Using Census Tract Demographics as Proxies for Individual Demographics in Analysis Using the Knowledge Networks Internet Panel: Draft Report

December 18, 2007

Jason Bell [1]

Note: This is a preliminary report and subject to modification pending further analysis.

Preamble

This report is the result of research conducted through direct funding from NCEE.

In our previous work on the valuation of water quality (EPA Cooperative Agreement CR823604 and Grant R827423, Economics of Environmental Improvement), we used survey data drawn from the Knowledge Networks panel. As part of that effort, we undertook several validity tests for the panel data. Did time in the panel affect survey responses? Was there sample selection bias in terms of which members of the panel took the survey? Did the time between offering the survey to the panel member and their completion of the survey affect their responses? Did subsequent panel attrition for the respondent influence valuations? On all these matters, the KN panel performed in a satisfactory manner. In the current research, we address the question of whether the panel itself is representative of the behavior and the demographic characteristics of its Census tract, and consequently of the U.S. population more generally.

This research had two primary goals.

The first goal was to address concerns about the use of Internet panels for surveys as a substitute for recruiting survey respondents using random digit dialing (we did not address other survey modes such as convenience samples or door-to-door surveys, which have sample bias or cost issues). Random digit dialing has had increasing difficulties achieving acceptable response rates, while the use of Internet panels has promise for significant savings in cost and time. This research addresses the question of whether Internet panel research produces biased results that might invalidate their other benefits.

The second goal was to establish a method to apply previous survey results to situations where individual demographic data are scarce by projecting results to broader samples. Earlier survey research conducted for EPA estimated values for recreational water quality improvements, and found that these values are sensitive to a range of demographic characteristics. That research had substantial demographic information on the respondents. The current research tests whether the same results would have been obtained had Census tract level data substituted for individual data. Thus, for the survey values of water quality that we obtained, would the regression equations be markedly different using the tract characteristics rather than the individual characteristics? In addition, using the regression equation for the original KN sample and

[1] This report uses data and some analysis from a previous EPA grant conducted with the assistance of W. Kip Viscusi and Joel Huber.

projecting the water quality values using the tract characteristics as variables, are the results similar to those obtained using the survey respondent values to estimate the water quality values? These sets of results will pertain to the ability to use KN survey results to project reliably the national benefit values.

Implications of results

The results of this research showed that individual demographics from a previously conducted survey and tract demographic characteristics from census tracts are highly correlated. KN sample members are representative of the tracts in which they live. This correlation requires that that surveyed individuals have a diverse set of demographic characteristics so that they can represent their tracts in an unbiased way. This pattern was supported by the similarity between sample enumerations using individual demographic values and tract values. These results support using Census tract demographics to estimate survey values for populations where individual demographics are not available (Goal 2).

There are several points where a person recruited for a survey panel can drop out. We call each of these points attrition levels. A person originally called with random digit dialing may never answer or may decline to be recruited, the person may fail to complete the panel demographic profile, the person may stop taking surveys after a period of time, the person may not have been selected from the panel for a particular survey, and the person may have chosen not to complete a particular survey. Tract demographic characteristics show small but statistically significant differences at each of these attrition points. However, these differences had negligible effects on estimates of survey values when included in a sample enumeration of regressions of survey values. These results give confidence that this survey did not experience serious sample bias due to the Internet panel survey mode (Goal 1). Even so, the differences that do exist between panel attrition levels call for caution wherever panels are used in surveys that estimate values that are correlated with underlying reasons for panel attrition.

Further work under direct NCEE funding will include examination of response differences associated with different survey administration modes, including panel-based internet, central location, mall-intercept, and phone-mail. The water quality project fielded surveys in each of these modes, and differences in responses and demographic characteristics between those surveys will serve as the units of comparison.

Disclaimer

This report has neither been reviewed nor approved by the U.S. Environmental Protection Agency. The views expressed by the author are entirely his own. The contents do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Introduction

Census data are available on a variety of demographic characteristics at the tract level. If a representative survey sample corresponds to their census tract demographics, results from a survey could be more easily and confidently generalized to the United States overall. Also, if survey respondents can be shown to be generally representative of their census tracts, survey results with incomplete demographic characteristics could be assigned their census tract average as a rough proxy for unknown individual characteristics.

