast Under \$20K	0.286
Male 35-54 Northeast \$20K-49.9K	0.825
Male 35-54 Northeast \$50K-99.9K	1.351
Male 35-54 Northeast \$100K+	1.07
Male 35-54 Midwest Under \$20K	0.34
Male 35-54 Midwest \$20K-49.9K	0.955
Male 35-54 Midwest \$50K-99.9K	1.486
Male 35-54 Midwest \$100K+	0.863
Male 35-54 South Under \$20K	0.646
Male 35-54 South \$20K-49.9K	1.641
Male 35-54 South \$50K-99.9K	2.453
Male 35-54 South \$100K+	1.463
Male 35-54 West Under \$20K	0.396
Male 35-54 West \$20K-49.9K	0.992
Male 35-54 West \$50K-99.9K	1.531
Male 35-54 West \$100K+	1.011
Male 55+ Northeast Under \$20K	0.351
Male 55+ Northeast \$20K-49.9K	1.091
Male 55+ Northeast \$50K-99.9K	0.991
Male 55+ Northeast \$100K+	0.572
Male 55+ Midwest Under \$20K	0.391
Male 55+ Midwest \$20K-49.9K	1.326
Male 55+ Midwest \$50K-99.9K	1.264
Male 55+ Midwest \$100K+	0.608

Male 55+ South Under \$20K	0.75
Male 55+ South \$20K-49.9K	2.166
Male 55+ South \$50K-99.9K	2.04
Male 55+ South \$100K+	1.066
Male 55+ West Under \$20K	0.448
Male 55+ West \$20K-49.9K	1.183
Male 55+ West \$50K-99.9K	1.174
Male 55+ West \$100K+	0.626
Female 18-34 Northeast Under \$20K	0.311
Female 18-34 Northeast \$20K-49.9K	0.656
Female 18-34 Northeast \$50K-99.9K	1.001
Female 18-34 Northeast \$100K+	0.516
Female 18-34 Midwest Under \$20K	0.415
Female 18-34 Midwest \$20K-49.9K	0.846
Female 18-34 Midwest \$50K-99.9K	1.335
Female 18-34 Midwest \$100K+	0.565
Female 18-34 South Under \$20K	0.745
Female 18-34 South \$20K-49.9K	1.5
Female 18-34 South \$50K-99.9K	2.352
Female 18-34 South \$100K+	1.095
Female 18-34 West Under \$20K	0.474
Female 18-34 West \$20K-49.9K	1.021
Female 18-34 West \$50K-99.9K	1.413
Female 18-34 West \$100K+	0.662
Female 35-54 Northeast Under \$20K	0.24
Female 35-54 Northeast \$20K-49.9K	0.784
Female 35-54 Northeast \$50K-99.9K	1.209
Female 35-54 Northeast \$100K+	0.696
Female 35-54 Midwest Under \$20K	0.295
Female 35-54 Midwest \$20K-49.9K	0.934
Female 35-54 Midwest \$50K-99.9K	1.714
Female 35-54 Midwest \$100K+	0.915
Female 35-54 South Under \$20K	0.555
Female 35-54 South \$20K-49.9K	1.625
Female 35-54 South \$50K-99.9K	3.054
Female 35-54 South \$100K+	1.851
Female 35-54 West Under \$20K	0.357
Female 35-54 West \$20K-49.9K	0.982
Female 35-54 West \$50K-99.9K	1.745
Female 35-54 West \$100K+	1.162
Female 55+ Northeast Under \$20K	0.389
Female 55+ Northeast \$20K-49.9K	1.385
Female 55+ Northeast \$50K-99.9K	1.203

