

## SUPPORTING STATEMENT

**U.S. Department of Agriculture  
Economic Research Service  
Rural Establishment Innovation Survey (REIS)  
OMB Control No. 0536-XXXX**

### **Part B. Collection of Information Employing Statistical Methods**

#### *1. Universe and Respondent Selection*

For the Rural Establishment Innovation Survey (REIS), the sample will be selected from the business establishment list maintained by the Bureau of Labor Statistics as part of its Quarterly Census of Employment and Wages (QCEW) program for those state employment security departments granting approval, and from a proprietary business list frame (Dunn and Bradstreet) for states not granting approval. Forty-six states and the District of Columbia have agreed to participate, 5 states have declined.

The sample will exclude business establishments with fewer than 5 employees, establishments that are not privately owned and establishments not included in 'tradable industries' defined as mining, manufacturing, wholesale trade, transportation and warehousing, information, finance and insurance, professional/scientific/technical services, arts, and management of businesses.

Sampling stratification will be based on North American Industry Classification System (NAICS) code, metropolitan/nonmetropolitan location, employment size class and whether or not the state has agreed to release their QCEW list frame through BLS for production. Establishments from the same strata in participating and nonparticipating states will have identical target sampling rates. The strata table below considers the national metro and nonmetro establishment populations by industry and is provided for illustrative purposes. Final sampling rates will be determined based on the most recent BLS Quarterly Census of Employment and Wages.

Establishment populations by strata are provided in the table below. The full study sample will be pre-selected based on an initial sample size of 30,000 for the telephone screening survey. This is the number of businesses that could be contacted and re-contacted multiple times and by multiple ways in a mixed mode survey protocol and stay within the survey budget. The target sampling rates were initially computed by compiling population establishment total across the 9 target industries for nonmetropolitan counties and metropolitan counties:

$$\text{Nonmetropolitan Target Sample Rate} = \frac{0.66667 \times 30,000}{\text{Nonmetropolitan Establishment Total}}$$

$$\text{Metropolitan Target Sample Rate} = \frac{0.33333 \times 30,000}{\text{Metropolitan Establishment Total}}$$

Examination of the establishment population data made it clear that the sample sizes for Management of Businesses (Headquarters) and Performing Arts Companies, Museums, Historical Sites, and Similar Institutions (Arts & Museums) would be insufficient to provide reliable statistics. In addition, the Finance and Insurance (Finance) establishment population is very large, particularly with respect to potentially tradable services in rural areas. Oversampling of Headquarters and Arts & Museums by a factor of 5.78 ensures reliable statistics and is offset by an undersampling of Finance establishments by a factor of 0.3.

Table 1. Population Universe by Strata for Rural Establishment Innovation Survey

Stratum: Industry	Stratum: Geography	Stratum: Estab. Size	Estab. Population <sup>1</sup>	Target Sampling Rate	Anticipated Sample
Mining	Nonmetro	5-19	3676	0.13595	500
Mining	Nonmetro	20-99	1813	0.13595	246
Mining	Nonmetro	100 +	422	0.13595	57
Mining	Metro	5-19	4192	0.01105	46
Mining	Metro	20-99	2222	0.01105	25
Mining	Metro	100 +	496	0.01105	5
Manufacturing	Nonmetro	5-19	18286	0.13595	2486
Manufacturing	Nonmetro	20-99	12139	0.13595	1650
Manufacturing	Nonmetro	100 +	5670	0.13595	771
Manufacturing	Metro	5-19	83603	0.01105	924
Manufacturing	Metro	20-99	54869	0.01105	606
Manufacturing	Metro	100 +	18844	0.01105	208
Wholesale Trade	Nonmetro	5-19	18409	0.13595	2503
Wholesale Trade	Nonmetro	20-99	5475	0.13595	744
Wholesale Trade	Nonmetro	100 +	464	0.13595	63
Wholesale Trade	Metro	5-19	120525	0.01105	1332
Wholesale Trade	Metro	20-99	46431	0.01105	513
Wholesale Trade	Metro	100 +	7524	0.01105	83
Transportation	Nonmetro	5-19	11667	0.13595	1586
Transportation	Nonmetro	20-99	4310	0.13595	586
Transportation	Nonmetro	100 +	675	0.13595	92
Transportation	Metro	5-19	44198	0.01105	488
Transportation	Metro	20-99	23320	0.01105	258
Transportation	Metro	100 +	6393	0.01105	71
Information	Nonmetro	5-19	7179	0.13595	976
Information	Nonmetro	20-99	2410	0.13595	328
Information	Nonmetro	100 +	230	0.13595	31
Information	Metro	5-19	32615	0.01105	360
Information	Metro	20-99	18550	0.01105	205
Information	Metro	100 +	5714	0.01105	63
Finance	Nonmetro	5-19	24309	0.04079	991
Finance	Nonmetro	20-99	3785	0.04079	154
Finance	Nonmetro	100 +	244	0.04079	10
Finance	Metro	5-19	130275	0.00332	432
Finance	Metro	20-99	30697	0.00332	102

Table 1. Population Universe by Strata (Cont.)

Stratum:	Stratum:	Stratum:	Estab.	Target	Anticipated
----------	----------	----------	--------	--------	-------------

<sup>1</sup> From County Business Patterns (2009).