This report examines the extent of such relationships between individual demographic characteristics and census tract demographics for a survey conducted using the Knowledge Networks Internet survey panel.

In addition, panel attrition within the Knowledge Networks survey panel is compared to census tract demographics to determine whether there are systematic factors associated with the various levels of survey panel membership. Substantial similarity across attrition levels would give confidence in using panels for surveys as a substitute for random digit dialing survey recruiting.

This report expands on previous work conducted under EPA Cooperative Agreement CR823604 and Grant R827423, Economics of Environmental Improvement. That work produced detailed estimates of the value of recreational water quality improvement based upon individual demographic characteristics using regression analysis of a large, nationally representative survey. This report demonstrates how those survey results can be projected to scenarios where individual demographics of affected populations are not available, using census tract demographics.

Summary

- Individual demographics of the sample are strongly correlated with tract demographic averages, implying that information at census tract level can be used to generate aggregate value estimates for goods estimated on an individual basis in the survey.

- Replacing individual demographics with tract averages in regression models produces quite similar estimates but larger standard errors around regression coefficients. This demonstrates the robustness of the water quality survey data that they can be used to estimate improvement values in policy scenarios without the need to collect individual demographic characteristics on populations affected by changes in water quality.

- There is substantial panel attrition that could be problematic for variables that are correlated with attrition factors. However, tract demographic averages predict almost identical (less than 3% difference) water quality values at each level of panel attrition, suggesting that the role of tract demographics in panel attrition does not affect the water quality survey results.

- Models of demographic outcomes estimated using tract demographic characteristics are quite similar across panel attrition levels. This is an encouraging sign that panel attrition does not occur in ways that make tracts in different attrition levels fundamentally different.

- Additional data from Knowledge Networks will allow analysis of whether individual demographic characteristics affect panel attrition differently than tract demographics, and whether those effects have any impact on water quality value estimates. Confirming the results found using tract data with individual demographics will support the assumption made in this report, that surveyed individuals can be considered as representatives of their census tract.

Our results show that tract-level demographic characteristics have substantial correlation with the individual demographics of the KN sample drawn for our EPA study of water quality values. Though replacing sample variables with the tract demographics does not produce statistically significant coefficients for tract level demographic variables in a regression model of water quality value, sample enumeration using tract data with an individual regression model's coefficients produces predicted values quite close to those generated using the individual data. This suggests that survey results might usefully be projected to a population where only average tract demographics are known, using the estimated coefficients based on the survey respondents.

However, estimates that can be projected from a survey sample to a broader context are only useful if the survey sample is not biased. Many tract-level demographic characteristics significantly predict a person's likelihood to reach each level of inclusion in Knowledge Networks panel, in both full population probits and descending population probits. These differences could affect the soundness of survey results themselves, in addition to interfering with the use of respondents as track representatives.

Censored regression analysis shows that the effects of these differences on the water quality values in our survey are very small (less than 3% for water quality values). In addition, Heckman selection models for both regional and national water quality values do not produce statistically significant sample selection effects, which gives us further confidence that the selection effects as measured using tract demographics do not significantly bias water quality survey results. Furthermore, models predicting tract demographic outcomes using tract demographic averages are substantially similar across all attrition levels.

This report is separated into the following sections.

1. We begin by examining how individual demographics collected for a water quality survey are correlated with tract demographic averages. This relationship between individual and tract demographics is important if tract demographics are to be used as proxies for individual characteristics.

2. Next, the report constructs regression models estimating key variables from a water quality survey using individual demographics, then tract averages as independent variables. Sample enumeration of those key variables demonstrate how tract averages perform when used as proxies for individual demographic characteristics.

3. Next, we examine the Knowledge Networks panel to determine how tract demographic averages are related to panel attrition, and we consider the effects of that attrition on

survey results. Serious differences here could call the panel into question as a tool for conducting unbiased surveys.

4. Finally, the report models tract demographic outcomes across attrition levels using tract demographic averages. Substantially similar results at each attrition level will provide reassurance that tracts at different attrition levels are not fundamentally different.

5. The report concludes with a summary of findings.

Section 1: Correlations Between Individual and Tract Demographics

Our results show that tract-level demographic characteristics have substantial correlation to individual demographics that were collected from a sample drawn for our EPA study of water quality values.

