

ECONOMIC RESEARCH SERVICE OMB CLEARANCE PACKAGE

**SECTION B. COLLECTION OF INFORMATION EMPLOYING
STATISTICAL METHODS**

for

**CLEARANCE TO CONDUCT THE SURVEY ON RURAL COMMUNITY
WEALTH AND HEALTH CARE PROVISION
FROM FY2012 THROUGH FY2014**

Prepared by

**Farm and Rural Household Well-Being Branch
Resource and Rural Economics Division
Economic Research Service
U.S. Department of Agriculture**

May 2012

TABLE OF CONTENTS

1. Potential respondent universe and respondent selection methods	1
2. Procedures for the collection of information	5
Sample selection	5
Estimation procedure	7
Degree of accuracy needed and confidence intervals	9
Unusual problems requiring specialized sampling procedures	16
Use of periodic data collection cycles	16
3. Methods to maximize response rates and deal with non-response issues	16
4. Tests of procedures or methods to be undertaken	19
5. Names and telephone numbers of people consulted on statistical methods; people who will collect the data	20

Annex I. Formulas for estimating means and variances

1. Potential respondent universe and respondent selection methods

The potential respondent universe will include health care providers, community leaders and other local stakeholders involved in recruitment and retention of health care providers in rural small towns of the nine states in the study (Arkansas, Louisiana, and Mississippi – Lower Mississippi Delta (LMD) region; Kansas, Oklahoma, and Texas – Southern Great Plains (SGP) region; and Iowa, Minnesota, and Wisconsin – Upper Midwest (UMW) region). These regions and states were selected to include areas in which poverty and lack of access to rural primary care are major problems (especially in the LMD and SGP) and areas where rapid growth in employment in the health care sector has been occurring since 2002 (true in parts of all three regions), as well as areas where growth has been less rapid. The UMW region provides a contrast to the LMD and SGP regions in access to rural primary care and health care utilization, as well as in the average level of poverty and other socioeconomic and demographic characteristics. These regions also include important variations (both across and within them) in health status of the population, presence of different racial and ethnic groups, social capital and other assets, presence of retirement destination communities, programs to improve access to rural health care, and other key factors hypothesized to be related to rural health care provision.

The rural towns in the universe of this study are identified by 2000 Zip Code Tabulation Areas (ZCTAs), which are based on Zip Codes as they were defined at the time of the 2000 Population Census.¹ We use ZCTAs as the primary sampling unit for this survey because this is the lowest geographic level at which data on health care access have been

¹ ZCTAs are defined and data on them are available at the Census Bureau website: <http://www.census.gov/geo/ZCTA/zcta.html>.

compiled by the Dartmouth Health Atlas² (some of which data are used in our sampling approach), and because ZCTAs correspond closely to towns, especially in rural areas.³ We use the Rural Urban Commuting Area (RUCA) codes to classify which towns are rural, excluding towns that are part of metropolitan urban core areas (RUCA codes 1 and 1.1), suburban areas with high dependence on commuting flows (greater than 30% commuting share) to urban areas (RUCA codes 2, 2.1, 4.1, 5.1, 7.1, 8.1, 10.1), and isolated rural areas not part of a metropolitan or micropolitan core and with low commuting dependence on such areas (RUCA codes 10, 10.3, 10.4, 10.5, and 10.6).

We include towns in ZCTAs with a population in 2008 of at least 2,500 and no more than 20,000 in the universe. This focus on small towns is to ensure that the communities are large enough that recruitment and retention of primary health care providers is a realistic prospect (96% of towns with less than 2,500 population and meeting the RUCA code criteria in the study states do not have any primary care physicians) and small enough so that recruiting and retaining primary care providers is likely to be a serious problem (96% of towns with more than 20,000 population meeting the RUCA code criteria in the study states have more than 5 primary care physicians, and 97% have a hospital). We expect that recruitment and retention of health care providers is more likely to be influenced by local community assets and investments in such small rural towns than in smaller rural settlements or larger towns and cities.

The universe of small towns in the three study regions meeting these criteria includes 809 small towns with a total population of 6.1 million in 2008 (about 9% of the U.S. rural

² Information on the Dartmouth Health Atlas and downloadable health sector data are available at <http://www.dartmouthatlas.org/>.

