Supporting Statement
Approval Request to Conduct Pilot Test of Consumer Tipping Survey (OMB #1545-xxxx)

A. JUSTIFICATION
1. Circumstances Making the Collection of Information Necessary

The IRS is charged with collecting revenue legally owed to the federal government. One important category of income comes in the form of tips. Previous empirical research has shown income from tips to be significantly underreported[1], limiting the IRS's ability to collect the proper amount of tax revenue. The IRS believes a new study of consumer tipping practices is needed in order to better understand current tip reporting behavior so tax administrators and policy makers can make the tax system fairer and more efficient. Therefore, the IRS wishes to develop updated estimates of consumer tipping revenue across numerous services where tipping is prevalent.

In support of this mission, IRS is seeking a standard clearance to conduct a one-month pilot test in preparation for a nation-wide consumer tipping survey. There exists a substantial difference in the cost per response between a probability and non-probability sample. Pilot tests are therefore necessary to determine the relative accuracy and selection bias of tipping data that are collected using these different sampling methodologies in order to determine if there is tradeoff between accuracy and cost. The results of the pilot will be used to determine the sampling method employed in a nation-wide survey.

This initiative flows from Goal 1 of the IRS Strategic Plan for FY 2014-2017: Deliver high quality and timely service to reduce taxpayer burden and encourage voluntary compliance. This is also a result of Executive Order 12862 (Setting Customer Service Standards, issued on September 11, 1993) that requires all government agencies to survey their customers. Pursuant to Executive Order 12862, agencies that provide significant services to the public must survey customers to determine the kind and quality of services they want and their level of satisfaction with existing services. Executive Order 13571 (Streamlining Service Delivery and Improving Customer Service, issued on April 27, 2011) expands the definition of "customer" and encourages the use of a broader set of tools to solicit actionable, timely customer feedback to capture insights and identify early warning signals. As used in Executive Order 13571, the term "customer" refers to any individual or to any entity, including a business, tribal, state or local government, or other agency, to which the agency directly provides significant services.

2. Purpose and Use of the Information Collection

The current research project is aimed at conducting a pilot test of a survey instrument using both probability and nonprobability samples. The data collected from these samples will be

---

[1] See IRS Publication 1530 (1990), Tip Income Study, "A Study of Tipping Practices in the Food Service Industry for 1984" and Robert B. Pearl and Kevin F. McCrohan, (1983), Estimates of Tip Income in Eating Places, 1982, IRS Research Bulletin, pp. 49-53.

adjusted for non-response and/or post-stratified to reflect the population of interest and used to generate estimates of tipping rates by industry category. Estimated restaurant tipping rates will be compared to those generated from point-of-sale data. The degree to which the estimates generated from the probability and non-probability samples agree with the POS estimates will determine which sampling method is employed for the nation-wide survey. In addition, an analysis of item non-response in the pilot surveys and the results of usability tests may be used to modify the survey instrument to maximize completion rates in the nation-wide survey.

The potential target population for the IRS tipping study includes all U.S. resident persons who use services that are commonly tipped. A precise estimate of the number of individuals in this population is unknown, but likely includes a majority of the U.S. adult population. Example settings where tipping is typical include: full-service restaurants, taxis, barber shops, beauty salons, hotels, and casinos.

The private nature of most transactions involving tipping makes it extremely difficult to collect reliable data that can be used to estimate total tip income. This difficulty is further compounded by the motivation of some individuals not to report tips received as taxable income. For these reasons, the IRS has concluded that surveying consumers about their tipping experiences is the most reliable way to collect quantitative data on tip income. Prior IRS research on consumer tipping behavior found tipping rates varied considerably by industry and by region. A 1982 study conducted by the University of Illinois for the IRS[2] found tipping rates to be roughly 14% of the total bill for restaurants, 12% for barber and beauty shops, 19% for bars, and 20% for taxis. On a regional basis, mean restaurant tipping rates ranged from a low of 12.5% in the West North Central to a high of 15% in the Northeast.

