Are Response Rates to Web-Only Surveys Spatially Clustered? Implications for Understanding Geographic Bias and Coverage in General Population Web Surveys.

Michael J. Stern[1], Ned English, and Lee Fiorio
NORC at the University of Chicago
55 East Monroe Street, 30th Floor, Chicago IL 60603

Abstract
Over the past decade, researchers have learned a great deal about the design and implementation of Web surveys. However, to date, we have virtually no empirical information about the role space and place has in estimating forms of error associated with Web-only surveys. The two types of error most often discussed when considering Web surveys are coverage and non-response; both of which are typically indicated as reasons for low response rates in these types of surveys. One way to pursue this issue of place is to use Geographic Information Systems (GIS) to spatially-model survey response rates. This will allow us to understand the impact of location on error in Web surveys. In this paper, we attempt to examine this gap in the literature by assessing the spatial clustering of response rates to a Web-only survey. The data come from a random, Address- Based Sampling Approach using the Delivery Sequence File (Valassis version) where respondents received a postal letter with a URL. We calculate response rates at several geographic scales, including county, state, and region, to determine the extent to which response rates are spatially clustered. While controlling for ACS demographics, internet availability, and postal characteristics, we then build a spatial lag model to measure spatial dependence of response rates observed. Preliminary findings show clusters of low response rates in the South that cannot be accounted for by other variables in the model.

Keywords: Web Surveys; Geographic Information Systems (GIS); Spatial Models; Coverage Error.

Introduction
With the decline in interview response rates for general population surveys, researchers have looked to the internet as the eventual replacement for the telephone. Yet, for a variety reasons the internet has not been able to fill this role as of yet. For example, household internet coverage as of 2012 is only about 75% (NTIA 2011) and in those households with connections, some individuals lack the skills to use it, are uncomfortable with it, or use it infrequently (Stern et al. 2009). Contacting individuals via the Web is also problematic given that no acceptable sample frame for email addresses exists with coverage equivalent to random-digit dial telephone. Additionally, when email addresses are known, studies show realizing an acceptable response rate through email contact is difficult as complicated by the inability to provide a pre-incentive.

One possible solution to the "mode of contact" conundrum has been to use mail to contact respondents and ask them to go to the Web to complete a survey. In this case, the contact letter could include a URL that the respondent may insert into their browser in addition to including a token incentive (e.g., a $5.00 bill) in the mailing. Such an approach has been proven effective in some cases especially when used in combination with another self-administered mode (Smyth et al 2010; Messer and Dillman 2011). However, numerous questions need to be addressed about using a mailed web address approach before we will understand its limitations. For instance, to date, we have virtually no empirical information about the role space and place has in estimating coverage or nonresponse error associated with Web-only surveys that use a mail contact.

[1] Direct all correspondence to Michael J. Stern at NORC at the University of Chicago, 55 East Monroe Street, 20th Floor, Chicago IL 60603 stern-michael@norc.org

One way to pursue this issue of place and spatial relationships is to use Geographic Information Systems (GIS) to model spatially survey response rates. Such an approach will allow us to understand the impact of location on error in Web surveys. In this paper, we attempt to examine this gap in the literature by assessing the spatial clustering of response rates to a Web-only survey. The data come from a random, Address- Based Sampling Approach using an extract of the United States Postal Service Delivery Sequence File[2] (DSF or CDSF) where respondents received a postal letter with a URL. We calculate response rates at several geographic scales, including county, state, and region, to determine the extent to which response rates are spatially clustered. While controlling for ACS demographics, internet availability, and postal characteristics, we then build a spatial lag model in several different ways to measure spatial dependence of response rates observed.