³ Larger towns sometimes include more than one ZCTA. In these cases, we combined these ZCTAs into a single town for the purposes of selecting the sample. Henceforth we refer to the primary sampling unit as “towns”.

population in 2008). 51% of these towns have a hospital and 78% have at least one primary care physician, according to the Dartmouth Health Atlas data. Among the towns that have at least one primary care physician, the mean population per primary care physician is 2,275, but this ranges as high as 16,173 and as low as 233. 36% of the towns in this universe have more than 3,500 people per primary care physician, potentially classifying these as primary care Health Professional Shortage Areas (HPSAs).⁴

Within this universe of small rural towns, the universe of potential respondents includes health care providers, community leaders and other local stakeholders involved in recruiting and retaining health care providers. For the purposes of this study, “health care providers” includes licensed physicians (with either a Doctor of Medicine (MD) or Doctor of Osteopathic Medicine (DO) degree), dentists (with a Doctor of Dental Surgery (DDS) or Doctor of Dental Medicine (DMD) degree), physician assistants or associates (with an appropriate graduate degree such as a Master of Physician Assistant Studies (MPAS)), nurse practitioners (Registered Nurses with an appropriate graduate degree such as a Masters in Nursing), nurse midwives (Registered Nurses with at least a Masters in Nursing and certification as a Certified Nurse Midwife (CNM)), and pharmacists (Pharm.D. degree). We focus on these categories of health care providers to keep the scope of the study manageable within the budget available, yet significantly broader than most studies of rural primary health care providers (usually limited to a focus on physicians). Other categories of health care providers, such as other nurses besides nurse practitioners and nurse midwives, mental health professionals, physical therapists, and others are also important and needed in rural areas, but will be beyond the scope of this study.

⁴ Other criteria besides the population per number of primary care physicians are also considered in designating primary care HPSAs: <http://bhpr.hrsa.gov/shortage/hpsas/designationcriteria/primarycarehpsacriteria.html>.

We do not have the data available to list or characterize the entire population of these health care providers within the universe of small rural towns that the study is focused on. Once the sample of towns is selected (more on this in the next section), a list of health care providers working in each of these sample towns will be compiled using secondary sources of information (e.g., directories of providers available online or through organizations such as the Federal Health Resources and Services Administration or State or local health departments), supplemented and confirmed as needed by telephone conversations with key informants knowledgeable about health care services in the selected towns. These lists will be used to select a census or simple random sample of each category of providers in each town, depending on the number of providers in the town (more on the sampling approach in the next section).

The population of community leaders and other stakeholders involved in recruiting and retaining health care providers is less well-defined than the population of health care providers, and simple quantitative criteria for selecting these respondents cannot be specified. Which leaders or stakeholders are involved or interested in activities to recruit and retain health care providers may vary greatly across communities (e.g, in one town the mayor may be the key person, while in another town, a banker or baker may be key). The population of relevant and important community leaders and other stakeholders for this issue will therefore be determined primarily through telephone conversations with key informants, but based upon our initial hypotheses about which types of leaders are likely to be relevant and important. This includes county health and economic development officials, the town mayor and/or city manager, the city council, administrators of hospitals and clinics, members of

local hospital or clinic boards or support organizations, and members of the Chamber of Commerce.

Given that the population of community leaders and other stakeholders that are potential respondents cannot be precisely defined, the rationale for and feasibility of using a random sample of this population is limited. Instead, these respondents will be selected purposively to represent people and organizations judged by key informants to have an important influence on or interest in the process of recruiting and retaining rural health care providers. We will not be able to claim statistical representativeness of this sample, and this caveat will be clearly stated in any analysis of the results of this survey. Despite this caveat, the views of community leaders and stakeholders are expected to lead to important insights and hypotheses for future research concerning the community level factors affecting recruitment and retention of health care providers and the influence of health care provision on community economic development.

The expected response rate is 80%. The methods that will be used to maximize the response rate are discussed in a subsequent section.

2. Procedures for the collection of information

Sample selection

150 of the 809 towns in the study universe will be selected using a stratified random sample. Six strata will be used: the three study regions (LMD, SGP, UMW) x whether or not the town has a hospital. The allocation of the sample among the strata will be based on minimizing the variance in the number of primary care physicians per population, which is expected to be strongly correlated with most of the variables investigated in the survey (e.g.,

local factors affecting health care provision, availability and quality of health care). Within each stratum, systematic sampling with a random start will be used, with the list of towns sorted by two variables: a variable based on ranges of the number of primary care physicians in the towns, and the population size of the town. The use of systematic sampling with a random start produces unbiased and consistent estimates of population parameters and is more efficient than simple random sampling if the variables used to sort the sample are correlated with the response variables.⁵

Of the 150 sample communities, 10 will be selected for inclusion in the pilot phase of the study, as noted in Part A. These 10 communities will be selected from the different strata discussed above. The data collected from the pilot phase communities will be combined with the data collected from the remaining 140 sample communities in the analysis of results, except for survey questions that are revised or dropped as a result of the pilot study.