Overall, of the nine demographic shown below, only one was not correlated with its tract percentage (Female). The other eight had significant positive correlations between individual demographics and the demographics of that individual's tract (Age, Black, White, Hispanic, Household Size, Income, Education, and Urban Setting).

Table 1: Comparisons Between Individual Demographics and Tract Averages

Demographic	Individual Mean	Standard Deviation	Tract Mean	Standard Deviation
Age	44.69 years	16.99	45.28 years	4.73
Black	13.48 %	0.3415	12.80 %	0.2246
White	80.16 %	0.3988	76.93 %	0.2505
Hispanic	10.48 %	0.3064	10.38 %	0.1721
Female	49.13 %	0.5000	52.16 %	0.0348
Household Size	2.66 persons	1.38	2.62 persons	0.45
Income	$ 51,002	36,005	$ 53,937	21,559
Years of Education	13.22 years	2.58	13.17 years	1.17
MSA Status (Urban)	83.15 %	0.3717	82.45 %	0.3856

Note: These correlations reflect a sample of 3,806 respondents for whom individual and tract demographics were available.

The implication of these results is greater confidence in the use of individual tract residents as representatives of their tract.

Section 2: Tract Demographics as Proxies for Individual Demographics in Regressions

However, replacing individual demographics with tract demographics leads to certain differences when replicating analysis done for the survey from which the individual data were collected. To some extent, these differences are to be expected since individual data allow for more information and individual variation than using aggregated variables.

As the table below shows, while individual demographics are effective for predicting the value of the analysis variable (log of the value of improvement to regional water quality), tract demographics are less so (the first eight variables are changed between equations). However, it is notable that most parameter estimates retain the same sign and similar magnitude between the two equations.

Table 2: Censored Regression of Regional Water Quality Value Using Individual vs. Tract Demographics

Log of Value of Water Quality Benefits	Individual			Tract	
	Parameter Estimate	Standard Error		Parameter Estimate	Standard Error
Log (income)	0.1231 ***	0.0241		0.2292 *	0.1282
Years of education	0.0469 ***	0.0084		0.0144	0.0400
Age	0.0061 ***	0.0013		-0.0033	0.0061
Race: black	-0.2311 ***	0.0620		-0.1724	0.1142
Hispanic	0.0811	0.0687		0.1002	0.1670
Gender: female	-0.0265	0.0413		-0.0952	0.7384
Household size	-0.0371 **	0.0162		-0.1352 *	0.0789
Metropolitan Area	-0.0518	0.0568		-0.0330	0.0610
Region: Northeast	0.0071	0.0644		0.0200	0.0660
Region: South	-0.0742	0.0582		-0.0603	0.0610
Region: West	-0.0489	0.0630		-0.0056	0.0690
State lake quality	0.0005	0.0008		0.0003	0.0008
Lake acres per State square mile	0.0045 **	0.0023		0.0050 **	0.0023
Intercept	0.5838 **	0.2680		0.5038	0.9398

Notes: * significant at .10 level, ** significant at .05 level, *** significant at .01 level, all two-tailed tests.
Results use 3,805 observations, including 359 left censored and 371 right censored.

A similar model for national water quality value shows similar effects between individual and tract demographics as the model for regional value..

Table 3: Censored Regression of National Water Quality Value Using Individual vs. Tract Demographics

Log (National Value)	Individual		Tract

	Parameter Estimate	Standard Error		Parameter Estimate	Standard Error
Log (income)	0.0668 ***	0.0208		0.0082	0.1136
Years of education	0.0288 ***	0.0075		0.0566	0.0353
Age	0.0033 ***	0.0012		0.0071	0.0055
Race: black	-0.1060 *	0.0551		0.0365	0.1012
Hispanic	0.0452	0.0618		0.2056	0.1483
Gender: female	-0.0007	0.0370		-0.9025	0.6667
Household size	-0.0351 **	0.0144		-0.0623	0.0695
Metropolitan Area	-0.0252	0.0504		-0.0282	0.0540
Region: Northeast	0.1158 **	0.0581		0.1279 **	0.0592
Region: South	-0.0381	0.0526		-0.0336	0.0548
Region: West	-0.0908	0.0568		-0.0784	0.0620
State lake quality	0.0008	0.0007		0.0007	0.0008
Lake acres per State square mile	0.0043 **	0.0021		0.0039 *	0.0022
Intercept	1.6898 ***	0.2322		2.2754 ***	0.8388

Notes: * significant at .10 level, ** significant at .05 level, *** significant at .01 level, all two-tailed tests.
Results use 2,732 observations, including 149 left censored and 127 right censored.