Theoretical Background

Challenges for Survey Methodology

Over the past two decades, contact rates for random digit dial (RDD) samples and other direct interview-based survey methods have experienced a considerable decrease. A primary reason for this decrease centers on low coverage for RDD samples, which generally include landline telephone numbers with a mobile telephone augmentation (Blumberg et al. 2011). Numerous studies have shown that there are significant declines in landline connections, particularly for young people, households with children, and for people with changing social relationships such as separation and divorce, make coverage a greater problem (e.g., Keeter et al. 2007). Although mobile phones numbers are often added to supplement landline sample frames to increase coverage, this dual frame approach introduces new challenges. For instance, in the United States, there are instances where incoming calls are counted against the respondent's prepaid minutes or data plans and prepaid plans are disproportionally used among low-income families (Brick et al. 2011). In addition, mobile phones produce geographic location challenges because individuals commonly keep the same mobile phone number after they relocate, rendering the area code plus exchange a poor indicator of residence.

With the recent decline in interview response rates, there was much optimism that the internet would become the survey replacement for the telephone. Yet, significant challenges and a large number of unknowns remain associated with web surveying. For instance, contacting respondents and driving them to the web has proven to be difficult. Specialized populations notwithstanding (e.g. attendees at conferences or recent doctorate graduates), very few email-only surveys have been effective at achieving high response rates. In addition, there is no acceptable general population sample frame for email addresses and professional norms prevent sending survey requests to individuals with which no prior relationship (e.g. client, student, and association member) exists. Still, one can use mail or telephone contact to prompt people to respond over the web. Using mail or telephone as such adds to the cost, but also has to be effective in convincing people to make the effort of going from that mode of contact to a computer where they enter a URL and password (Anderson and Tancreto 2011). In addition, there has been little research with respect to how well surveys fare in areas of poor internet penetration or low computer literacy. Nonetheless, increasingly researchers are using this mail approach, relying on address-based sampling (ABS), to contact potential web respondents.

ABS has become a popular method of executing single- or multi-mode studies in recent years. One reason is that over the past decade it has become possible to license extracts of the United States Postal Service Delivery Sequence File (DSF or CDSF) from a particular set of vendors. The DSF embodies a list of all housing units in the United States that receive mail (O'Muircheartaigh, Eckman, and Weiss 2003, Amaya et al. Forthcoming). Survey research and government organizations have been researching the use of the DSF as a replacement for traditional listing, due to the implications for cost savings (O'Muircheartaigh et al, 2003, Iannacchione et al. 2003, O'Muircheartaigh et al. 2007, Link et al. 2008). The sum total is that the DSF is often adequate itself in urban areas, but may not be so in rural areas with non-city-style delivery (Staab and Iannacchione, 2003, Link et al. 2008, O'Muircheartaigh et

[2] Provided by the Valassis vendor

al. 2009). Consequently, there needs to be some kind of listing or other augmentation in places where the DSF is insufficient. Besides handling non-coverage, users also need to understand various delivery types and their implications for conducting a given type of survey.

Being a raw extract of all addresses that receive mail in the United States means that there are address types that are more or less likely to be conducive to an ABS survey. One primary issue is generally the presence of non-city-style addresses, generally the form of post office (PO) or rural route (RR) boxes on the DSF. While such addresses receive mail, they do not contain enough information to link to the requisite dwelling unit. Post-office boxes may also be leased by respondents in urban areas, requiring the distinction between those where it is the "only way to get mail" from others to avoid the potential for duplication. In addition, the DSF contains any delivery points that may be vacant, seasonal, throwbacks, or drop-delivery. We would expect such complexities to affect both how likely a survey is to reach the intended household as well as the likelihood they would respond to it. The sum total is that using a raw extract of the DSF may introduce variable levels of response based on urbanicity and housing types. Furthermore, ABS designs also cause a shift in the type of people who respond. ABS surveys cover the vast majority of the population, but they do not always achieve response rates from a representative sample. Using data from the Census Barriers Attitudes and Motivators Survey, Bates & Pan (2009) found that Asian and Hispanic individuals who speak a language other than English in their homes were significantly less willing to respond to the Census than were English-speaking Asian and Hispanics.

What is Spatial Inequality?