The sample of respondent health care providers within each sample town will be selected using a stratified simple random sample, in which the types of providers are the three strata (i.e., physicians, dentists, and other providers (physician assistants, nurse practitioners, certified nurse midwives, and pharmacists), each as one stratum). The number selected within each stratum will be proportional to the total number of providers in that stratum working in the town, and the total number selected will be limited to a maximum of 8. If there are 8 or fewer providers working in the town, a complete census of providers will be interviewed. The reason for limiting the sample to a maximum of 8 providers per sample community is to limit the cost of the survey and the burden on members of any particular town, while allowing a sufficient number of observations to be representative of the

⁵ If the variables used to sort the sample are uncorrelated with the variables of interest and the sort order is random, systematic sampling and simple random sampling are of similar efficiency (see Särndal, et al. (2003), section 3.4).

providers in the town. A complete census of all providers within each town is not necessary for the purposes of this study, and would not be efficient because the information collected from multiple providers in the same town is not likely to be completely independent.⁶

The sample of respondent community leaders and other stakeholders will be selected purposively, as described in the previous section. A maximum of 8 community leaders and stakeholders will be interviewed. In towns where fewer than 8 leaders and stakeholders are identified as interested and important in influencing recruitment and retention of health care providers, the number interviewed will be less than 8.

Considering the maximum sample of 8 health care providers and 8 community leaders and other stakeholders per sample community, and the expected response rate of 80%, the maximum number of potential respondents in the 150 sample communities is 3000 (150 x 16/80%), as noted in Part A.

Estimation procedure

The sample means and variances of the quantitative survey response variables will be estimated using the formulas provided in Annex B, which reflect the complex two stage sample design. Comparisons of means of different subpopulations (e.g., communities with vs. without a hospital, communities in different regions) will be based on the formula (which follows from the Central Limit Theorem):

$$1) \frac{(\widehat{y}_i - \widehat{y}_j) - (\bar{y}_i - \bar{y}_j)}{\sqrt{V(\widehat{y}_i - \widehat{y}_j)}} \xrightarrow{D} N(0,1),$$

where \widehat{y}_i and \widehat{y}_j are the sample means of y in subpopulations I and J , respectively (with $i \in I$ and $j \in J$); \bar{y}_i and \bar{y}_j are the subpopulation means of y in subpopulations I and J ;

⁶ We discuss the issue of non-independence and its implications for sampling efficiency further in a subsequent subsection.

" $\xrightarrow{D} N(0,1)$ " means "converges in distribution to a standard normal distribution"; and $V(\cdot)$ denotes the variance operator.

The variance of the difference in subpopulation means is estimated by the formula:

$$2) \hat{V}(\hat{y}_i - \hat{y}_j) = \hat{V}(\hat{y}_i) + \hat{V}(\hat{y}_j) - 2\widehat{Cov}(\hat{y}_i, \hat{y}_j),$$

where \hat{V} and \widehat{Cov} refer to the estimated variance and covariance, respectively.

In cases where the subpopulations being compared are sampled independently, such as when they are from separate strata (e.g., if comparing communities with vs. without a hospital or communities in different regions), the covariance term in equation 2) is zero, simplifying the estimation. To compare subpopulations that are not from separate strata (e.g., communities with low vs. high numbers of health care providers per capita), multiple regression analysis will be used to account for covariance among the subpopulations.

An equivalent procedure to estimate the differences in means between subpopulations is to use an ordinary least squares (OLS) regression, with dummy categorical variables for each of the subpopulations except one (and using the appropriate estimators of the regression coefficients and variance matrix to reflect the complex sample design).⁷ The coefficients of each of these dummy variables represent the difference in means between the subpopulation represented by the dummy variable and the excluded category.