<b>Female 55+ Northeast \$100K+</b>	<b>0.579</b>
<b>Female 55+ Midwest Under \$20K</b>	<b>0.484</b>
<b>Female 55+ Midwest \$20K-49.9K</b>	<b>1.642</b>
<b>Female 55+ Midwest \$50K-99.9K</b>	<b>1.486</b>
<b>Female 55+ Midwest \$100K+</b>	<b>0.632</b>
<b>Female 55+ South Under \$20K</b>	<b>0.844</b>
<b>Female 55+ South \$20K-49.9K</b>	<b>2.683</b>
<b>Female 55+ South \$50K-99.9K</b>	<b>2.413</b>
<b>Female 55+ South \$100K+</b>	<b>1.125</b>
<b>Female 55+ West Under \$20K</b>	<b>0.462</b>
<b>Female 55+ West \$20K-49.9K</b>	<b>1.518</b>
<b>Female 55+ West \$50K-99.9K</b>	<b>1.373</b>
<b>Female 55+ West \$100K+</b>	<b>0.672</b>



## Appendix B: Quality Control Checklist

### Online Survey QC Checklist and Timeline

**Dataset to be Reviewed:**

**Reviewer:**

Date Begun:

Date Completed:



Online Survey QC Checklist and Timeline							
Project Name:			Responsible:				
Timeframe	Task	Minimum # of days*	Ops Team	Research Team	ALWAYS REQUIRED*:	Checked by:	Date Completed:
I. Before online survey is programmed, finalize the AQ:	1. Finalize Annotated Questionnaire (AQ) content	2 days		X	Yes		
	a. Check that question type matches the client needs and project team's expectations	-		X	Yes		
	b. Check that skip patterns match the client needs and project team's expectations	-		X	Yes		

	c. Check that skip patterns make logical sense looking for errors that would cause never ending loops, missed questions, or incorrect branching paths	-		X	Yes		
	d. Check that data requirements match the client needs and project team's expectations	-		X	Yes		
	e. Consult with programming team to determine days needed to program based on schedules and complexity	-	X	X	Yes		
	f. Research team builds a Skip Logic Checking Matrix (see example tab)	-		X	Yes		
	2. Technical editing review of AQ (no content changes; just spelling, grammar, punctuation)	1 day		X	Yes		
	3. Make edits to the AQ based on technical edit, to create the final AQ	1 day		X	Yes		
	4. After tech edits are finalized, Project Lead sends final AQ to Ops for programming	-	X	X	Yes		

II. After AQ is FINAL:	1. Meeting with survey checking team to review AQ and discuss testing objectives	1 day (first day of programming)	X	X	Yes		
		# of days based on earlier consultation in I.1.e. (see Online Survey Scheduling Guidelines for minimum business day requirements based on complexity)					
	2. Program the online survey		X	X	Yes		
III. After online survey is programmed, verify online question content match AQ:	1. Checking team prints paper AQ, and checks word by word for exact match of AQ and online survey items. Check for any typos, grammatical errors, and/or format issues:	4 days	X	X	Yes		
	a. Typos include misspelled words and spacing issues	-	X	X	Yes		
	b. Blatant grammatical errors	-	X	X	Yes		

	c. Format issues should include questions being consistent with other questions.	-	X	X	Yes		
	i. For example: If all questions except Q3 is in bold, this should be mentioned to programmer	-	X	X	Yes		
	ii. Generally, we follow the same format as the AQ. If something is underlined/in bold/italicized on the AQ, it should appear so in the online survey as well.	-	X	X	Yes		

	iii. Ensure that scales with the same response options (e.g., agree to disagree) are always oriented in the same manner (e.g., horizontal). In other words, do not display several horizontal agree/disagree scales and then a vertical agree/disagree scale. This causes issues with response patterns.	-	X	X	Yes		
	d. Check that question type online matches AQ	-	X		Yes		
	e. Check that question type matches the client needs and project team's expectations	-		X	Yes		
	g. Check that skip patterns online match AQ and Skip Logic Matrix	-	X		Yes		
	h. Check that skip patterns match the client needs and project team's expectations	-		X	Yes		