For another variable collected in our survey, trips the respondent has taken to lakes or rivers in the last 12 months, individual demographic characteristics can be used to model its value. Unlike the previous variables, however, tract demographics do significantly predict the number of such trips. Additionally, several demographic variables differ in sign or magnitude between the individual and tract models for number of trips.

Table 4: Regression of Trips to Lakes or Rivers Using Individual vs. Tract Demographics

Number of Trips to Lakes or Rivers in Region, Last 12 Months	Individual			Tract	
	Parameter Estimate	Standard Error		Parameter Estimate	Standard Error
Log (income)	0.2047 ***	0.0668		-0.6122 *	0.3407
Years of education	0.0718 ***	0.0232		0.0580	0.1063
Age	-0.0022	0.0036		0.0143	0.0163
Race: black	-1.4966 ***	0.1718		-1.6999 ***	0.3062
Hispanic	-0.8934 ***	0.1908		-3.6671 ***	0.4425
Gender: female	-0.2995 ***	0.1147		-3.1231	1.9478
Household size	0.0179	0.0449		0.4207 **	0.2096
Metropolitan Area	-1.1042 ***	0.1574		-0.8174 ***	0.1637
Region: Northeast	-1.1682 ***	0.1790		-0.9312 ***	0.1771
Region: South	-1.1752 ***	0.1618		-0.9133 ***	0.1639
Region: West	-1.2809 ***	0.1750		-0.8903 ***	0.1839

State lake quality	0.0027	0.0023		0.0005	0.0022
Lake acres per State square mile	0.0230 ***	0.0063		0.0229 ***	0.0063
Intercept	2.2391 ***	0.7434		10.8316 ***	2.4967

Notes: * significant at .10 level, ** significant at .05 level, *** significant at .01 level, all two-tailed tests.
Results use 3,797 observations.

Most of the significant individual demographic characteristics are also significant in the tract level model. But many of the parameter estimates differ in sign (particularly income) or magnitude from their counterparts in the individual demographics model. Hispanic is over three times greater in the tract model, household size is significant and much larger in the tract model, and the effect of gender is starkly larger in the tract model.

Sample Enumeration Using Individual Model Coefficients

Using sample enumeration to reproduce the models predicted using individual demographics by instead using tract level demographic data as proxies leads to a value that is remarkably close to the results from the individual regression.

The sample enumeration from the individual demographics regression (the sum of each person's individual demographics multiplied by the parameter estimate for that demographic variable from the individual regression model) leads to an average value of $32.29.

Using tract demographics as proxies in the individual regression (tract demographics multiplied by the parameter estimates for the corresponding demographic variable in the individual regression model) leads to an average value of $32.98, a difference of less than 3% from the individual estimate.

Table 5: Predicted Water Quality Values Using Individual Regression, Tract Averages as Proxies, and Tract Averages as Proxies Projected to All United States Census Tracts

Sample Enumeration	Individual Demographics	Tract Demographics	All Tracts in US
Mean	$32.29	$32.98	$33.46
Median	$14.13	$14.67	$13.82

This result is interesting because it suggests that values estimated using individual demographics could be generalized to situations where less precise data are available (such as where only tract or population demographics are known). Making the proxy calculation for the almost 65,000 tracts for which we have full tract demographics, the average calculated value for water quality improvements was $33.46, weighted by population of the tract.

The same trends found in the regional water quality values above are repeated for the analysis variable of national water quality values, with less than 2% difference in national values between the individual model and tract demographics used as proxies.

Sample enumeration using tract demographics as proxies also provides close approximations to the individual model for number of trips to lakes or rivers in the past 12 months.