We present this work within the framework of spatial inequality. Lobao (2004: 1) defines spatial inequality as "stratification within or between territorial units." Therefore, studying spatial inequality requires the empirical examination of variation across spatial units and the consideration of how varying degrees of access to resources affect different segments of society. Most of the studies addressing spatial inequality and internet use have focused on the problems associated with unequal access in rural places to the internet. For example, Warren (2007) argues that rural residents who have the most to gain by the declining significance of physical distance and geographical boundaries resulting from internet use. Yet, they also have the most to lose with poor or inefficient access to and low proficiency with the technology. What is more, rural populations tend to have lower levels of education, which results in less experience with information technologies; thus, while all rural residents may be at some disadvantage, the most vulnerable segments of rural populations including the poor, those with lower levels of education, will be left even farther behind.

The current running through much of this research is a focus on technological diffusion, in particular high-speed access. In terms of spatial inequality, there are two interrelated angles from which one must consider this issue: the demand-side and supply-side (Khatiwada and Pigg 2010). In terms of the supply-side, the argument has been that the expense of bringing high-speed, broadband access to remote places is not offset by the profits that could be made in doing so because residents are not motivated or capable of paying for the services (Whitacre 2010). However, Whitacre and Mills (2007) have used longitudinal data to examine policies that promote broadband infrastructure. What they find is that these programs are most successful after residents understand the benefits of the technology (See also Whitacre 2007). Mosseberger et al (2008) and Stern et al (2009) provide support for this position by showing that when broadband is available in communities, technological proficiency increases as does the diversity of uses. Nonetheless, there is still variation within and between spatial units (e.g., regions) in terms of broadband availability.

In order to measure variation of a particular phenomenon across a given area, each observation must be ascribable to a location, either to a point such as a latitude/longitude coordinate or to a subdivision of the larger area of interest such as a county within a state. With locational data, it is then possible to define the spatial relationship between each observation and the remaining observations either by calculating distance or indicating a shared border between each pair. The resulting $n \times n$ spatial relationship matrix would be considered a geographic information system or GIS and could be operated on to conduct spatial autocorrelation or spatial regression analyses.

In this study, we are interested in using GIS to perform a spatial autocorrelation analysis for the purposes of identifying regions or clusters of subdivisions with high or low internet connectivity and high or low response to a web-only ABS survey. To do so, the unit of observation would be some kind of subdivision of the United States like counties or amalgamations of counties for which we can calculate internet connectivity rate , study response rate, or other independent variables. Using a spatial relationship matrix indicating contiguity between subdivisions, we can then measure the degree of spatial dependency of our phenomena of interest. The theory behind such an analysis is the oft cited Tobler's 1st law of Geography, "Everything is related to everything else, but near things are more related than distant things" (1970). Working under the null hypothesis that the rates are distributed randomly among subdivisions, we can conduct a spatial autocorrelation analysis to determine if subdivisions with similar rates neighbor each other. Visualizing this data will then allow us to observe regions in the United States that are statistically different from the global average. To take this further, we can use the same spatial relationship matrix in subsequent regression analyses to control for the influence of geography.

Analytic Approach

We were faced with several conceptual obstacles in measuring the spatial clustering of response to our web-only survey. For one, because survey prompts were mailed to a nation-wide sample of 10,000 addresses stratified by Census region, the spatial distribution of selected addresses naturally mirrored population density, such as of urban areas and the higher overall housing unit density east of the Mississippi River. With response rates to the web-only survey relatively stable by state, any cluster analysis performed on the spatial distribution of survey response would simply reflect the *a priori* distribution of housing units in the United States. Without taking into account the distribution of non-response, our findings regarding the distribution of response would be fairly meaningless. Thus, it follows logically that we should investigate the spatial clustering of the response rate.

While in theory the use of response rate as a dependent variable allows us to control for the *a priori* distribution of the sample, it exposed us to a number of additional conceptual concerns, chief among them the "modifiable areal unit problem" (Cressie, 1996). In order to calculate a response rate, we needed to identify zones into which completes and non-completes could be aggregated. Any zone selection process is ultimately arbitrary and has the potential to introduce unwanted bias into a spatial analysis. For example, Cook and DuPage counties are both neighboring in suburban-Chicago, Illinois. Each county had relatively high response rates to the web-survey, but what conclusions we derive about spatial clustering of response rate in Illinois will differ depending on whether Cook and DuPage are put into a single zone and compared with neighboring counties or split into separate zones and compared with each other. To avoid the modifiable areal unit problem, we decided to conduct our preliminary analysis at a number of different geographic scales and zone schema with the goal of identifying patterns that exist despite changes in the unit of analysis.