	2. Check all DATA RESTRICTIONS for open-ended boxes to make sure input matches question requirements	2 days	X	X	Yes		
	a. Check that data requirements match the client needs and project team's expectations	-		X	Yes		
	b. Check that data requirements online match AQ (data requirements should be specified in AQ by research team)	-	X		Yes		
	3. Log all survey edits/changes during QC process	ongoing during QC	X	X	Yes		
	4. Programmer reviews change log and makes revisions to online survey	1 day	X		Yes		
	5. Double Checker ensures that all previous changes listed in change log have been corrected and signed off on, as well as review for additional edits/errors in survey	2 days	X	X	Yes		
	6. If any additional changes have been found resolve with programmer and restart Question Verification Process	restart at III. 1. until no further revisions; then move to IV.	X	X	Yes		

IV. After online survey content is verified, take steps to verify the underlying data:	1. Verify all of the underlying data in the online survey: (note, in some online survey platforms, programmer does not have control over data labels)	2 days	X	X	Yes		
	a. Make sure to RECORD ALL ANSWERS when testing each survey (all testing should be done by two independent testers, 1 can be programmer if time permits)	-	X	X	Yes		
	b. Compare data output from programmer to make sure all data has been recorded properly	-	X	X	Yes		
	e. Add discrepancies to change log	-	X	X	Yes		
	c. Resolve discrepancies between tester's and data output's answers with programmer	-	X	X	Yes		
	d. Programmer makes edits/revisions	-	X		Yes		
	2. Check that all previous changes listed in change log have been corrected and signed off on	1 day	X	X	Yes		
	3. If any additional changes have been found resolve with programmer and restart data verification process	restart at IV. 1. until no further revisions	X	X	Yes		

	4. Open survey to check email campaign and survey using live data to ensure correct values are piped in.		X	X	Yes		
	a. Use 5 cases from the actual upload file that have been prepared but send internally to QC team responsible for review.		X	X	Yes		
	b. Make sure piping looks correct in both emails and on survey (if applicable). Identify and review all areas that piping occurs.		X	X	Yes		
	5. If any additional changes have been found resolve with programmer and restart data verification process	restart at IV. 4. until no further revisions	X	X	Yes		
	6. Remove test data from production survey						



## Appendix C: Descriptive Statistics Tables from Pilot

**Table 1 - Unweighted Descriptive Statistics - GfK Sample**

<b>Respondent-Level Variables</b>	<b>N</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Minimum</b>	<b>Maximum</b>
Full Service Restaurant Transactions in Last Day	5,663	0.20	0.44	0.00	4.00
Male	5,663	0.49	0.50	0.00	1.00
Age, Excluded Category = <i>18-24</i>					
<i>25-34</i>	5,663	0.16	0.37	0.00	1.00
<i>35-44</i>	5,663	0.15	0.35	0.00	1.00
<i>45-64</i>	5,663	0.39	0.49	0.00	1.00
<i>65+</i>	5,663	0.22	0.42	0.00	1.00
Age, Continuous	5,663	49.93	17.29	18.00	94.00
Educational Attainment, Excluded Category = <i>No High School Degree</i>					
<i>High School Graduate</i>	5,663	0.30	0.46	0.00	1.00
<i>Some College</i>	5,663	0.20	0.40	0.00	1.00
<i>Associate Degree</i>	5,663	0.09	0.29	0.00	1.00
<i>Bachelors Degree</i>	5,663	0.18	0.39	0.00	1.00
<i>Graduate Degree</i>	5,663	0.13	0.33	0.00	1.00
Race/Ethnicity, Excluded Category = <i>White</i>					
<i>Black</i>	5,662	0.10	0.30	0.00	1.00
<i>Hispanic</i>	5,662	0.10	0.30	0.00	1.00
<i>Other</i>	5,662	0.07	0.25	0.00	1.00
Income, Excluded Category = <i>Less than \$10,000</i>					
<i>\$10,000-\$14,999</i>	5,663	0.05	0.22	0.00	1.00
<i>\$15,000-\$24,999</i>	5,663	0.09	0.28	0.00	1.00
<i>\$25,000-\$34,999</i>	5,663	0.10	0.30	0.00	1.00
<i>\$35,000-\$49,000</i>	5,663	0.13	0.33	0.00	1.00
<i>\$50,000-\$74,999</i>	5,663	0.19	0.39	0.00	1.00
<i>\$75,000-\$99,999</i>	5,663	0.14	0.34	0.00	1.00
<i>\$100,000-\$149,000</i>	5,663	0.17	0.37	0.00	1.00
<i>\$150,000+</i>	5,663	0.08	0.27	0.00	1.00
% of Respondent's					