Industry	Geography	Estab. Size	Population	Sampling Rate	Sample
Finance	Metro	100 +	7648	0.00332	25
Prof/Sci/Tech Serv.	Nonmetro	5-19	19742	0.13595	2684
Prof/Sci/Tech Serv.	Nonmetro	20-99	2624	0.13595	357
Prof/Sci/Tech Serv.	Nonmetro	100 +	255	0.13595	35
Prof/Sci/Tech Serv.	Metro	5-19	171249	0.01105	1893
Prof/Sci/Tech Serv.	Metro	20-99	51105	0.01105	565
Prof/Sci/Tech Serv.	Metro	100 +	9479	0.01105	105
Headquarters	Nonmetro	5-19	1415	0.78581	1112
Headquarters	Nonmetro	20-99	879	0.78581	691
Headquarters	Nonmetro	100 +	248	0.78581	195
Headquarters	Metro	5-19	13521	0.06388	864
Headquarters	Metro	20-99	10792	0.06388	689
Headquarters	Metro	100 +	5115	0.06388	327
Arts & Museums	Nonmetro	5-19	637	0.78581	501
Arts & Museums	Nonmetro	20-99	127	0.78581	100
Arts & Museums	Nonmetro	100 +	20	0.78581	16
Arts & Museums	Metro	5-19	3379	0.06388	216
Arts & Museums	Metro	20-99	1523	0.06388	97
Arts & Museums	Metro	100 +	496	0.06388	32
Totals			1051885	0.02852012	30000

The relatively small cell sizes for some of the “Large” establishment strata raises the possibility of oversampling the Large establishment strata. However, the main interest in including all establishment size classes is to ensure the ability to make inferences on the tradable sector nationally. The main focus of the study is on innovation in small and medium sized establishments. The Large establishment strata sample size should be sufficient for producing reliable statistics when aggregated across industries.

The sample for the pilot study will be comprised of roughly 2,600 respondents from the previous 1996 ERS Rural Manufacturing Survey and 1,400 respondents drawn from the BLS sample frame and from Dunn and Bradstreet, a commercial sample vendor.

For the states that do not approve BLS providing sample, the Dun and Bradstreet sample (DB) is expected to be less current and of lesser quality compared to the BLS sample. For the pre-screening effort, it is anticipated that the DB sample will be updated less frequently and with less authority compared to the BLS provided sample. The screening survey is very short and since anyone answering the phone can provide this information there will be lesser limitations to responding with contact information (Attachment J). The screening survey design and full study questions for the REIS are very similar to the 1996 Rural Manufacturing Survey that was administered and validated by SESRC.

## 2. Procedures for Collecting Information

For participating establishments, REIS will be a one-time survey collection and will occur mainly in 2013. This is a voluntary government sponsored survey and will be conducted by an academic survey organization based at a Land Grant University. All establishments will be contacted through an initial telephone screening effort to determine if businesses are eligible (currently “in-business” and having 5 or more employees) for the study. During this contact, information will be obtained to identify a knowledgeable and appropriate respondent for the business and to collect all of this individual’s contact information (Attachment J). Screening calls improve the quality of the survey sample and improves likelihood of reaching a target respondent and gaining more accurate survey responses for the business.

A letter of introduction signed by USDA Undersecretary Catherine Woteki of the Research, Education, and Economics mission area will be sent to all potential respondents prior to the survey participation contacts (Attachment D). The purpose of this advance letter is to notify businesses about the study and why we need their participation.

For the REIS, respondents will be asked to complete questionnaires in at least one of three possible survey modes (telephone, web, or mail, Attachments A, B and C). All survey instruments across modes will be carefully aligned to provide the same information and explanations of the survey. The web version of the survey is to be located on the SESRC WSU website with a specific URL. Each question screen will carry a banner with the survey title “National Survey of Business Competitiveness” and USDA ERS sponsor. The telephone survey introduction will be used by interviewers to explain the purpose and the sponsorship of the study. The mail surveys will use a cover letter to provide this information. All modes of contacting respondents will provide information on how to contact SESRC or ERS if they have questions or need clarifications about the study.

The survey methodology literature over the last decade has addressed the use of incentives as a means to improve response rates in household and person based surveys. However, there remain gaps in this literature with respect to detailed description of the establishment survey response process, the effectiveness of survey mode sequencing and how incentives interact within these processes to impact establishment survey respondents. The most important aspects of survey implementation shown to increase response rates in business surveys respectively are: 1) “Response Required By Law” message; 2) multiple contacts; and 3) cash incentives. A pilot study will use an experimental design to test various interventions on survey response that can be used to improve response. The experimental testing framework used in this study (see Table 2 and Table 3) is important because it will offer insights into how non-mandatory (voluntary) survey response is impacted by process components and strategies. There are a number of objectives to be tested: 1) alternative survey mode sequencing (telephone sequence first versus mail sequence first); 2) the effectiveness of each mode; 3) the combination of postal class and packaging (first class postage versus two day priority mail class and mail envelope packaging cardboard mailer versus brown paper envelope);

and 4) early stage, later stage, and repetitive application of a small token \$2 cash incentive with mail questionnaire; and 5) the timing of offering the web mode as an alternative response option for survey completion. Depending on the experimental group assignment and intervention, the business respondent will be contacted by telephone and/or by mail and will be offered one of three ways (telephone, mail, or web) to complete the survey.

These results collected within a voluntary survey environment reflect a more generalizable survey structure than those realized under mandatory government collections. We hope to capitalize on respondents' awareness of web surveys and the offering of a choice as a means to accommodate completing the survey in a mode of their preference to determine if this is an important element of survey strategy. There is research in the household respondent survey literature that suggests offering more than one survey mode at a time can decrease survey response rather than enhance response (Millar and Dillman, 2011). This is an aspect that has not been tested in the establishment survey arena. This study specifically incorporates the idea of offering a web link at specific junctures in the contact process and then following this with email augmentation to those respondents with an email address that was gained during the telephone prescreening contact of the business.