The analysis using OLS regression will be extended to include other observed covariates (x) (some of which will come from other available data sources) that are expected to be correlated with the y variables (e.g., population size of the community; proximity to a metropolitan area, a highway, or natural amenities) as explanatory variables. There are three

⁷ Appropriate regression estimators for complex two stage sample designs are available in the software program Stata, which will be used for the analysis. The variance estimator used by Stata assumes a stratified simple random sample in the first stage, which yields a conservative estimate of the variance for the case of a stratified systematic sample (which we are using in the first stage).

purposes of this extension: 1) to account for the influence of other potentially confounding factors on the differences that are being estimated (e.g., accounting for the fact that differences in mean responses between communities with vs. without a hospital may be due to differences in their population size, access to a city or other factors); 2) to investigate the effects on the response variables of these other factors, many of which represent different types of assets and whose influence are part of the study objectives; and 3) to reduce the residual unexplained variance, which will improve the efficiency of the estimated differences between target subpopulations.

One drawback of using OLS regressions to estimate the effects of factors on a binary response variable is that the predicted mean values of the response variable can be outside of the range of feasible values; i.e., greater than 1 or less than 0. Maximum likelihood discrete choice models, such as probit and logit models, are superior in this regard.⁸ Hence we will also use probit models to estimate the influence of covariates on the probability of a positive response for binary response variables.⁹ For multiple valued ordered response variables, ordered probit models will be used. Almost all of the response variables collected in the survey will be either binary or ordered variables, so these models will be commonly used.

Degree of accuracy needed and confidence intervals

Our objective is to maximize the accuracy and precision (or, in statistical terms, to minimize a combination of bias and variance, such as mean square error) of estimates within the available budget, and the sampling approach has been developed with this objective in mind.

⁸ Furthermore, if the distribution function assumed in the maximum likelihood model is the correct one, maximum likelihood is the most efficient estimator. However, violations of distributional assumptions, which can't be tested in binary response models, can cause maximum likelihood estimators to be inconsistent and inefficient. See Maddala (1983) for a discussion of maximum likelihood discrete choice and ordered response models.

⁹ The difference between a probit and logit model is in the univariate distribution function assumed in the maximum likelihood model (normal vs. logistic). In practice, the results of these models are quite similar.

We can provide an indication of the accuracy and precision that can be expected by constructing 95% confidence intervals of the form:

$$CI_{95}(\bar{y}_i - \bar{y}_j) = (\hat{y}_i - \hat{y}_j) \pm 1.96 \sqrt{\hat{V}(\hat{y}_i - \hat{y}_j)},$$

using the formulas for the means and variances in equation (2) and Annex B. We do not have the information necessary to estimate these means and variances prior to our study, but we can estimate ranges of these confidence intervals based upon the sampling approach, conservative assumptions about unknown factors determining the variances, and sensitivity analysis using a range of assumed true population means.

Our task is simplified because almost all of the response variables measured in the survey are either binary valued (0/1) variables or ordered categorical variables (typically with three to five responses, such as values from one to five representing “strongly disagree” to “strongly agree”). For binary response variables, the estimation is simplified because the variance of such a variable is determined by the mean; i.e., if p is the probability of a 1 for a binary response variable in a population, this is also the population mean of the variable and $p(1-p)$ is the population variance. We use this fact to estimate the sample variances.

Ordered response variables can be represented by a set of binary response variables, so the approach to estimating confidence intervals for binary response variables can be generalized to ordered response variables. For example, four binary response variables can be defined to represent a single ordered response variable (y) having five response levels from 1 to 5: $d_5 = 1$ if $y=5$ (e.g., if “strongly agree”), 0 otherwise; $d_4=1$ if $y \geq 4$ (“agree” or “strongly agree”), 0 otherwise; $d_3=1$ if $y \geq 3$ (“neutral”, “agree”, or “strongly agree”), 0

otherwise; and $d_2=1$ if $y \geq 2$ (“disagree”, “neutral”, “agree”, or “strongly agree”), 0 otherwise.¹⁰

Then comparisons of means of these binary variables could be used, for example, to test whether a larger fraction of one subpopulation agrees or strongly agrees (i.e, the proportion for which $d_4=1$) with a particular statement than another population.

We estimate confidence intervals using a conservative assumption related to the estimation of the variance in equation 2); namely, we assume no independence of observations taken from different respondents in the same community. In this case, the variance of means estimated for the respondent population is the same as the variance of means estimated as if there were only one respondent per community, and is larger than the variance that would be estimated if there is some independence of the within community observations. To illustrate this, consider the approximate formula for the variance of the sample mean for a two stage simple random sample (without replacement in each stage), with an equal population size (M) and sample size (m) within each cluster (this is simpler than our sampling approach, but serves to illustrate the point):¹¹

$$3) V(\hat{y}_i) \cong \left[\left(1 - \frac{n}{N}\right) \frac{(1+(M-1)\rho)}{nM} + \left(1 - \frac{m}{M}\right) \frac{(1-\rho)}{nm} \right] S^2,$$

where n is the number of clusters sampled in the first stage, N is the total number of clusters in the population of clusters, ρ is the intraclass correlation coefficient (ρ ranges from 0 for completely independent observations within a cluster to 1 for no independence of observations within a cluster), and S^2 is the population variance of y .