County Which is Foreign Born	5,658	0.12	0.10	0.00	0.51
Urbanization Status of Respondent's County, Excluded Category = <i>Metro areas of 1 million population or more</i>					
<i>Metro areas of 250,000 to 1 million population</i>	5,658	0.23	0.42	0.00	1.00
<i>Metro areas of fewer than 250,000 population</i>	5,658	0.10	0.30	0.00	1.00
<i>Nonmetro areas</i>	5,658	0.14	0.35	0.00	1.00
Census Division, Excluded Category = <i>New England</i>					
<i>Middle Atlantic</i>	5,658	0.13	0.34	0.00	1.00
<i>Midwest</i>	5,658	0.16	0.37	0.00	1.00
<i>West North Central</i>	5,658	0.08	0.27	0.00	1.00
<i>South Atlantic</i>	5,658	0.20	0.40	0.00	1.00
<i>East South Central</i>	5,658	0.05	0.23	0.00	1.00
<i>West South Central</i>	5,658	0.10	0.30	0.00	1.00
<i>Mountain</i>	5,658	0.07	0.26	0.00	1.00
<i>Pacific</i>	5,658	0.15	0.36	0.00	1.00
<b>Transaction-Level Variables</b>					
Was Transaction Tipped?	1,147	0.91	0.28	0.00	1.00
Tip Rate	924	0.18	0.06	0.01	0.42

**Table 2 - Weighted Descriptive Statistics - GfK Sample**

<b>Respondent-Level Variables</b>	<b>N</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Minimum</b>	<b>Maximum</b>
Full Service Restaurant Transactions in Last Day	5,663	0.20	0.45	0.00	4.00
Male	5,663	0.48	0.50	0.00	1.00
Age, Excluded Category = 18-24					
25-34	5,663	0.19	0.39	0.00	1.00
35-44	5,663	0.17	0.37	0.00	1.00
45-64	5,663	0.36	0.48	0.00	1.00
65+	5,663	0.17	0.38	0.00	1.00

Age, Continuous	5,663	46.87	17.36	18.00	94.00
Educational Attainment, Excluded Category = <i>No High School Degree</i>					
<i>High School Graduate</i>	5,663	0.30	0.46	0.00	1.00
<i>Some College</i>	5,663	0.20	0.40	0.00	1.00
<i>Associate Degree</i>	5,663	0.09	0.29	0.00	1.00
<i>Bachelor's Degree</i>	5,663	0.17	0.38	0.00	1.00
<i>Graduate Degree</i>	5,663	0.12	0.32	0.00	1.00
Race/Ethnicity, Excluded Category = <i>White</i>					
<i>Black</i>	5,662	0.11	0.32	0.00	1.00
<i>Hispanic</i>	5,662	0.15	0.36	0.00	1.00
<i>Other</i>	5,662	0.08	0.27	0.00	1.00
Income, Excluded Category = <i>Less than \$10,000</i>					
<i>\$10,000-\$14,999</i>	5,663	0.04	0.20	0.00	1.00
<i>\$15,000-\$24,999</i>	5,663	0.07	0.26	0.00	1.00
<i>\$25,000-\$34,999</i>	5,663	0.10	0.30	0.00	1.00
<i>\$35,000-\$49,000</i>	5,663	0.12	0.33	0.00	1.00
<i>\$50,000-\$74,999</i>	5,663	0.18	0.39	0.00	1.00
<i>\$75,000-\$99,999</i>	5,663	0.16	0.36	0.00	1.00
<i>\$100,000-\$149,000</i>	5,663	0.18	0.38	0.00	1.00
<i>\$150,000+</i>	5,663	0.08	0.27	0.00	1.00
% of Respondent's County Which is Foreign Born	5,658	0.12	0.10	0.00	0.51
Urbanization Status of Respondent's County, Excluded Category = <i>Metro areas of 1 million population or more</i>					
<i>Metro areas of 250,000 to 1 million population</i>	5,658	0.22	0.41	0.00	1.00
<i>Metro areas of fewer than 250,000 population</i>	5,658	0.09	0.28	0.00	1.00
<i>Nonmetro areas</i>	5,658	0.15	0.36	0.00	1.00
Census Division, Excluded Category = <i>New England</i>					
<i>Middle Atlantic</i>	5,658	0.14	0.34	0.00	1.00
<i>Midwest</i>	5,658	0.14	0.35	0.00	1.00