The observed variation in tipping rates implies larger sample sizes are required in order to produce accurate estimates of tipping rates. Other things being equal, a larger sample size means greater cost. This constraint may be met in two ways: (1) limiting the scope of the study to focus on fewer industries/regions or (2) finding a more cost-effective mode of data collection. Because of the previous study's finding on the variance of tipping rates by industry and region, the IRS believes it would be inappropriate to limit the scope in these manners.

With respect to lowering the cost of data collection, an increasingly common alternative is the use of non-probability Internet samples. The costs of sampling from an opt-in Internet panel may be substantially lower than the costs associated with sampling from a telephone- or mail-based frame, or a panel recruited from such frames (e.g., probability based web panel). In addition, there might be additional costs or non-response associated with pushing individuals sampled from the telephone or mail frame to the Internet survey instrument. The chief drawback of using a non-probability sample from an Internet opt-in panel is that such panels could produce a realized sample that is less representative of the target population than the phone or mail frames[3]. However, given the high rates of non-response associated with sampling from phone or mail frames, it is not clear to what degree respondents from probability samples are more representative with respect to tipping behavior than respondents contacted through an opt-in Internet panel, particularly after post-stratifying on observed demographic characteristics. Although non-response can be mitigated through follow-up contacts, this exacerbates the

---

[2] Pearl, R. B., & Sudman, S. (1983, June). *A survey approach to estimating the tipping practices of consumers* (Final Report to the Internal Revenue Service under Contract TIR 81-52); Pearl, R. B. (1985, July). *Tipping practices of American households: 1984* (Final Report to the Internal Revenue Service under Contract 82-21).

[3] AAPOR (2013). "Report of the AAPOR Task Force on Non-Probability Sampling." https://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/ NPS_TF_Report_Final_7_revised_FNL_6_22_13.pdf

differences between the probability and non-probability sampling strategies with respect to the cost of obtaining a sample of a given size. Consequently, given a fixed budget it is unclear whether the reductions in bias in the estimates of mean tipping and stiffing rates that result from using a probability sample is worth the increase in the variability in these estimates that results from a smaller sample size, especially for relatively infrequent tipping transactions.

Past studies that have examined the relative degree to which estimates derived from probability and non-probability conform to benchmark administrative data have often concluded that probability based estimates produce more accurate estimates[4]. One limitation of these prior studies, however, is that they only examine a limited set of benchmarks and surveys. Other studies, particularly of election outcomes, have found that non-probability samples compare favorably with respect to the accuracy of their estimates of benchmark outcomes[5]. Without knowing how the observable and unobservable characteristics of the populations from which the probability and non-probability samples are drawn differ and relate to tipping, it is difficult to draw conclusions from this prior literature concerning relative bias.

Given the uncertainty in the tradeoff between variance and bias in estimated tipping rates between a probability and non-probability sample, this consumer tipping study will follow Office of Management and Budget (OMB) guidelines[6] by using a pilot survey to resolve this conflict. Specifically, we will conduct a pilot study to determine if the results generated by two different Internet-based data streams—a probability-based sample derived from the GfK KnowledgePanel and a non-probability based sample taken from Ispos' i-Say online opt-in panel—produce equivalent estimates. This will allow the IRS to estimate the degree to which there is a difference in bias that results from the use of a non-probability sample versus a probability sample. One benefit of using these two panels is that they both make use of a web-based interface which should facilitate reduced respondent burden, lower item non-response rates, and greater response accuracy than mail- or phone-based surveys.

Fors Marsh Group (FMG), who will administer the pilot test (if approved), worked in close collaboration with a number of outside Subject Matter Experts throughout the course of this project. These experts were: Dr. Michael Lynn, a professor at Cornell who has published extensively on tipping related behaviors; Dr. James Alm, a professor at Tulane who specializes in public finance and tax compliance research; and Dr. John Vidmar, a chairman for Ipsos who has a long history of research studies and has worked closely with the IRS on a number of tipping-related reports. These SMEs consulted on numerous aspects of the project including the initial background research, the analysis reports, and questionnaire design.

3. **Consideration Given to Information Technology**

The study will use a web survey to collect data.