One obvious technique for selecting zones and calculating response rates was to use pre-existing political boundaries like states and counties. At the state level, the average response rate (unadjusted, non-AAPOR) was 7.8% with a standard deviation of 5.3%. Lows of 0% were observed in Alaska (0/20) and Delaware (0/21) while the highest response rate, 26.9%, was observed in North Dakota (7/26). Simple statistics quickly belie problems with working at the county level. The average response rate (unadjusted, non-AAPOR) among counties was 6.9% with a standard deviation of 18.8%. Over 75% of the 1,783 counties sent a survey request returned zero completed surveys while 2.6% of counties completed at a rate of 100%. Beyond the issues with distribution of response rate, working at the county level posed additional problems for spatial analysis. For instance, despite sampling 10,000 addresses, many counties were not sent a survey request and thus had no observation. The resulting patchwork of contiguous and discontiguous counties could not be reliably used for a nearest-neighbor cluster analysis. In general, states present themselves as a useful but coarse unit of analysis; going forward, we needed a geographic scale that was somewhere between state and county in size.

To develop medium-sized zones, we took two distinct approaches. First, we manually grouped counties together based on the number of sampled addresses, which was the denominator in a simple response rate calculation. To ensure sufficient sample per district, it was decided that 70 districts of roughly 143 sampled addresses each would be optimal. Among the 70 districts, the mean response rate (unadjusted, non-AAPOR) was 7.45% with a standard deviation of 3.14%. The lowest observed rate among districts was 2.23% in western Louisiana/eastern Texas while the highest was 15.0% in north central Washington. The second method we used to create zones was to overlay a 250 km^2 grid and use each panel or grid cell to calculate response rates. This approach created zones that were variable in number of observations but constant in geographic area. Among the 157 panels created, the mean response rate was 7.61% (unadjusted, non-AAPOR) with a standard deviation of 10.44%. Nearly 30% of the grid cells returned no completes, most of which were west of the Mississippi River with one panel in South Dakota completing two of two survey requests. The majority of cells had a rate between 40% and 6.3%. The spatial pattern observed in other maps was less clear using the 250 km grid approach; while we still observed high rates in the Dakotas and Nebraska, they were intermixed with spots of low response. Such a finding would be indicative of a more random distribution of response. Low spots were observed in the south, but as with the Dakotas, the pattern was less clear.

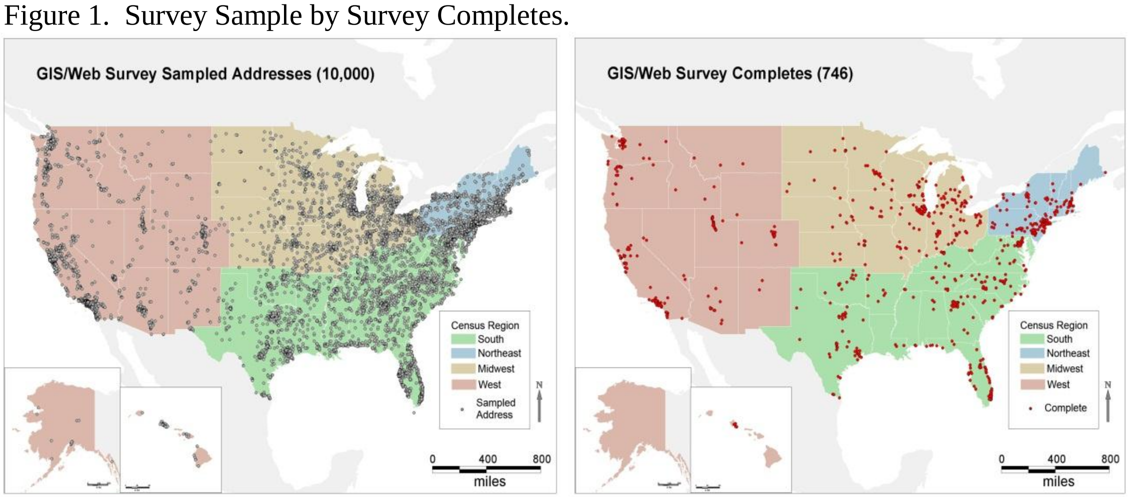
To further elucidate the clustering of response rates, we examined spatial autocorrelation using Moran's I (Moran, 1950). Calculated for each of the separate geographic scales described above, Moran's I measures the degree to which properties of a set of areal units are similar to the properties of their nearest neighbors. In this case, we used it to measure the average correlation between the response rate for a given zone (state, county, district, grid) and the response rate of the zones directly neighboring it. Ranging between -1 and 1, Moran's I when positive indicates clustering of similar properties across space while a score of zero indicates a random distribution. A negative score indicates a non-random non-clustered distribution conceptually similar to a checkerboard. To help visualize clusters and identify regions with response rates that differ statistically from the global average, we created a series of local indicators of spatial association or LISA maps. Created by Luc Anselin (1995), LISA statistics allow for the decomposition of Moran's I into each areal unit and make it possible to identify "hot spots" of statistically different behavior.