<i>West North Central</i>	5,658	0.07	0.26	0.00	1.00
<i>South Atlantic</i>	5,658	0.20	0.40	0.00	1.00
<i>East South Central</i>	5,658	0.06	0.23	0.00	1.00
<i>West South Central</i>	5,658	0.11	0.32	0.00	1.00
<i>Mountain</i>	5,658	0.07	0.26	0.00	1.00
<i>Pacific</i>	5,658	0.16	0.37	0.00	1.00
<b>Transaction-Level Variables</b>					
Was Transaction Tipped?	1,147	0.90	0.30	0.00	1.00
Tip Rate	924	0.18	0.06	0.01	0.42

**Table 3 - Unweighted Descriptive Statistics - Ipsos Sample**

<b>Respondent-Level Variables</b>	<b>N</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Minimum</b>	<b>Maximum</b>
Full Service Restaurant Transactions in Last Day	6,920	0.17	0.43	0.00	8.00
Male	6,878	0.46	0.50	0.00	1.00
Age, Excluded Category = 18-24					
25-34	6,878	0.18	0.39	0.00	1.00
35-44	6,878	0.16	0.36	0.00	1.00
45-64	6,878	0.44	0.50	0.00	1.00
65+	6,878	0.12	0.32	0.00	1.00
Age, Continuous	6,878	46.30	15.78	18.00	105.00
Educational Attainment, Excluded Category = No High School Degree					
High School Graduate	6,828	0.21	0.40	0.00	1.00
Some College	6,828	0.26	0.44	0.00	1.00
Associate Degree	6,828	0.12	0.32	0.00	1.00
Bachelor's Degree	6,828	0.25	0.43	0.00	1.00
Graduate Degree	6,828	0.14	0.34	0.00	1.00
Race/Ethnicity, Excluded Category = White					
Black	6,781	0.08	0.26	0.00	1.00
Hispanic	6,781	0.08	0.28	0.00	1.00
Other	6,781	0.08	0.27	0.00	1.00
Income, Excluded Category = Less than					

\$10,000-\$14,999	6,530	0.06	0.23	0.00	1.00
\$15,000-\$24,999	6,530	0.12	0.32	0.00	1.00
\$25,000-\$34,999	6,530	0.11	0.31	0.00	1.00
\$35,000-\$49,000	6,530	0.14	0.34	0.00	1.00
\$50,000-\$74,999	6,530	0.19	0.40	0.00	1.00
\$75,000-\$99,999	6,530	0.12	0.33	0.00	1.00
\$100,000-\$149,000	6,530	0.12	0.33	0.00	1.00
\$150,000+	6,530	0.06	0.24	0.00	1.00
% of Respondent's County Which is Foreign Born	6,914	0.12	0.10	0.00	0.51
Urbanization Status of Respondent's County, Excluded Category = <i>Metro areas of 1 million population or more</i>					
<i>Metro areas of 250,000 to 1 million population</i>	6,914	0.22	0.42	0.00	1.00
<i>Metro areas of fewer than 250,000 population</i>	6,914	0.09	0.29	0.00	1.00
<i>Nonmetro areas</i>	6,914	0.13	0.34	0.00	1.00
Census Division, Excluded Category = <i>New England</i>					
<i>Middle Atlantic</i>	6,914	0.16	0.36	0.00	1.00
<i>Midwest</i>	6,914	0.18	0.38	0.00	1.00
<i>West North Central</i>	6,914	0.07	0.25	0.00	1.00
<i>South Atlantic</i>	6,914	0.20	0.40	0.00	1.00
<i>East South Central</i>	6,914	0.05	0.22	0.00	1.00
<i>West South Central</i>	6,914	0.08	0.28	0.00	1.00
<i>Mountain</i>	6,914	0.07	0.25	0.00	1.00
<i>Pacific</i>	6,914	0.14	0.35	0.00	1.00
<b>Transaction-Level Variables</b>					
Was Transaction Tipped?	1,144	0.88	0.32	0.00	1.00
Tip Rate	909	0.18	0.06	0.01	0.48