The mode sequence selected for the full study will be contingent on findings from the pilot and the factors surrounding this decision will be fully elaborated in the pilot study assessment report submitted to OMB. The three general outcomes anticipated are statistically significant higher response rate of one mode sequence over all others, statistically significant higher response rates for two or more mode sequences over remaining mode sequences without identification of a clear dominant mode sequence, or failure to discern statistically significant differences in response rates across all mode sequences. The mode sequence selected in the first and last case would be the one with highest response rate or lowest cost, respectively. In the middle case mode sequences with statistically significant lower response rates would be abandoned and the survey would be administered by allocating an equal share of potential respondents to the remaining mode sequences. The likelihood that the pilot study will identify substantive differences between mode sequences if they exist is high: the power of detecting a difference in response rates of 0.05 between two mode sequences with a sample size of 1600 is 0.953 at the 0.01 level of significance.

For the web survey version, the website for the survey will be secure and respondents can only access the website by entering their specific project assigned identification code. It is anticipated most respondents will be able to complete the questionnaire in one session. However, business respondents will be allowed multiple reentries to the survey website if needed to complete the questionnaire in multiple sessions.

Upon receipt of completed questionnaires, SESRC will download, enter, compile, and aggregate survey responses from each survey (mode version and interventions) and analyze all survey responses. Respondents will all be addressed with the same survey questions about their business environment, activities and revenues thus providing

uniform data across survey venues.

All contact materials and survey questionnaires have benefited from expert consultation (internal and external) and peer review by stakeholder groups. Cognitive interviews to test the survey questionnaire were conducted in September 2013 (Attachment F). The letters and reminders were developed in collaboration with internal and external survey methodologists.

## **DATA EDITING PROCEDURES**

### **Telephone screening and telephone interviewing**

Survey data for all REIS samples – landline, listed, and cell– will be collected using the same computer-assisted telephone interview (CATI) system for both screening telephone survey phase and extended full interviews collected over the telephone. While the screening interview may vary somewhat by sample, the same editing procedures will be followed for all REIS cases. In a CATI environment, the data collection and interview process is controlled using a series of computer programs to ensure consistency and quality. At SESRC WSU, the commercial CATI software used is Voxco and this software has been used more than 15 years. SESRC has more than 25 years experience with CATI software. For the telephone survey administration, the CATI system programming determines which questions are asked based on business characteristics, composition, respondent characteristics, or preceding answers, and the order in which the questions are presented to interviewers. The system also presents the response options that are available for recording answers. CATI range and logic edits do much to help ensure the integrity of the data during the collection process by telephone. This editing at the time of the interview greatly reduces the need for post-interview editing and allows most questionable entries to be reviewed in real time with the respondent as part of the collection process. Although the CATI system virtually eliminates out-of-range responses and many other anomalies, some consistency and edit issues may arise. For example, interviewers may note concerns or problems that must be handled by data analysts or preparation staff after the interview is complete. Updating activities require that both manual and machine editing procedures be developed to correct interviewer, respondent, and CATI program errors and to check that updates made by data management staff were input correctly. Because data editing may result in changes to the survey data, specific quality control procedures will be implemented. REIS survey data will be carefully examined and edited before delivering final data files to ERS USDA.

Additional data quality assurance occurs through survey supervision of interviewer performance. Quality checking is implemented by survey monitors and survey supervisors that listen and visually screen check coding of live interview answering between interviewers and respondents while they are being conducted. Any problems in question delivery, interview performance, or entry will be noted and the interviewer will be notified of performance problems. SESRC has a performance management scoring system for interviewers. This process includes meeting with each interviewer to discuss performance, review outcomes, and plan for improvement. Interviewers are routinely

monitored with a goal of once a week during calling, within the first few days of calling on any given project, and to meet contractual agreements. Routinely, as part of contractual agreements, SESRC monitors between 5 % and 10% of all interviews for quality. If needed an interviewer will be retrained and systematically monitored for improvement. If an interviewer is not capable of meeting performance objectives they are terminated from calling. If an error in the data recorded by an interviewer is detected a data correction will be made to the case. If the errors detected are severe then all cases by a given problematic interviewer will be reviewed for completeness and accuracy. If any cases are suspect, then cases will be recalled and/or particular answers verified with the business.

One critical step during the data collection process for telephone interviews includes a process whereby at the completion of an interview, each interviewer answers a set of questions about the interview. If the interviewer detects concerns with quality such as compromised respondent ability, extreme distractions, or other issues these are noted at this time. Survey supervisors routinely review these results to detect poor or suspect interviews. Quality control procedures associated with data corrections may also involve limiting the number of staff who make updates, using the CATI specifications to resolve issues in complex questionnaire sections, carefully checking updates, and performing computer runs to identify inconsistencies or illogical patterns in the data associated with the current questionnaire.

The data editing procedures for REIS will consist of four main tasks: (1) managing and resolving problem cases (error checking), (2) reading and using interviewer comments to make data updates, (3) coding questions with open ended text strings (i.e., “other, specify” responses), (4) verifying data editing updates, (5) survey supervisor review of interviewer response outcomes on interviews. The final step will be to convert the edited data from the CATI system to the SAS data delivery files.