If $\rho = 1$, equation 3) reduces to:

$$4) V(\hat{y}_i | \rho = 1) \cong \left(\frac{1}{n} - \frac{1}{N} \right) S^2,$$

¹⁰ There is no need to create a fifth binary variable with $d_1=1$ if $y \geq 1$ and 0 otherwise, because this variable will always equal 1.

¹¹ This formula was derived from the formulas in Chapter 4 of Särndal, et al. (2003).

which is the estimator for the variance of y with a single stage simple random sample (without replacement).

At the other extreme, if $\rho = 0$, equation 3) reduces to:

$$5) V(\hat{y}_i|\rho = 0) \cong \left[\frac{1}{nm} - \frac{1}{NM} \right] S^2 = \frac{\left(1 - \frac{n}{N}\right) \left(\frac{m}{M}\right)}{m \left(1 - \frac{n}{N}\right)} V(\hat{y}_i|\rho = 1).$$

Equation 5) implies that $V(\hat{y}_i|\rho = 0) \leq 0.5 V(\hat{y}_i|\rho = 1)$ if $M \geq 2$ and $m \geq 2$.¹² For values of ρ between 0 and 1, $V(\hat{y}_i|\rho)$ varies linearly with ρ between these extreme values.

Using equation 2) and the formulas for our variance estimators in Annex B, we estimated the half width of the confidence interval (“CI half width”) for the difference between subpopulation means conservatively assuming zero independence of observations within communities for ranges of assumed values of the subpopulation means. We consider comparing means of response variables for communities with vs. without a hospital, and for communities in the different regions of the study. The results of these calculations are shown in Figures 1 to 4.

The results for comparisons of communities with vs. without a hospital show that the maximum CI half widths (or “sampling error bounds”) are about 0.15 or less for all parameter values considered, and largest at the midrange of probabilities, because the variance of a binary random variable is largest at $p = 0.5$. The sampling error bounds are somewhat larger for comparing means from different regions – but generally less than 0.20 – since the sample size within regions is smaller than the number of sample communities with or without a hospital.

¹² Proof: If $M = m = 2$, $\frac{\left(1 - \frac{n}{N}\right) \left(\frac{m}{M}\right)}{m \left(1 - \frac{n}{N}\right)} = 0.5$. This ratio is a declining function of both M and m (for $m \leq M$), so if m (and therefore M) is greater than 2, $V(\hat{y}_i|\rho = 0) < 0.5 V(\hat{y}_i|\rho = 1)$.

These results imply that we will have fairly good statistical power to detect a true difference between means of a binary response variable of at least 0.20 across regions or between communities with vs. without a hospital, under conservative assumptions. For example, the statistical power to detect a true difference between a population mean of 0.5 for communities without a hospital and 0.7 for communities with a hospital is 0.759, assuming the size of the test is 0.05 and no independence among observations within communities.¹³ If there is some independence among observations within communities (which we expect), the actual sampling error bounds will be smaller and the statistical power larger than this.¹⁴ Furthermore, use of a multiple regression approach will also reduce the error bounds, by reducing the residual variance not accounted for by observed covariates.

¹³ It can be shown that the statistical power to detect a true difference in means (i.e., the probability of correctly rejecting the null hypothesis of no difference, when it is false) is related to the CI half width and the size of the true difference in means, by the formula: $Power = 1 - \Phi(Z_\alpha(1 - \frac{|\Delta|}{CI_{hw}}))$, where $\Phi(\cdot)$ is the standard normal distribution function, Z_α is the critical value of the standard normal distribution function, which depends on α , the size of the test (= 1.96 if $\alpha = 0.05$), CI_{hw} is the CI half width, and Δ is the true difference in means.

¹⁴ For example, equations 3) and 4) imply that if $n/N \ll 1$, $V(\hat{y}_i | \rho = 0) \cong \frac{V(\hat{y}_i | \rho = 1)}{m}$. Hence, if $m = 4$ in this example, the variance would be about one fourth as large and the sampling error bounds would be about one half as large if $\rho = 0$ than if $\rho = 1$. If $m = 4$ and $\rho = 0.5$, the variance would be about 5/8 as large as if $\rho = 1$ (since 5/8 is the midpoint between 1/4 and 1), and the sampling error would be about $\sqrt{5/8} = 0.79$ as large.