**Table 4 - Weighted Descriptive Statistics - Ipsos Sample**

Respondent-Level Variables	N	Mean	Standard Deviation	Minimum	Maximum
Full Service	6,824	0.17	0.44	0.00	8.00

Restaurant Transactions in Last Day					
Male	6,824	0.48	0.50	0.00	1.00
Age, Excluded Category = 18-24					
25-34	6,824	0.18	0.38	0.00	1.00
35-44	6,824	0.15	0.36	0.00	1.00
45-64	6,824	0.44	0.50	0.00	1.00
65+	6,824	0.11	0.31	0.00	1.00
Age, Continuous	6,824	45.74	15.96	18.00	105.00
Educational Attainment, Excluded Category = No High School Degree					
High School Graduate	6,824	0.37	0.48	0.00	1.00
Some College	6,824	0.20	0.40	0.00	1.00
Associate Degree	6,824	0.09	0.29	0.00	1.00
Bachelor's Degree	6,824	0.18	0.39	0.00	1.00
Graduate Degree	6,824	0.11	0.31	0.00	1.00
Race/Ethnicity, Excluded Category = White					
Black	6,757	0.11	0.32	0.00	1.00
Hispanic	6,757	0.15	0.35	0.00	1.00
Other	6,757	0.07	0.26	0.00	1.00
Income, Excluded Category = Less than \$10,000					
\$10,000-\$14,999	6,530	0.05	0.22	0.00	1.00
\$15,000-\$24,999	6,530	0.11	0.32	0.00	1.00
\$25,000-\$34,999	6,530	0.11	0.31	0.00	1.00
\$35,000-\$49,000	6,530	0.13	0.33	0.00	1.00
\$50,000-\$74,999	6,530	0.19	0.39	0.00	1.00
\$75,000-\$99,999	6,530	0.11	0.31	0.00	1.00
\$100,000-\$149,000	6,530	0.15	0.35	0.00	1.00
\$150,000+	6,530	0.07	0.25	0.00	1.00
% of Respondent's County Which is Foreign Born	6,818	0.13	0.11	0.00	0.51
Urbanization Status of Respondent's County, Excluded Category = Metro areas of 1 million population or more					
Metro areas of 250,000 to 1 million	6,818	0.22	0.41	0.00	1.00

<i>population</i> Metro areas of fewer than 250,000	6,818	0.08	0.28	0.00	1.00
<i>population</i> Nonmetro areas	6,818	0.15	0.36	0.00	1.00
Census Division, Excluded Category = New England					
<i>Middle Atlantic</i>	6,818	0.14	0.35	0.00	1.00
<i>Midwest</i>	6,818	0.16	0.36	0.00	1.00
<i>West North Central</i>	6,818	0.06	0.23	0.00	1.00
<i>South Atlantic</i>	6,818	0.22	0.41	0.00	1.00
<i>East South Central</i>	6,818	0.06	0.23	0.00	1.00
<i>West South Central</i>	6,818	0.10	0.29	0.00	1.00
<i>Mountain</i>	6,818	0.08	0.27	0.00	1.00
<i>Pacific</i>	6,818	0.16	0.36	0.00	1.00
<b>Transaction-Level Variables</b>					
Was Transaction Tipped?	1,144	0.88	0.32	0.00	1.00
Tip Rate	909	0.18	0.06	0.01	0.48