---

[4] See Yeager S., Krosnick J.A., Chang L., Javitze H.S., Levendusky M.T. Simpser A., Wang R. (2011) "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples." Public Opinion Quarterly 75(4): 709-747.

[5] Wang W., Rothschild D., Goel S., and Gelman A. (2014) "Forcasting elections with non-representative polls." International Journal of Forecasting, forthcoming.

[6] See Office of Management and Budget (2006). *Questions and answers when designing surveys for information collections*. Page 16, Section 22: "An agency may also use a pilot study to examine potential methodological issues and decide upon a strategy for the main study."

The consumer tipping questionnaires include screening criteria and pull down-menus for transaction types to reduce respondent burden. Copies of the survey questionnaire and sample invitation letters are attached.

4. Duplication of Information

This study is the only one being conducted to inform IRS staff about national tipping rates by industry.

5. Reducing the Burden on Small Entities

N/A

6. Consequences of Not Conducting Collection

Not conducting a pilot study would cause the IRS to make a less informed decision concerning the best sampling methodology to use for efficient data collection in the nation-wide survey and, ultimately, for obtaining accurate estimates of consumer tipping behavior by industry. The usefulness of the nation-wide survey data for tax administration purposes would be seriously compromised.

7. Special Circumstances

There are no special circumstances. The information collected will be voluntary. These statistics will be used in the development of a larger annual survey to be conducted at a later date.

8. Consultations with Persons Outside the Agency

IRS is working with Fors Marsh Group (FMG) to administer the survey.

9. Payment or Gift

The IRS is not providing payment or gift to survey respondents. However, the Ipsos panel has a reward system set up with their panel members based on their response to any survey administered through the Ipsos panel. GfK also has a reward system in place for members of their panel who participate in a survey.

10. Confidentiality

The security of the data used in this project and the privacy of survey respondents to the extent allowed by the law will be carefully safeguarded at all times. Security requirements are based on the Computer Security Act of 1987 and Office of Management and Budget Circular A-130, Appendices A & B. Physical security measures include a locked, secure office. Notes are stored in locked cabinets or shredded.

The IRS will apply and meet fair information and record-keeping practices to ensure privacy protection of all respondents to the extent allowed by law. This includes criteria for disclosure—laid out in the Privacy Act of 1974, the Freedom of Information Act, and Section 6103 of the Internal Revenue Code—all of which provide for the protection of respondent information as well as its release to authorized recipients.

The survey data will not contain any respondent names, Social Security Numbers, or Taxpayer Identification Numbers. Participants will not be identified in any of the documents or files used for this project. FMG will limit and control the amount of information we collect to those items that are necessary to accomplish the research questions. FMG will carefully safeguard the security of data utilized as well as the privacy of the survey respondents to the extent allowed by law. FMG will apply the fair information and record-keeping practices to ensure protection of all survey respondents. The criterion for disclosure laid out in the Privacy Act, the Freedom of Information Act, and section 6103 of the Internal Revenue Code provides for the protection of information as well as its releases to authorized recipients.

The survey will be administered electronically; however there are no cookies involved. Survey participants will be provided a link/web address via a secure website. Transmission to/from the secure website for the survey will be encrypted.

All panel groups involved in this study have privacy policies for their members to explain when and how any personal information can be shared, either for client use or as required by law. Ipsos and GfK privacy policies for their panel members can be viewed in full at the following locations:

GfK: http://join.knpanel.com/privacy2.html
Ipsos: http://www.i-say.com/Footerlinks/PrivacyPolicy/tabid/282/language/en-US/Default.aspx

11. Sensitive Nature

No questions will be asked that are of a personal or sensitive nature.

12. Burden of Information Collection

IRS will attempt to obtain 10,000 complete surveys for both the probability and non-probability web-panels, for a total of 20,000 complete surveys. Burden hours were calculated using a 13% overall response rate estimate.