Data and Results

Here we present our results thus far using maps, quantitative geography, and multivariate models as an overview of our initial findings.

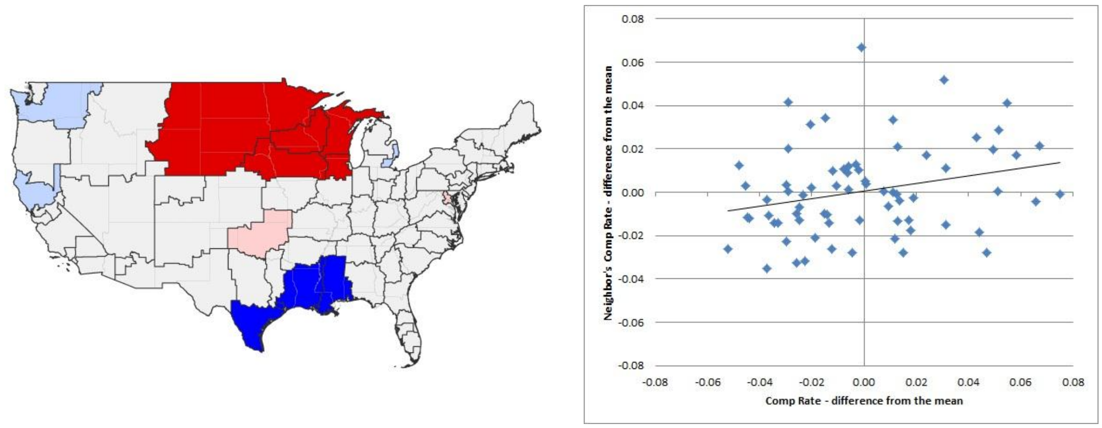
Our first set of maps show the overall geographic dispersion of the sample, as well as the spatial distribution of responses. In general, it appears that areas with high concentrations of sample produce the most returns, as one would expect.

Figure 1. Survey Sample by Survey Completes.