### **Mail returns, review, hand coding, and hand data entry**

For completed mail questionnaires, the data entry process consists of three main stages: 1) initial data entry by one clerical staff, 2) verification (second pass data entry) performed by a different clerical staff, and 3) the final validation step is to account for all questionnaires by ID number and ensure all observations have been entered, verified and to correct any errors that may have occurred during this process. The data entry program consists of a computerized online system that prompts clerical personnel for valid responses to every question in the survey. The data entry program has the same features and operational features as the CATI questionnaire software for range checks and question branching/skipping logic.

Prior to the initial data entry, data editing and data cleaning will occur once a large number of returned completed questionnaire are received and a coding manual has been developed. During this initial phase several hundred paper questionnaires and question answers will be reviewed for: 1) respondents’ adherence to following question branching and skip instruction patterns, 2) marks and comments written in the margins or on questions; 3) completeness and open-ended numeric answers with anomalies; 4) straight lining on question banks; 5) selective checking in question banks; and any other types of



errors indicating the need for data cleaning and data edits. Once a large number of paper questionnaires have been reviewed a coding manual will be drafted and reviewed with principal investigators and researchers at ERS prior to hand coding and data entry. Cleaning decisions will be documented in the coding manual and instructions for specific questions and problems developed for coders. Coding will be performed by a limited numbers of coder staff to ensure accuracy and consistency of coding. A data manager/analyst will review coding. Once questionnaires are coded data entry will be performed by data entry staff.

### **Web surveys**

In the web survey environment, all questions allow voluntary responses and there is no insistence built into the web questionnaire program that requires an answer to maintain progression through the survey by the respondent. This also meets the best practices for human subject's research. Allowing the respondent to "not answer a question" also prevents abandonment of the interview as it reduces respondent's frustration if they are unwilling to answer a given question. In order to reduce instances of questions being skipped over without answering special screen prompts will be programmed and shown that will prompt for an answer. The goal of this functionality is to persuade the respondent to answer the question by describing the importance of the response or the purpose of the question. The types of questions that most often experience item non-response are open ended numeric questions. These questions will be carefully reviewed and pretested to determine if they require specific instructions for inclusions or exclusions. If it is found during the early stages of the study that respondents are skipping particular questions, these particular questions will be reviewed for sensitivity, wording and or comprehension issues. If needed the question will be changed or information added such as an instruction, definition, or a screen prompt.

### **Estimation procedures**

The analytical approach for addressing the study's central research questions are discussed below:

1. What percentage of rural establishments in tradable industries introduced product, process or practice innovations in the previous 3 years?
2. What percentage of self-reported innovative establishments also demonstrates behaviors consistent with substantive innovation?
3. How do self-reported and ostensibly substantive innovation rates differ by urban/rural location, industry and establishment age?
4. What establishment and community characteristics are associated with self-reported and ostensibly substantive innovation?
5. Do ostensibly substantive innovators demonstrate faster rates of employment growth or higher survival rates than claimed innovators and non-innovators?

Questions 1-3 will be addressed using descriptive analysis. Questions 4-5 will be addressed using multivariate regression techniques. In addition, questions 2-5 will require a method for classifying innovative establishments as either claimed innovators or substantive innovators.

To address the first question, the percentage of rural respondents that report product, process or practice innovations will incorporate information from the complex sample design to the entire sample to produce valid estimates of mean and variance and pseudo-maximum likelihood methods for generating population weighted frequency tables. Within the rural stratum, comparison of innovation rates across settlement types ranging from micropolitan counties to entirely rural counties will use domain analysis to take into account the randomness of the sample size across settlement types. As the first quantitative assessment of rural innovation in the U.S., valid variance estimation will be critical in describing the phenomenon across the rural continuum.

However, past efforts examining measures of self-reported innovation in the European Union have identified a problem of over-reporting (North and Smallbone 2000). Lacking the resources to qualitatively assess the innovativeness of each respondent, the analysis will utilize auxiliary information on various establishment characteristics believed to be strongly associated with substantive innovation. For example, a question designed to correct for social desirability bias will ask about failed innovations at the establishment. Comparing the percentage of claimed innovators that acknowledge failed innovations to the percentage of claimed innovators that do not acknowledge failed innovations will provide one measure of possible over-reporting. Other characteristics, such as safeguards for protecting intellectual property or practices that facilitate data-driven decision-making, may also differentiate substantive innovators from claimed innovators. Variation in these observed variables may reflect variation in an unobserved factor related to substantive innovation.

Mixture models such as latent class models are well-suited to the problem of describing and analyzing observations hypothesized to come from different unobserved subgroups in the population. The two conceptual classes of most interest are substantive innovators and nominal innovators with non-innovators identified as respondents opting out of the innovation questions. However, the data could support four subgroups in the population with a subgroup of advanced non-innovators being identified; i.e., respondents that did not introduce new or significantly improved products but did utilize data-driven decision-making tools or possessed intellectual property worth protecting. Recent research examining the use of latent class models with complex survey design data (Patterson, et al. 2002; Vermunt 2007; Wedel, et al. 1998) has made it possible to apply these tools when the assumption of simple random sampling is violated.

The validity of the latent class structure will be assessed in the short-run by comparing the industry distribution of substantive innovators with known innovation intensive industries. If ostensible substantive innovators are much more likely to be in innovation intensive industries, then this would provide prima facie evidence of the validity of the class structure. In the long-run, linking REIS to the Business Employment Dynamics

data at BLS (see below) will provide longitudinal performance data to compare substantive with nominal innovators that would provide outcome based evidence of the validity of the class structure.