RAS estimates that the total burden hours for this project come to 4,549.9 hours. The burden hour estimates breakdown as follows:

Category of Respondent/Activity	No. of Respondents	Participation Time	Burden Hours
Read Invitation Email	154,000	0.5 minute	1,283.3 hours
Read Reminder Email*	144,000	.25 minute	600 hours
Complete Survey	20,000	8.0 minutes	2,666.6 hours
		TOTAL BURDEN HOURS	**4,549.9 hours**

* The estimate for the Reminder emails is based on the assumption that 50% of the needed respondents will complete the survey online in time to not receive the Reminder email. Estimated time to complete survey is based on assuming 0.5 minutes per question (5 demographic questions + time to read and understand 7 tipping occasion fields) and an average 3.5 minutes per recorded expenditure with an average number of occasions within an assumed 1 day recall of .527 (sum of mean daily occasions across 6 service categories in Table 4 of Task Order #1

Final Report).

13. Costs to Respondents

N/A

14. Cost to Federal Government

The cost is $95,935. This cost is a portion of the total contract cost, which includes survey questionnaire development and purchase and analysis of electronic point of sale (POS) data.

15. Reason for Change

N/A

16. Tabulation of Results, Schedule, Analysis Plans

Once the survey instrument is finalized, the contractor will administer surveys using an online panel through Ipsos and GfK. The survey administration will include an initial invitation email with a link to the survey as well as reminder emails, as necessary for each panels' setup.

Based on estimates gathered in Task Order 1 concerning the likelihood of any service expenditure of interest within the previous 24 hours, it is estimated that up to 10,540 surveys will have a recorded transaction in tipping-related industries that would be captured by the survey. These captured expenditures will include non-tipped expenditures for services like fast-food and take-out dining. These estimates were based on an average of .527 occurrences per day, the summed likelihood that was derived from the final Report for Task Order 1[7]. The IRS requires 3,000 consumers to complete the survey over the course of a month with some form of tipping expenditure, 1,500 from Ipsos' non-probability system, and another 1,500 from GfK's probability-based panel. A restaurant frame obtained from a third party vendor with information for restaurant annual sales, mean bill size, and number of full service restaurants was used to calculate an estimate of the total number of transactions for full service restaurants in the United States. Using an estimated stiffing rate for full service restaurants obtained from the 1982[8] tipping study to obtain an estimate the total number of daily tips and the total population of the U.S. 18 and older taken from the ACS, we calculated a daily tipping incidence rate for full service restaurants of approximately 15%. Using this tipping incidence rate, we estimate that 10,000 respondents will be required to complete the survey in order to gather 1,500 surveys that have at least one tipped expenditure for each panel. Additionally, 10,000 respondents per panel will be the minimum necessary to obtain estimates of the frequency of certain infrequent transactions such as casino gambling. These frequency estimates will be used to determine the target sample size for the final survey.

Response rates for non-probability samples cannot be calculated since the probability is unknown. However, response rates can be calculated for the GfK's Knowledge Panel. In a pair of previous studies that were conducted using GfK's panel they reported that 14.6% and 13.2%

---

[7] *Estimating Consumer Tipping Behavior: Review and Recommendations*, February 2014. Internal report prepared for the Internal Revenue Service by Fors Marsh Group under contract TIRNO-13-Z-00021-001.

[8] Pearl, R. B., & McCrohan, K. F. (1984). Estimates of tip income in eating places, 1982. Statistics of Income Bulletin, 3(4), 49–53.

of respondents accepted the invitation to join the panel for that project, leading us to estimate that 14% would comparable for this effort. In those studies, 65.5% and 65.2% of invitations were accepted by panelists, so a selection rate of approximately 65% could be expected. The screening criterion for this study, that the respondents have had at least one tipped service occasion within the past 24 hours, will likely to lead to a 15% qualification rate. This qualification rate corresponds to the prior estimate that it would be necessary to gather 10,000 completed survey in order to gain 1,500 completed surveys with a tipped expenditure. This qualification criterion, when coupled with the recruitment and selections rates, will likely lead to a cumulative response rate of 1.4% for the probability-based sample.

Following the survey's administration, the contractor will provide FMG and the IRS with the survey data in the form of a generic respondent number (numerical; i.e. 1, 2, 3, 4, 5 …) and responses to the survey questions.