Figure 1. Maximum CI half widths for communities with vs. without a hospital

(for a range of $P(\text{yes}|\text{without hospital})$, $P(\text{yes}|\text{with hospital}) - P(\text{yes}|\text{without hospital}) = 0.1$)

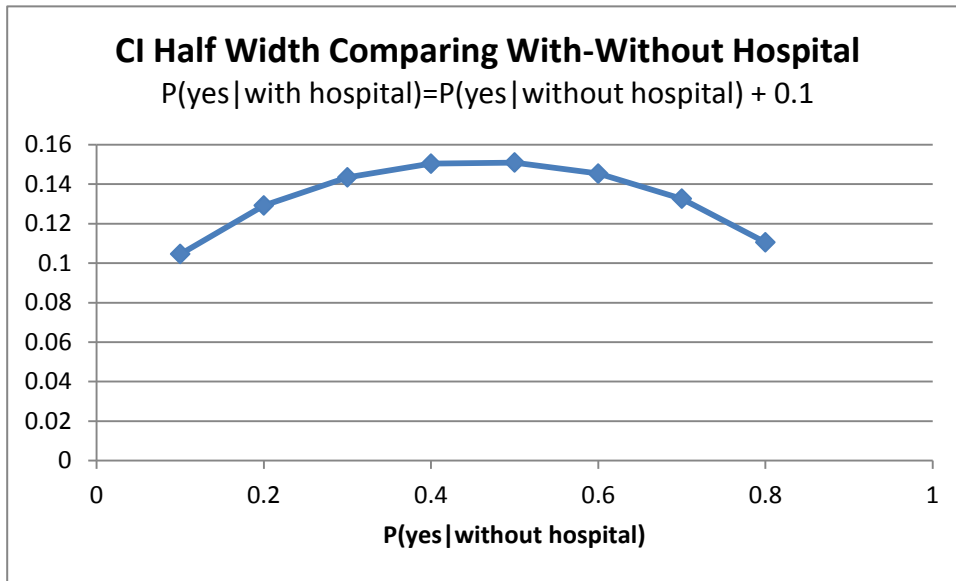


Figure 2. Maximum CI half widths for communities with vs. without a hospital

(for a range of $P(\text{yes}|\text{with hospital}) - P(\text{yes}|\text{without hospital})$, $P(\text{yes}|\text{without hospital}) = 0.5$)

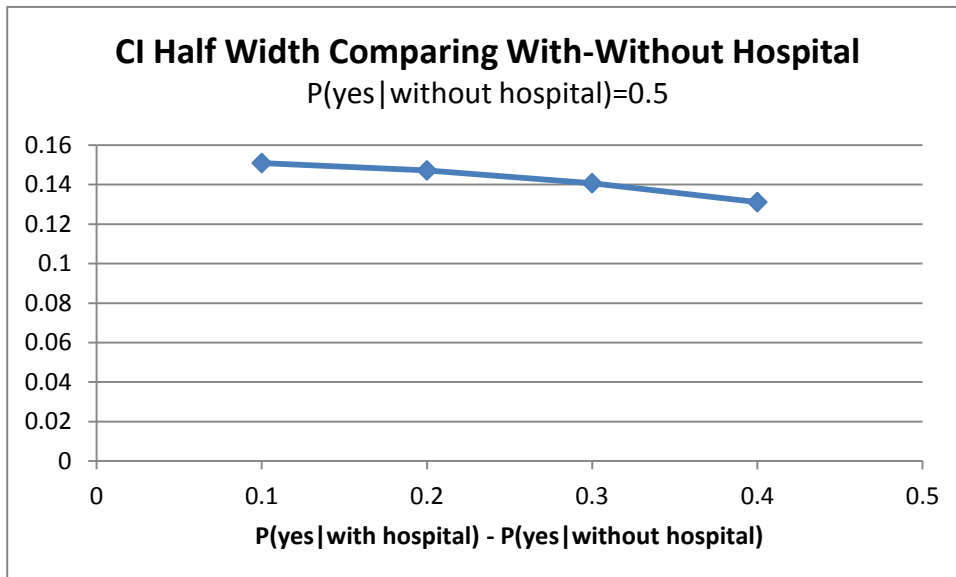


Figure 3. Maximum CI half widths for communities in different regions
 (for a range of $P(\text{yes}|\text{region } j)$, $P(\text{yes}|\text{region } i) - P(\text{yes}|\text{region } j) = 0.1$)