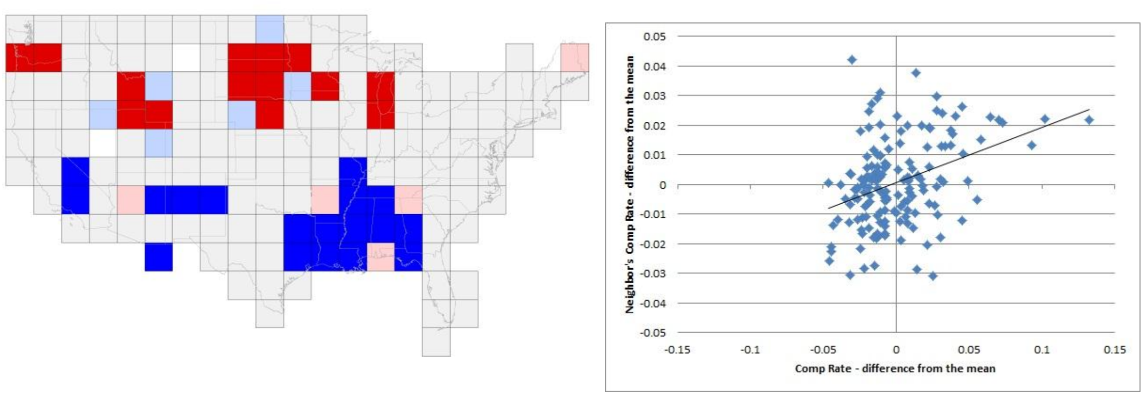
However, when we divide the country into districts and calculate the response rate for each as described above, a different story begins to be presented. In fact, response rates do appear to be spatially clustered considering region. In Figure 2, the red indicates significant clusters of high response rate and blue indicates significant clusters of low response rate. The LISA map indicates a cluster of low response rate in Mississippi Valley and TX and a cluster of high response rate in the upper Midwest. The Moran's I is a positive 0.25 with a p-value of 0.018 indicating clustering.

Figure 2. Spatial Clustering of Response Rates and Moran's I by District.



It is clear that the manually created districts are heterogeneous in their constitution. For example, the Northwest in essentially divided into only three districts even though there are vast expanses of virtually uninhabited areas clustered with few cities of over 100,000 people; this lack of specificity is potentially problematic and leaves open the possibility that patterns we observe are influenced by the scale of area units into which we aggregated the data. An alternative and more robust model may be built using a grid where each square represents an equal area. As described above, being too precise by using a small distance grid leaves us with too few cases in each cell whereas going too great in distance creates the district issue. Through our own experimentation, we have found that 250 Kilometers performed well with this dataset. Not surprisingly, the LISA map indicates significant low cluster in the Mississippi Valley and Appalachia and significant high cluster in eastern Dakotas. The Moran's I is a positive 0.19 with a p-value of 0.002 indicating clustering (Figure 3).

Figure 3. Spatial Clustering of Response Rates and Moran's I on 250k Grid.



Finally, we use multivariate models to assess the impact of the diffusion of technologies, socio-spatial demographics, and other controls on overall response as well as number of responses (Table 1). At a high level, we find that in places with high levels of fixed high-speed access we are more likely to get a response in general and higher levels of response overall. We also find that percent poverty has an effect on the overall likelihood of received a response but not the number of responses when people are inclined to respond. Linguistic isolation and percent white non-Hispanic both play a positive role in response whereas urbanicity shows a significant relationship while controlling for percent vacant. Although these results are preliminary, it does appear that is much to learn from the maps and models. With the spatial weight matrix created using the 250 kilometer grid, we can fit multivariate models to assess the impact of spatial autocorrelation, demographics, technology diffusion and other controls on overall response rate as well as the number of responses. At a high-level, it is apparent from our first group of models (Model 0 and Model 0A) that even while controlling for obvious factors urbanization, education levels and poverty, high-speed access exhibits strong spatial clustering behavior. That is, regions in the United States differ in their access to high-speed technology despite regional differences in education, urbanization, race, language and housing stock age. The strong spatial influence holds in our modeling of response; in both the generalized linear models predicting raw completion rate and the negative binomial models predicting count of completes, the spatial lag coefficient is the most powerful explanatory variable even while controlling for other factors.

Conclusions and Discussion

In this paper, we explored spatial clustering of response rates to a Web-only survey. Our goal was to understand the impact of location on error in Web surveys. In addition, we sought to investigate various ways to scale our analysis, thus adding to the literature on methodological approaches that combine spatial analysis and cross-sectional household-level data. The results of our work suggest several important conclusions and areas for future research.

As the extant research suggests Web access is spatially uneven. Thus, it is no surprise that we find responses to our Web only survey to follow this non-uniform, spatial pattern. Furthermore, the results here show that the representativeness of Web survey respondents is most often realized in areas of affluence, which are concomitant with significant levels of high-speed internet access (e.g., the Northeast). We can thus confidently argue that Web-only surveys are not an optimal, standalone approach for the general population. It was only through our GIS approach that we could both visualize and model clusters or “hot-spots” for Web survey response. The findings suggest that even when using an ABS approach responses disproportionality come from area with educated and high SES households. As a result, there is a risk of systematically excluding regions of the country that do not conform to these socio-demographic characteristics.