FMG will use the data to assess the relative validity of tipping rate estimates that result from the the GfK probability sample and the Ipsos non-probability sample. The results of this analysis will be provided to the IRS and used to adjudicate the use of a probability or non-probability sample for a large annual consumer tipping survey.

There is a question included in the survey that asks specifically what payment method they use for any voluntary tips that are made.[9] Respondents receive this question if they report that they made a voluntary tip for a service transaction, and are asked for the payment method for each voluntary tip they report. These questions will allow us to observe if there is any difference in magnitude between individuals who reported leaving a tip with an electronic payment methods (such as debit or credit cards, both of which are response options) or cash.

The analysis will make two comparisons, a "Differences" comparison where statistically significant differences in the tip rate between the probability and non-probability sample and a "Differences in Differences" test. Both tests can be read in full in the report supplied by FMG to the IRS outlining these tests.[10]

The GfK KnowledgePanel represents a benchmark with respect to probability-based panels because of its combination of a representative frame. Under the assumption that an estimate derived from a probability sample is at least as unbiased as that derived from a non-probability sample with respect to tipping behavior, then the choice of whether to use the probability or non-probability sample is, assuming equal variability in tipping rate between the populations represented in the two samples, reduced to a bias versus variance trade-off. Given that it is known that the cost-per complete will be lower with the non-probability sample, then if the samples do not differ with respect to tipping behavior, the non-probability sample can be said to be superior because of the larger potential sample size, and thus lower degree of sampling-related error in the final estimates. To test for similarities in tipping behavior between the two samples, what will subsequently be referred to as a "Difference in Samples" test, the following model will be estimated separately for each industry:

$$1 ¿ \hat{T}_{tj} = \delta\, Ipsos_{tj} + Constant$$

---

[9] The exact text for this question is as follows: "Q1_H. What payment type(s) did you use to pay the voluntary tip? (select all that apply)" with possible responses of 01 Cash, 02 Debit, 03 Credit, 04 Check, 05 Gift Card, 06 Smartphone credit or app, 07 Paper or online coupon (e.g., Groupon), 08 Non-monetary*, 09 Other.

[10] *Comparison of Estimates of Tipping Behavior Produced using Probability and Non-Probability Samples*, January 2015. Internal report prepared for the Internal Revenue Service by Fors Marsh Group under contract TIRNO-13-Z-00021-0002.

In Equation 1, $\hat{T}_{tj}$ is a tip rate greater than 0 of positive tip, positive bill size transaction t for an individual residing in location j; *Ipsos* is an indicator variable that takes a value of 1 if the respondent was part of the Ipsos i-Say panel and 0 if part of the GfK KnowledgePanel. Equation 1 allows for a test of an unconditional difference in tipping rates, i.e. systematic differences in tipping rates between the samples that can be driven by either differences in observed or unobserved demographic or geographic characteristics of respondents in the two samples. Specifically, a δ that is significantly different from 0 is consistent with unconditional differences in behavior between respondents from the two samples. Because of the small number of parameters (*k*=2) of this model, it allows for precise estimates of this unconditional difference even with small samples. Given that some industries are likely to have a small number of tipping transactions represented in the data, the parsimony of Equation 1 becomes a big advantage. Note that the assumption that the variance of the two samples is constant can be tested using standard tests for heteroscedasticity (e.g. Breusch-Pagan, Brown-Forsythe). The test for bias in the in the non-probability sample can be made robust to violation of equal variances through the use of robust standard errors.

The second test is a "Differences in Differences" test that compares the difference between mean tip rate of the non-probability sample and the benchmark POS data to the difference in the mean tip rate of the probability sample and the POS data. To estimate the unconditional "Differences in Differences," we can pool data for tipped transactions at full-service restaurants from the probability, non-probability, and POS samples (or a random subsample of the latter to mitigate computational complexity) and estimate the following model separately for each industry:

$$3 ¿ \hat{T}_{tj} = \delta\, Ipsos_{tj} + \vartheta\, GfK_{tj} + Constant$$

Equation 3 differs from Equation 1 in that it includes GfK_{tj}, an indicator variable that takes a value of 1 if a given transaction was extracted from the GfK sample and 0 otherwise. The excluded category is now transactions taken from the POS sample. Our null hypothesis can be stated as:

$$|\delta| = ¿\, \vartheta \vee ¿$$

This null hypothesis can be tested using a Wald or Likelihood Ratio Test. If the null hypothesis is rejected, we may conclude that the survey sample with the smaller absolute difference more closely matches the mean tip rate implied by the POS data.