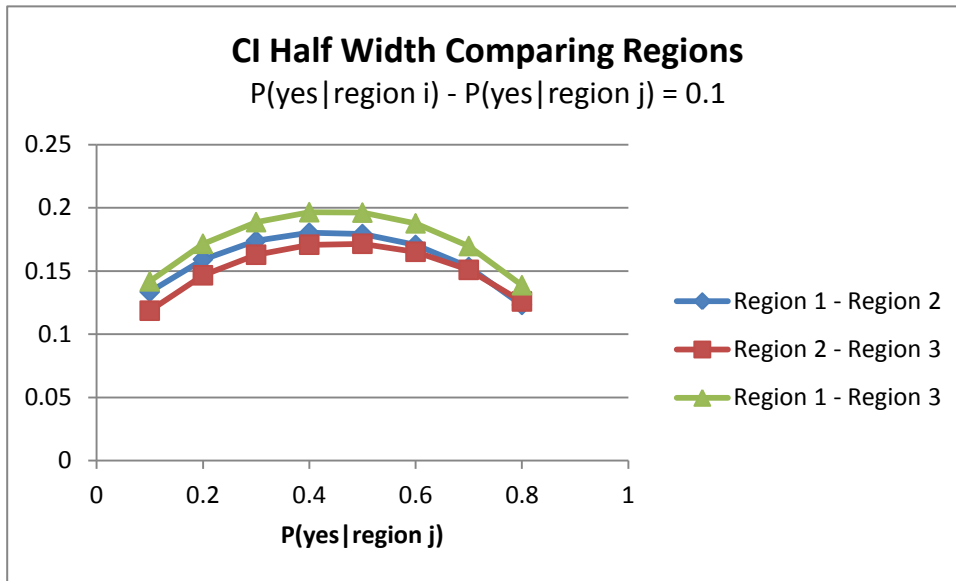
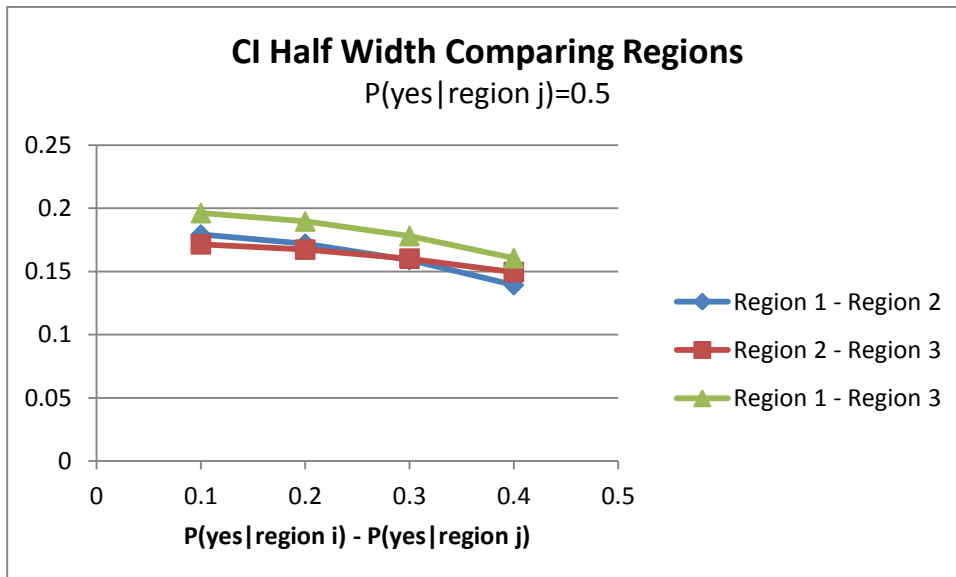


Figure 4. Maximum CI half widths for communities in different regions
 (for a range of $P(\text{yes}|\text{region } i) - P(\text{yes}|\text{region } j)$, $P(\text{yes}|\text{region } j) = 0.5$)



Unusual problems requiring specialized sampling procedures

The only unusual problem requiring specialized sampling procedures is that the relevant population of community leaders and other stakeholders cannot be defined in advance, as discussed above. We therefore will select these respondents purposively, based upon information collected during the sampling frame development in each community. We cannot claim statistical representativeness of information collected from this stratum. Hence, this stratum will be analyzed separately, and our inability to draw representative conclusions about the underlying population will be stated clearly in any publications based upon this survey.