There were important limitations with this study. For instance, the response rate for this work was quite low even by Web survey standards. The reason for this response rate most likely can be found in the use of post incentives, which have historically performed below par. In a national-level survey that seeks to examine space as a key variable, a low response rate means that very few responses for large swaths of geographic territory. Using the 250k grid ameliorated sum of the problems but did not allow for the granularity we would have preferred. A second limitation was that we did not impute nor use any post stratification. There were so few responses that perhaps a form of imputation and weighting would have allowed for a more robust picture to be drawn but the fear of overgeneralizing to a given area prevented us from taking this step. In short, this was a reasonable first step but there is much work to be done in the area.

This paper has shown that Web-only surveys can be effective if one is interested in a particular type of household. However, as tool for the general population there are coverage issues at the household and geographic level. As a result, Web surveys are not the panacea once envisioned; at least not at this point in time. Researchers must continue to operate in the multimode environment when seeking a general population.

References

Amaya, A., F. LeClere, L. Fiorio, and N. English. Forthcoming. "Improving the Utility of the DSF Address-Based Frame through Ancillary Information." *Field Methods*.

Anderson, A. and J. Tancreto. 2011. "Using the Web for Data Collection at the United States Census Bureau." Full text: http://isi2011.congressplanner.eu/pdfs/450221.pdf Downloaded on February 24, 2013.

Bates, N. and Pan Y. 2009. "Motivating Non-English-Speaking Populations for Census and Survey Participation." Paper Presented at the Federal Committee on Statistical Methodology. Washington, DC, November 2-4, 2009.

Blumberg, S.J. and Luke, J.V. 2012. "Wireless substitution: Early release of estimates from the National Health Interview Survey, July–December 2011. National Center for Health Statistics. Available from: http://www.cdc.gov/nchs/nhis.htm. Downloaded on February 20, 2013.

Brick, J.M., D. Williams, and J. Montaquila. 2011. "Address-based sampling for subpopulation surveys." Public Opinion Quarterly, 75, 409-428.

Cressie, N. A. 1996. "Change of support and the modifiable areal unit problem." *Geographical Systems*, 3 (2-3), 159-180.

Iannacchione, V. G., Staab, J. M., & Redden, D. T. 2003." Evaluating the use of residential mailing addresses in a metropolitan household survey." *Public Opinion Quarterly*, 67 (2): 202-210.

Keeter, S., C. Kennedy, A. Clark, T. Tompson and M. Mokrzycki. 2007. "What's Missing from National Landline RDD Surveys?: The Impact of the Growing Cell-Only Population." *Public Opinion Quarterly* 715: 772-792.

Khatiwada L and Pigg K. 2010 "Internet service provision in the U.S. counties: is spatial pattern a function of demand?" *American Behavioral Scientist* 53(9): 1326–1344.

Lobao L. 2004. "Continuity and change in place stratification: spatial inequality and middle-range territorial units." *Rural Sociology* 69(1): 1–30.

Anselin, L. 1995. "Local Indicators of Spatial Association—LISA." *Geographical Analysis* 27: 93-115.

Messer, B. and D. Dillman. 2011. "Surveying the General Public Over the Internet Using Addressed-Based Sampling and Mail Contact Procedures." *Public Opinion Quarterly* 75(3):429-57.

Moran P.A.P. 1950. "Notes on continuous stochastic phenomena." *Biometrika* 37: 17–23.

Mossberger K, Tolbert C and McNeal R. 2008. *Digital Citizenship: The Internet, Society, and Participation*. Cambridge, MA: MIT Press.

NTIA, 2011. Exploring the Digital Nation: Computer and Internet Use at Home. Report prepared for the U.S. Department of Commerce. Full text: http://www.ntia.doc.gov/files/ntia/publications/exploring_the_digital_nation_computer_and_internet_use_at_home_11092011.pdf Downloaded on February 10, 2013.