Under the simplifying assumption that the observations that comprise the probability and non-probability samples are drawn independently from a normal distribution, with each sample containing approximately 1500 cases and each and that, due to the large number of observations, the POS mean can be treated as the population mean (i.e. not subject to sampling variability), then our analysis implies that the null hypothesis will be rejected approximately 80% of the time if the true standardized effect size of approximately .05 at alpha=.05 for the "difference" or "differences in differences" test.

The Point-of-Sale (POS) data was supplied by a company who specializes in POS devices and software. As part of their business they have access to a large database of restaurant transactions, so only this industry can be assessed using this POS data. An initial dataset of restaurant transaction data covering one month in 2014 was purchased to assess whether this

data could be used to assess the validity of tipping estimates in the pilot study. Two dimensions on which the POS data were assessed in the course of making this determination were (1) the degree to which the transactions represented in the POS data were subject to measurement error, or internal validity; and (2) the degree to which the transactions represented in the POS data were geographically representative of the general population of transactions, as determined by a comprehensive restaurant frame.

In assessing the degree of measurement error in the POS data, it is important to understand that the bill size (i.e., the check total minus any automatic gratuities) does not necessarily capture all the revenue obtained by the restaurant directly or indirectly from the customer. Services such as Groupon purchase meals from the restaurants to sell to customers, but the check total typically only represents transfers directly from the customer to the restaurant. The presence of such transactions in the data is supported by the fact that there were 10,226 transactions out of 66,508,684 with a recorded bill size of zero. The inability to record all of the revenue associated with a transaction can bias the tipping percentage and tip rates derived from the POS data if customers determine the amount tipped based on a conventional tipping rate applied to the total amount paid to obtain a meal (i.e., Groupon + direct payment to the restaurant). Specifically, this will tend to increase the estimated tipping rate and tipping percentage by underestimating the total cost of the meal (the denominator in Equation 1). Although we may mitigate the influence of these extreme tip amounts by weighting transactions by bill size, this will essentially remove the transaction from the data. Similarly, if the restaurant receives no direct revenue from the customer, than the denominator will be zero and both the tipping percentage and tipping rate will be undefined, and the tip rate will be excluded from the analysis. If the unobserved "true" tip rates for these excluded transactions are systematically different from tips with a larger observed bill size, then the tipping estimate will be biased, though it is unclear whether the mean tip rate absent these exclusions would differ significantly from that observed in the trimmed data set given the small number of exclusions.

With respect to measurement error, examination of the data and consultation with the POS vendor indicated potential error in both the amount tipped and/or the bill size for some transactions. Specifically, approximately 100% of cash transactions had a zero tip. Under the assumption that, conditional on the bill being paid in cash, the tip is also likely to be paid in cash, this indicates that cash tips are likely underreported in the POS data, leading to some tip amounts being erroneously recorded as zero. Even restricting the transactions to those with positive tips, the exclusion of cash tips may bias the aggregate estimate of the tip rate if tip rates systematically differ when the tip is paid in cash versus electronically. In addition, this problem makes it infeasible to use the POS data to estimate the stiffing rate and tipping percent (including zero and nonzero tips). With respect to measurement error in bill size, some transactions had negative, zero, or very small positive bill sizes. Consultation with our POS data vendor indicates that these transactions could result from refunds, prepayment methods such as Groupon, or training exercises conducted by the POS vendor. After excluding transactions that have very low or very large bill sizes relative to other transactions in the establishment, the average tip rate for full-service restaurants converges at a 21–23% tip rate, which is qualitatively similar to a 20% tipping convention, though higher than the 14.4% tip rate reported in Pearl and McCrohan (1984). However, to the degree that the transactions with zero or low tip rates represent real transactions, and not training data, and that tip rates for these transactions with mismeasured bill sizes differ from those with bill sizes closer to the establishment average, this aggregate tip rate may be a biased estimate of the population tip rate. High variability in between-establishment tip rates for quick-service restaurants and a relatively small number of such restaurants means that a precise estimate of the mean tipping rate cannot be estimated for quick-service restaurants.