Questions 2 and 3 will apply the relevant innovator classification to all respondents and then estimate mean and variance of percentages as was done for the self-reported innovation variable in Question 1. Domain analysis will be used when estimating parameters across groups such as settlement type, industry or establishment age.

Question 4 will be addressed using a binary response model to investigate the relationship between innovative activity and establishment and community characteristics. Nonlinear logit or probit models able to incorporate complex survey design information are available in statistical software packages allowing unbiased estimation of parameter variance. Domain analysis will allow investigating similarities or differences with respect to innovative activity across settlement types or industry groups providing critical information for designing rural innovation policy.

The analysis will also provide an assessment of the value of the ostensibly substantive innovation classification. It is anticipated that the explanatory power of the substantive innovation model will be significantly higher than the self-reported innovation model since the latter is thought to include establishments over-reporting innovative activity due to social desirability bias. Alternatively, if the substantive innovation model does not demonstrate better explanatory power then it is less likely that the observed characteristics thought to be related to substantive innovation are correlated with the hypothesized unobserved factor.

Questions 1-4 will be addressed as soon as cleaned data from the REIS becomes available. Addressing Question 5 will not be possible until several years later when a sufficient amount of quarterly employment data is available to support survival analysis. It is anticipated that the REIS will be linked with the Business Employment Dynamics data at the Bureau of Labor Statistics that will allow examining the medium and long-term effects of innovative activity on establishment survival and employment growth.

To examine employment growth we will use a two-stage model that incorporates information from an establishment exit model to correct for the nonrandom selection of surviving establishments. This model has been widely adopted in manufacturing studies (Doms et al., 1995; Jarmin, 1999, Acs 2002). The two stages are specified as:

$$(1) \quad Exit_i = \text{Pr}(X_i \beta_1) + \mu_{1i}$$

$$(2) \quad Growth_i = X_i \beta_2 + \sigma_{12} \varphi(X_i \beta_1) / (1 - \Phi(X_i \beta_1)) + \mu_{2i}$$

where  $\beta_1$  is the parameter vector from the exit equation,  $\beta_2$  is the parameter vector from the growth equation,  $\sigma_{12}$  is the covariance between the disturbance terms of the two equations and  $\varphi(X_i \beta_1) / (1 - \Phi(X_i \beta_1))$  is the inverse Mills ratio—derived from the first

stage regression and used as an instrument to control for selection bias in the second stage. We estimate equation (1) using standard limited dependent variable techniques. We identify equation (2) via the nonlinearity of the Mills ratio as do Evans, 1987, and Doms et al., 1995.

Establishment survival will be assessed using a proportional hazard specification that is widely-used and designed to account for the censored nature of the data. Our dependent variable, whether an establishment is continuing or has exited, is reported quarterly for each establishment, is modeled as:

$$(3) \quad Y_{it} = \alpha + \beta X_i + \mu_i,$$

where  $i = 1, \dots, N$  establishments,  $t = 1, \dots, T$  quarters during the specified period and  $Y_{it}$  is 0, 1.

The quarterly dependent variables are regarded as a panel of binary variables; each quarter, for each establishment, there is an indicator variable for whether or not the establishment has any employees. Each establishment is viewed as contributing several observations to a larger logit likelihood function, the product of each of the (3) logit models:

$$(4) \quad \prod_{i=1, t=1}^{i=N, t=T} Y_{it}$$

Treating the data as a panel data set facilitates estimating flexible hazard functions because the complicated likelihood maximization problem is replaced with a familiar logit estimation problem (equation 4), which can be estimated with standard software.

Integrating complex survey design information into the analysis required to address Question 5 is now possible using the `svyset` functionality in Stata 11. Both 2-stage selection models and proportional hazard models can now be estimated using the `svy` command that incorporates survey design information and allows performing domain analysis on selected subpopulations to produce valid variance estimates.

### **Degree of Accuracy Needed**

Comparing innovation rates between urban and rural establishments is a primary focus of the study. The most challenging aspect of this question with respect to sample size is comparing conventional measures of innovative or inventive activity such as patent application rates as these tend to be rare in both urban and rural environments. Unfortunately we have not been able to locate previous studies that have examined patent application rates at the establishment level. However, it is possible to combine information from different sources to arrive at a reasonable estimate of differences in patent application rates we would expect to observe. We would want a sample large

enough to detect a significance difference between these expected application rates.

We combine survey results from the 1996 Rural Manufacturing Survey with the 2008 BRDIS results to arrive at an expected patent application rate for manufacturing establishments. We then use European data on differences in patent application rates between manufacturing and services to estimate patent application rates for our entire sample. We incorporate information from both differences in rural and urban patent application rates and differences in the mix of manufacturing and services to arrive at expected patent application rates for urban and rural areas among all tradable sectors.

The results from the 2008 BRDIS suggest that 1 in 5 firms with R&D units applied for at least one patent. Findings from the Rural Manufacturing Survey demonstrate that 30% of urban establishments and 22% of rural establishments had an R&D unit. For manufacturing we would expect that 6% of urban establishments applied for a patent compared with 4.4% of rural establishments. Given a likely rural manufacturing sample of 4,907 and urban manufacturing sample of 1,738, and assuming a 60% response rate, the power of the test for two proportions fails to make the threshold of 0.8 of a powerful test at 0.655. This example is instructive because it is the one industry for which we have the best information and also where the events are anticipated to be less rare. However, the low power is not a problem for the study objectives of comparing rural and urban innovation rates for the tradable sector.