O'Muircheartaigh, C., Eckman, S., and Weiss, C. 2003. "Traditional and Enhanced Field Listing for Probability Sampling." *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

O'Muircheartaigh and C., Eckman, S. 2007. "Efficiency and Bias in a Two-Phase (Subsampling) Model for Nonresponse in a Face-to-Face Survey." Paper presented at the European Survey Research Association conference, Prague, Czech Republic.

Smyth, J D., D. Dillman, L. Christian, and A. O'Neill. 2010. ''Using the Internet to Survey Small Towns and Communities: Limitations and Possibilities in the Early 21st Century.'' *American Behavioral Scientist* 53:1423–48.

Staab, J. and V. Iannacchione. 2003. "Evaluating the Use of Residential Mailing Addresses in a National Household Survey." *Proceedings of the American Statistical Association, Survey Methodology Section (CD-ROM),* 4028–33. Alexandria, VA: American Statistical Association.

Stern, M., A. Adams, and S. Elsasser. 2009. Digital Inequality and Place: The Effects of Technological Diffusion on Internet Proficiency and Usage across Rural, Suburban, and Urban Counties. *Sociological Inquiry* 79:391–417.

Warren, M. 2007 "The digital vicious cycle: links between social disadvantage and digital exclusion in rural areas. " *Telecommunications Policy* 31(6–7): 374–388.

Whitacre, B. 2008. "Factors influencing the temporal diffusion of broadband adoption: evidence from Oklahoma." *Annals of Regional Science* 42(3): 661–679.

Whitacre, B. 2010 "The diffusion of Internet technologies to rural communities: a portrait of broadband supply and demand." *American Behavioral Scientist* 45(3): 1283–1303.

Whitacre B. and Mills B. 2007. "Infrastructure and the rural urban divide in high-speed residential Internet access." *International Regional Science Review* 30(3): 249–273.

		GLM High-speed access		GLM Raw Completion Rate			Negative Binomial Count of Completes		
		Model 0	Model 0A	Model 1	Model 1A	Model 1B	Model 2	Model 2A	Model 2B
		β (se)	β (se)	β (se)	β (se)	β (se)	β (se)	β (se)	β (se)
Spatial Lag	Spatial Lag	0.987*** (0.103)	0.561*** (0.087)	0.616** (0.186)	0.544** (0.202)	0.471* (0.211)	0.098*** (0.019)	0.059** (0.019)	0.063*** (0.019)
Diffusion of Technology	High-speed access	NA	NA	-	0.010 (0.110)	0.031 (0.128)	-	2.337* (1.109)	1.922 (1.185)
Socio-Spatial Demographic Factors	Percent of HH that speak Spanish	-	-0.105 (0.063)	-	0.012 (0.102)	0.028 (0.110)	-	-0.837 (1.083)	-1.538 (1.110)
	Percent of population below poverty	-	-0.447* (0.196)	-	0.261 (0.320)	0.253 (0.328)	-	-0.676 (3.464)	0.809 (3.568)
	Percent of population with bachelors degree or higher	-	0.556*** (0.115)	-	0.216 (0.195)	0.163 (0.203)	-	5.092** (1.935)	1.706 (2.087)
	Percent non-Hispanic white	-	-0.063 (0.060)	-	0.075 (0.092)	0.076 (0.102)	-	-1.623 (0.838)	-0.935 (0.916)
Other Factors Controls	Percent HUs that are vacant	-	-	-	-	-0.210 (0.173)	-	-	-3.947 (2.034)
	Percent of HUs built before 1939	-	0.098 (0.072)	-	-	0.025 (0.127)	-	-	-0.349 (1.110)
	Percent HUs in urban areas	-	0.166*** (0.038)	-	-	-0.027 (0.087)	-	-	1.757* (0.773)
R Squared		0.372	0.694	0.066	0.082	0.095	NA	NA	NA
AIC		-313.07	-413.49	-267.4	-260.0	-256.3	637.86	599.86	597.45

**** p < 0.001; ** p < 0.01; * p < 0.05, ' p < 0.1*