With respect to temporal coverage of the POS data, the data comes from one calendar month, and thus may not be comparable to transactions conducted in other months. To mitigate this issue, for the pilot study, a separate POS data set will be purchased that covers the same span of time represented in the survey data. With respect to the geographic coverage of the POS data, the frequency of tipped transactions across Direct Marketing Areas (DMAs) was compared to the frequencies of all transactions as indicated by benchmark frame of full service restaurants obtained from a separate third party vendor. The geographic distribution of the POS tipped transactions did not systematically differ from that implied by the restaurant frame. There is thus little indication that the POS data is biased with respect to the geographic coverage.

Collectively, our analysis of the POS data implies that the data may be utilized to validate the tip rates reported in the pilot surveys for some transactions. Specifically, the POS data can be used to validate tip rates reported for transactions that took place at full-service restaurants where the tip and bill were both paid electronically at the point of sale. It would likely be less appropriate to compare responses for other types of transactions to the national and subnational mean tipping rates derived from the POS data. The consumer surveys are thus still likely the only source of information we will have about the relative frequency of tipped and non-tipped occasions.

17. Display of OMB Approval Date

N/A

18. Exceptions to Certification for Paperwork Reduction Act Submissions

N/A

19. Dates collection will begin and end

The data collection period for survey administration is scheduled to take place within two weeks of receiving survey approval, with a current schedule from June 1, 2015 to June 30, 2015.

B. STATISTICAL METHODS
1. Universe and Respondent Selection

The research will be conducted using Internet panels maintained by subcontractors Ipsos and GfK which have been designed to be representative of the adult population.

2. Procedures for Collecting Information

GfK Knowledge-Panel®: The KnowledgePanel is an Internet-based panel that uses a probability-based sampling strategy where the survey frame is derived from the USPS Delivery Sequence File. Individuals are invited to join the GfK KnowledgePanel by mail, followed by telephone calls for those who do not respond to the initial invitation. Once they have joined the panel they are invited to surveys and other projects via email. Households are sampled without replacement, avoiding potential bias that may result from respondents participating in the panel twice. For those individuals selected for participation without computers or an Internet connection, a netbook is provided. This process attempts to mitigate the selection bias associated with web surveys while preserving the benefits associated with a computer interface. The primary benefit of the KnowledgePanel relative to the opt-in panels described below is that knowing the probability of selection allows researchers to estimate error. However, these

estimates will always be deficient capturing all aspects of non-response unaddressed by demographic post-stratification. Further, the procedures used to setup and maintain panel membership and participation serve as an additional component of error difficult to fully model and correct for. Email invitations will be the only method used to sample respondents from the GfK KnowledgePanel for the Consumer Tipping Survey.

Blended Online Sample (Ipsos Ampario): Ipsos' blended sample approach combines the use of its Ampario online sampling method in addition to its ISAY online panel—an online panel of 800,000 members and their households. Ampario is a new nonprobability sampling procedure Ipsos has developed that invites respondents by invitations, banner ads, and other means on 100 to 400 websites that have partnered with Ipsos. These two methods are combined into a single sample using Ipsos' proprietary Cortex routing system, which allocates and reallocates a sample given respondent eligibility. Simply put, when respondents are not eligible for one survey, they are immediately redirected to other surveys in progress. In traditional one-off opt-in surveys, noneligible respondents are lost, representing a considerable cost. Finally, Bayesian methodology, which requires previous information regarding the overall sample of interest in order to mix with current information for the final distribution of results, is used to form the final distribution. As is the case with a traditional online sample, Ipsos' blended sampling could work with several different data collection modes, but it is best served with an online-based questionnaire, which could include a cross-sectional administration or a longitudinal diary approach. However, because of the opt-in nature of the Blended Sample, it is not possible to model the probability of response, and thus to account for that source of potential bias in survey estimates.