The POWER Procedure  
Pearson Chi-square Test for Two Proportions

**Fixed Scenario Elements**

<b>Distribution</b>	Asymptotic normal
<b>Method</b>	Normal approximation
<b>Number of Sides</b>	1
<b>Group 1 Proportion</b>	0.06
<b>Group 2 Proportion</b>	0.044
<b>Group 1 Sample Size</b>	1043
<b>Group 2 Sample Size</b>	2944
<b>Null Proportion Difference</b>	0
<b>Alpha</b>	0.05

**Computed Power**

**Power**

0.655

To apply the power analysis to the entire sample we use patent application data from Europe to arrive at a reasonable ratio of services to manufacturing patent application

rates. We then apply this ratio to our estimates of rural and urban US manufacturing patent application rates to derive the services patent application rates. The assumption is that the ratio between manufacturing and services application rates is the same in both entities without requiring the more restrictive assumption that patent application rates in Europe and the US are equal.

The services patent application rate in Europe is 41.5% the manufacturing patent application rate. Thus, in the US we estimate that the rural services patent application rate would be 0.01826 (or 41.5% of 4.4%) and the urban services patent application rate would be 0.025 (or 41.5% of 6%). The fact that manufacturing makes up a larger share of the tradable sector in rural areas reduces the expected difference between rural and urban patent application rates overall. For the urban tradable sector overall the patent application rate is expected to be 0.03183 and 0.024575 for the rural tradable sector. Assuming a 60% response rate an initial sample size of 30,000 will produce a test of adequate power of 0.872.

The POWER Procedure  
Pearson Chi-square Test for Two Proportions

**Fixed Scenario Elements**

<b>Distribution</b>	Asymptotic normal
<b>Method</b>	Normal approximation
<b>Number of Sides</b>	1
<b>Group 1 Proportion</b>	0.03183
<b>Group 2 Proportion</b>	0.024575
<b>Group 1 Sample Size</b>	6000
<b>Group 2 Sample Size</b>	12000
<b>Null Proportion Difference</b>	0
<b>Alpha</b>	0.05

**Computed Power**

**Power**

0.872

By positing the magnitude of innovation events we expect to be rare in the sample we are able to demonstrate that an initial sample size of 30,000 will be sufficient for detecting expected difference between rural and urban establishments.

*3. Methods to Maximize Response*

Efforts to maximize response and still remain within the survey budget will use token cash incentives (\$2), higher class postage and distinctive mailers in the mail modes of contact. For all modes and mode sequences this study will utilize multiple contacts as a best practice to reach the respondent and achieve response. The use of mixed mode design, with a telephone sequence with 20 call attempts and the use of a mail sequence are also known strategies to increase survey response. In addition, in the mailing portion of the study an additional special contact will be mailed to sampled businesses that refused during telephone contact or by mail. This letter will be specially designed to appeal and persuade based on known psychological messaging to emphasize the importance of the survey request.

#### *4. Tests of Procedures or Measures*

After the initial design phase, the telephone version of the questionnaire was tested by internal SESRC and ERS expert review, mock interviews over the telephone between SESRC and ERS USDA staff. The CATI telephone instrument was tested with one ineligible known innovative business from the local WA State population to assess: questionnaire length, usability, workability, question understanding, and to behavior code respondent clarifications.

After the initial testing, mail and telephone versions of the survey were tested using cognitive interviewing protocols with 6 establishments (see Attachment F for the detailed report). A special focus of the cognitive interviewing was auxiliary questions that will be used to differentiate substantive from nominal innovators. All of the auxiliary questions were easily understood and answered by the six respondents. The cognitive interviewing was also invaluable for assessing how industries outside of manufacturing would respond to questions and resulted in significant modifications to the survey instruments. Finally, the cognitive interviewing helped identify opportunities for decreasing respondent burden (e.g., allowing firms with no debt to avoid questions on borrowing).

The questionnaires will undergo comprehensive testing and usability testing by internal SESRC experts, supervisors, and interviewers during pretesting with actual respondents in a pilot phase of this study after OMB clearance. Usability pretesting during the pilot will include monitoring interviews to observing participants' probes and clarification behaviors, noting difficulties and comments, and conducting post-testing interviews with interviewers to gain qualitative feedback about potential confusions. In addition, quantitative measures will also be gathered, including time to complete the survey, evaluating paradata and navigation patterns from the web questionnaire.

The pilot study will also be used to assess item nonresponse along with problems of very limited response variation. A focus of this analysis will be to identify systematic nonresponse within particular industry or establishment size strata. With the proposed sample size of 4000 only two of the 54 strata have empty cells and four strata have an initial sample of two. This coverage should be sufficient to identify significant nonresponse problems prior to the full study.

This study includes a pilot study that has experimental components that are designed to evaluate impacts on less cooperative respondents that require more contacting to gain cooperation. The study tests the impact of survey mode sequencing (mail, telephone, and web) and interactions with other interventions as shown in Tables 2 and Table 3. The pilot sample frame is randomly assigned to experimental groups 1 to 5. Each group varies on sequence and timing of treatments. Two-fifths of the sample is assigned to first receive the telephone sequence of survey contacts which is then followed by questionnaire mailings for main data collection. Three-fifths of the sample frame will be contacted first by mailings with questionnaires followed by telephone follow-ups for survey completion. Next the groups vary on when (which specific day and mailing) a web link is offered to do the survey over the internet. For those respondents with an email address an email contact will follow that is designed to augment the postal letter contact as it offers a web link that can be clicked on to go directly to the survey. Also the interventions of \$2 cash incentives and the use of higher class two day priority mail compared to first class postage and the number of applications will be used at varying phases in the multi-contact sequence. The overall goal is to evaluate whether any of these interventions comparatively improve response propensity and/or bring in more of the “hard to reach” establishment respondents.