Use of periodic data collection cycles

Not applicable; this is a one-time data collection effort.

3. Methods to maximize response rates and deal with non-response issues

Several research procedures will be incorporated to maximize the telephone survey response rate. Potential respondents will be provided with advance information about the project from a variety of sources. Notices and/or articles about the upcoming project will be placed in several applicable publications, such as the Journal of Rural Health and Rural Roads magazine, and on websites such as RuralHealthWeb.org and the National Rural Health Association website. Advance letters and a colorful informative sheet/brochure will be mailed to sampled individuals prior to telephone contact. A project website will be available with additional information, and a toll-free number will be provided for those who have questions or concerns. Potential respondents will be advised that project results will be

available to them and other members of their community through the project website within 6 months after the completion of telephone data collection.

Operationally, limited samples will be drawn from the frame, with replicates added only as necessary. The data collection staff will be prepared to accommodate a wide range of respondent schedules and needs. In addition, paper or online copies of the survey will be made available to respondents who are unable or unwilling to complete a telephone interview.

The pilot study in 10 communities will be particularly useful in identifying and evaluating any potential response rate and non-response issues. Should any such problems arise during the pilot study, procedures can be adapted to address the issues at hand so that response rates for the ensuing 140 communities will be maximized.

We have a target response rate of 80% and believe that this target is achievable based upon similar surveys conducted by the Survey and Behavioral Research Services (SBRS) group at Iowa State University – the organization that will implement the survey. In 2009, SBRS completed data collection for a community-based study that focused on the impact of the retail sector in small rural communities (*Community Resiliency: Role of the Retail Sector in Easing Sudden and Slow Motion Economic Shocks*). A sample of 32 communities in 8 states was chosen, and a total of 1,161 interviews were conducted with retail business owners, community leaders, and community residents (general population). The sample for the retail owners and community leaders was developed in a manner similar to the process for the proposed project. The response rate for the community leaders was 76.4%, with a 61.8% response rate for the retail business owners. For the proposed project, it is expected that the response rate will exceed that of the *Community Resiliency* project and could reach

the target of 80%, given the significant follow-up measures planned (more than in the *Community Resiliency* project) and the high saliency of the topic (health care in rural communities) for the respondents.

If the unit response rate in the pilot study is less than 80 percent, an investigation of potential nonresponse bias will be planned and implemented in the full survey, using data available from secondary sources, the sample frame, and the survey. For example, we could investigate whether potential respondents having higher expected salaries were less likely to participate as respondents in the survey (due to higher opportunity costs of their time). The analysis in this example would combine data from secondary sources such as the Bureau of Labor Statistics on salary ranges existing for particular health care occupations in particular locations, information from the sample frame on the location and specific occupations of the potential respondents, and information from the survey on which potential respondents decided not to participate. Data from the survey could also be used to investigate the extent to which the responses to key survey questions – such as the importance of different types of community assets in affecting health care providers’ decisions to work in the community – differ between respondents from different occupational groups or different locations. Combining these types of analysis will enable an assessment of the two components of non-response bias: i) differences between respondents and non-respondents in characteristics hypothesized to influence response variables of interest; and ii) the extent to which such characteristics are correlated with the response variables of interest.

4. Tests of procedures or methods to be undertaken

The pilot study will be used to evaluate the survey protocol (e.g., the procedures used to identify, inform and contact respondents, schedule interviews, etc.) and to estimate the response rates for individual questions and the unit response to the survey as a whole. If the response rate is below 70 percent for individual questions, these questions will be revised to improve response rates or dropped in the full survey. If the response rate to the pilot survey as a whole is below 80 percent, efforts to maximize the response rate will be applied to the fullest extent possible, and an investigation of nonresponse bias will be conducted, as described in the preceding section.

5. Names and telephone numbers of people consulted on statistical methods; people who will collect the data

Name	Position	Telephone
Cindy Yu, PhD	Project Statistician, Assistant Professor of Statistics, Iowa State University	515-294-6885
Shirley Huck	Assistant Director, Survey & Behavioral Research Services (SBRS), Iowa State University	515-294-1652
Janice Larson	SBRS Survey Unit Director	515-294-3451
Allison Anderson	SBRS Project Manager	515-294-1949
Jody Fox	SBRS Data Collection Supervisor	515-294-4289
Debbie Bahr	SBRS Data Collection Supervisor	515-294-3104
Kihwan Kim	Programmer & Statistical Support	515-294-8197