3. Methods to Maximize Response

We are utilizing an established online panel for survey administration. Survey administration will include an invitation email and up to one reminder email (as needed) in an effort to maximize response rate. The expected response rate for the Ipsos sample is 13%. This is based on the response rate for the 2012 survey concerning the last Presidential election that Ipsos conducted using their ISAY panel and Ampario system.

4. Testing of Procedures

Prior to finalizing the survey instrument, FMG conducted a usability study with 35 adults. Testing occurred in November and December of 2014. These participants tested the survey language (to ensure survey respondents understand the industry/service) as well as tipping (monetary/in-kind) attribute language and can accurately recall their tipping activity. The survey findings indicated that some minimal wording changes were required. The changes did not include any of the following: an increase in the kind or amount of information sought; an increase in coverage; an increase in the timing or frequency of reporting; a change in the sample design or collection method; or a change in the purpose for which the information is collected or required to be maintained.

This testing included a review of possible differences in recall time used for the survey and led to the conclusion that the recall period should be 1 day for all transactions. One of FMG's

findings were that respondents appeared to more heavily lean on the use of estimation heuristics as the recall period was lengthened from 1 to 3 to 5 days (e.g., "It wouldn't really be that difficult for me to recall [longer period of time] since I usually tip about 15%"). More information on the findings from FMG's cognitive and usability testing on this project can be found in the full report highlighting findings and edits made to the survey instrument.[11]

A recall period longer than 1 day could also lead to other forms of recall errors. One such type of error that has been found in research concerning major events or large consumer purchases is telescoping, which occurs when the respondent fails to accurately remember when an irregular event occurs or even "remembers" events that never occurred. This could be very problematic during analysis as it would be impossible to determine the rate at which these services are occurring and on what days. Increasing the number of observations of certain, less frequent services would be ideal, but not at the expense of significantly reduced data quality per observation that could result from these recall errors or reliance on estimation heuristics.

Lengthening the recall period beyond 1 day would also increase respondent burden and could lead to reporting bias. Respondents would have to carefully read instructions for each series of questions to determine when different recall lengths are being asked, and then would have to remember specifically which day the expenditure occurred in order for it to be useful for analysis. Such a significant change in survey instructions would require additional testing to ensure comprehension. Confusion could also result from determining which was the most recent expenditure, and could allow for reporting biases concerning which expenditure they choose to report. This would be particularly problematic for service categories where multiple expenditures could occur at the one establishment (like hotels, beauty salons, or casinos). In such scenarios, respondents might choose to report a service that they can recall easily rather than reporting the one that they paid for most recently.

The survey will be administered electronically; however there are no cookies involved. Survey participants will be provided a link/web address via a secure website. Transmission to/from the secure website for the survey will be encrypted.

Survey respondents will be selected from the subcontractor's panel members and non-panel Internet users. Potential respondents will be sent an email invitation to participate in a survey to understand their preferences for how to get help for tax-related service needs they may encounter. Participants will be provided a link/web address to a secure website with their unique survey URL that corresponds to their survey questions. The subcontractor hosting the panel and survey will maintain a secure survey control system that will document the correspondence and track the status of all sample members by giving each sample member a unique sample ID. The sample ID is used in place of name, address, or other personally identifiable information.

5. Contacts for Statistical Aspects and Data Collection

For questions regarding the study or questionnaire design or statistical methodology, contact:
Brian K. Griepentrog, Ph.D.
Director of Research Studies
Fors Marsh Group LLC

---

[11] *IRS Tipping Report on Cognitive and Usability Testing*, January 2015. Internal report prepared for the Internal Revenue Service by Fors Marsh Group under contract TIRNO-13-Z-00021-0002.

bg@forsmarshgroup.com

<u>Attachments</u>
1) Survey Instrument and example recruitment emails
2) Privacy Impact Assessment