Table 2 shows the overall tests for each group and inclusion of specific treatments. Table 3 shows each group and the specific details of implementation by days across data collection. Early responders from the screening portion of the survey will not be allowed in the pilot so that all respondents experience the experimental treatments. Early responders will be encouraged in the full study.



**Figure 1: Overview of the Rural Establishment Innovation Survey Pilot Study Implementation Process**

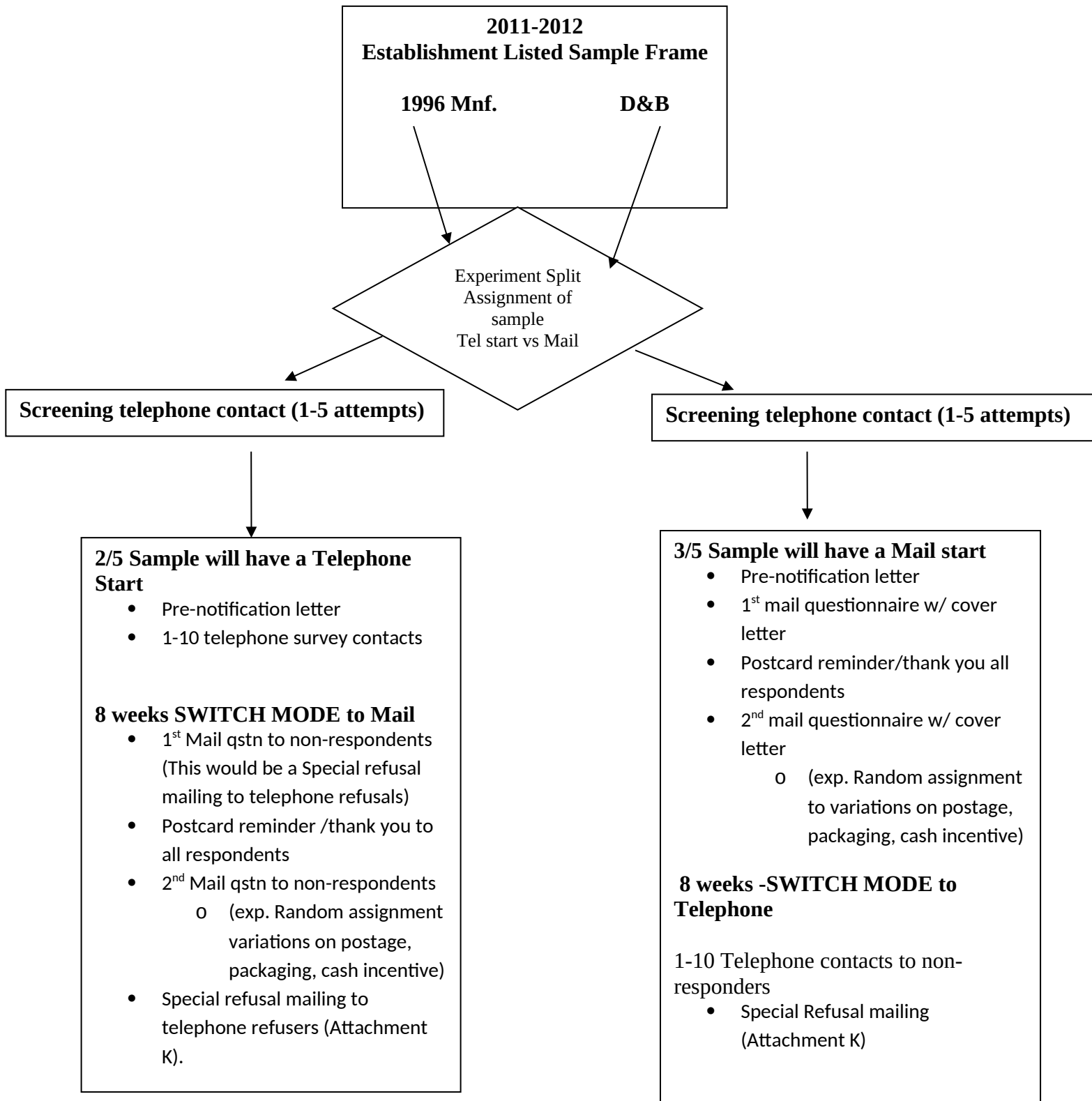


Table 2. REIS Pilot Study Experimental Design and Stimuli

	Tel Pre-screen	Pren. letter	Pren. letter has web link	Mode Sequence test	Web link timing	Web link day	When web link is offered	Web link times	Email Augm	\$2 Incentive yes/no	Incentive times	Incentive day(s)	Incentive timing	Priority mail	Priority mail timing
G1	Yes	Yes	No	Mail first	Early	day 7	1st mail qstn and after	3	day 14	Yes	2x	day 7 & day 35	early	1x	late day 35 2nd qstn
G2	Yes	Yes	No	Tele first	Late	day 42	1st mail qstn and after	3	day 49	Yes	2x	day 42 & day 56	late	1x	late day 56
G3	Yes	Yes	Yes	Mail first	Very early	Day 1	Prenotice letter and all mailings after	4	day 7	Yes	2x	day 1 & day 28	very early	1x	day 28 1st qstn
G4	Yes	Yes	No	Mail first	Early	day 7	1stn mail qstn and after	4	day 7	Yes	2x	day 7 & day 35	early	2x	day 7 1st qstn and day 35 2nd qst
G5	Yes	Yes	No	Tele first	Late	Day 42	1st qst mail qstn and after	4	day 49	No	None	None	None	None	None

Table 3. REIS Pilot Study Experimental design specific interventions and details of implementation procedures across data collection.

Group	Exp. Design	Sample size	Tel Prescreen	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8	Phase 9	Phase 10	Phase 11
			4 weeks	Day 1	Day 7	Day 14	Day 21	Day 28	Day 35	Day 42	Day 49	Day 56	Day 63	Day 70-77
1	Mail First	800	tel. prescrn	Pre-notice <sup>1</sup> letter <b>NO Web link</b>	1st Qstn <b>Web link \$2 First class</b>	Email Augm		Postcard thank you reminder	2nd Qstn <b>Weblink \$2 Priority Mail</b>	Tel 1-2	Tel 3-4	Tel 5-6	Tel 7-8	Tel 9-10
2	Telephone First	800	tel prescrn	Pre-notice letter <b>NO Web link</b>	Tel 1-2	Tel 3-4	Tel 5-6	Tel 7-8	Tel 9-10	1st Qstn <b>Web link \$2 First class</b>	Email Augm	Postcard thank you reminder	2nd Qstn <b>Weblink \$2 Priority Mail</b>	Refusal mailing
3	Early Web Push Mail first	800	tel. prescrn	Pre-notice letter <b>Web link \$2</b>	Email Augm		paper follow-up reminder letter	1st Qstn <b>Weblink \$2 Priority mail</b>	Postcard reminder	Tel 1-2	Tel 3-4	Tel 5-6	Tel 7-8	Tel 9-10
4	All stimulus Mail 1st Qstn	800	tel. prescrn	Pre-notice letter <b>NO Web link</b>	1st Qstn <b>Weblink \$2 Priority mail</b>	Email Augm		Postcard reminder	2nd Qstn <b>Weblink \$2 Priority Mail</b>	Tel 1-2	Tel 3-4	Tel 5-6	Tel 7-8	Tel 9-10
5	Control Tel 1 <sup>st</sup> No cash First class only	800	tel prescrn	Pre-notice letter <b>NO Web link</b>	Tel 1-2	Tel 3-4	Tel 5-6	Tel 7-8	Tel 9-10	1st Qstn NO Web No Cash First class	Eamil Augm	Postcard Thank you reminder	2nd Qstn NO Weblink NoCash First Class	Refusal mailing

<sup>1</sup> All prenotice contacts will have an enclosure from the USDA undersecretary Catherine Woteki of the Research, Education, and Economics Mission area.

5. *Contact(s) for Statistical Aspects and Data Collection*

For questions on statistical methods described above, please contact  
Timothy R. Wojan  
Regional Economist  
Farm and Rural Business Branch  
Economic Research Service, USDA  
355 E Street SW  
Washington, DC 20024  
Tel. 202-694-5419  
twojan@ers.usda.gov

For questions on the data collection described above, please contact:  
Danna L. Moore  
Social and Economic Sciences Research Center  
Washington State University  
Pullman WA 99164-4014  
Tel. 509-335-1117  
moored@wsu.edu

**Attachments**

Attachment A	Draft Rural Establishment Innovation Survey (sent out as National Survey of Business Competitiveness)
Attachment B	Final CATI Script
Attachment C	Screen shots of the Rural Establishment Innovation Survey Internet Application
Attachment D	Draft Rural Establishment Innovation Survey Letters
Attachment F	Cognitive interview Report 12-051: National Survey of Business Competitiveness
Attachment J	Pre-screening Telephone Script
Attachment K	Mail Short Form for Telephone Refusals
<b>Attachment Not Referenced in Supporting Statement</b>	

Attachment G	ERS Response to NASS Comments
--------------	-------------------------------

## References

- Acs, Z. 2002. *Innovations and the growth of cities*. Northampton, MA: Edward Elgar.
- Doms, M., Dunne, T. and Roberts, M.J.. 1995. "The role of technology use in the survival and growth of manufacturing plants," *International Journal of Industrial Organization*. 13: 523-542.
- Evans, D.S. 1987. "The relationship between firm growth, size, and age: estimates for 100 manufacturing industries," *The Journal of Industrial Economics*. 35(4):567-581.
- Hsieh, F.Y. 1989. "Sample size tables for logistic regression," *Statistics in Medicine* 8:795-802.
- Jarmin, R.S. 1999. "Government technical assistance programs and plant survival: the role of plant ownership type," CES Discussion Paper 99-2 February.
- Millar, M.M. and Dillman, D.A. 2011. "Improving response rates to web and mixed-mode surveys," *Public Opinion Quarterly* 75(2):249-269.
- North, D. and Smallbone, D. 2000. "The innovativeness and growth of rural SMEs during the 1990s," *Regional Studies* 34(2):145-157.
- Patterson, B., Dayton, C.M., and Graubard, B.I. 2002. "Latent Class Analysis of Complex Survey Data: Application to Dietary Data," *Journal of the American Statistical Association* 97(459): 721-741.
- Vermunt, J.K. 2007. "Latent Class Analysis with Sampling Weights: A Maximum Likelihood Approach," *Sociological Methods and Research* 36(1):87-111.
- Wedel, M., ter Hofstede, F. and Steenkamp, J.-B.E.M. 1998. "Mixture Model Analysis of Complex Samples," *Journal of Classification* 15(5):225-244.