Table 6: Predicted Trips Using Individual Regression, Tract Averages as Proxies, and Tract Averages as Proxies Projected to All United States Census Tracts

Sample Enumeration	Individual Demographics	Tract Demographics	All Tracts in US
Mean	3.42 trips	3.45 trips	3.53 trips
Median	3.36 trips	3.28 trips	3.26 trips

Section 3: Panel Attrition and Tract Demographics

Probit Models Estimating Tract Level Demographics at Each Level of KN Panel Attrition

We estimated probit models of the likelihood that a person who Knowledge Networks attempted to contact (based upon the working residential phone numbers KN called) was *recruited* (reached and recruited via telephone), was *profiled* (whether KN sent a core profile survey preceding panel membership), are still *active* in the KN panel, was in our *sample* (whether the person was drawn to be a respondent in our water quality survey), and was in our *final* data (whether the person completed the survey). The independent variables for these models consisted of tract level demographic characteristics.[2]

Table 7: Number of Observations in Each Subset.

	Total	Recruit	Profile	Active	Sample	Final
N	1,159,220	260,678	125,549	40,060	5,276	4,016

The probit models were done first with the full population of persons KN attempted to contact.

These results show that census tract demographics significantly predict whether a person reaches a particular attrition level. Income had a significant negative coefficient in each probit, suggesting that more wealthy people are less likely to join and remain in the panel. Education had a positive effect for recruitment and profile collection, suggesting that educated people were more interested in the idea of taking surveys than others. The effect was negative for inclusion in our sample and completion, suggesting this interest might not translate as well into participation. Age had a negative coefficient when significant, suggesting that younger people are more likely to join and participate in the survey panel than older people (perhaps reflecting greater comfort with the technology). Black and Hispanic people were less likely to be retained at every attrition level. Women were more likely to be successfully recruited and profiled, though this effect's significance was lost for the other attrition levels. People with larger household sizes were more likely to be retained at every level. This might indicate a problem if multiple household members (with different demographic characteristics) shared the survey-taking responsibilities. Geographically, every region was significantly less likely to be recruited, profiled, and remain active (compared to the Midwest). Finally, those in states with higher water quality had a very slight positive coefficient for recruitment, remaining active in the panel, and inclusion in our sample. Finally, lake density in the panelist's state led to higher recruitment and profile coefficient, but lower coefficients for inclusion in the sample and survey completion.

Overall, these results indicate that care should be taken before assuming that a panelist at a given attrition level is as random a representative of that panelist's census tract as when that panelist was first contacted.

[2] A similar analysis using census factors was conducted by Cameron and DeShazo: "Sample Selection in a Major Consumer Panel: Assessment and Correction Using Year 2000 Census Tract Characteristics and County-Level Presidential Voting Patterns" draft, June 4, 2005.

Table 8: Full Probits (only coefficients and significance shown)

	Recruit	Profile	Active	Sample	Final
N	1,159,220	1,159,220	1,159,220	1,159,220	1,159,220
Log (income)	-0.1549 ***	-0.1750 ***	-0.1418 ***	-0.1674 ***	-0.1622 ***
Years of education	0.0227 ***	0.0228 ***	0.0010	-0.0253 ***	-0.0227 **
Age	-0.0056 ***	-0.0069 ***	0.0004	-0.0031 **	-0.0029 *
Race: black	-0.1192 ***	-0.3714 ***	-0.3479 ***	-0.1131 ***	-0.1665 ***
Hispanic	-0.5065 ***	-0.5932 ***	-0.6107 ***	-0.2585 ***	-0.2980 ***
Gender: female	0.5424 ***	0.4224 ***	-0.0327	0.1663	0.1566
Household size	0.2121 ***	0.1552 ***	0.1018 ***	0.0416 **	0.0318 *
Metropolitan Area	0.0884 ***	0.1125 ***	0.0094	-0.0109	0.0015
Region: Northeast	-0.1772 ***	-0.2113 ***	-0.1508 ***	-0.0017	-0.0116
Region: South	-0.1109 ***	-0.1193 ***	-0.1128 ***	-0.0094	-0.0145
Region: West	-0.0782 ***	-0.1044 ***	-0.0961 ***	-0.0118	-0.0186
State lake quality	0.0001 **	2.55e-5	0.0004 ***	0.0006 ***	0.0005 **
Lake acres per State square mile	0.0012 ***	0.0009 ***	-0.0002	-0.0011 **	-0.0013 **
Intercept	0.1104 **	0.1564 **	-0.3948 ***	-0.4660 **	-0.6164 ***
Pseudo R^2	0.0061	0.0085	0.0085	0.0051	0.0047

The analysis was repeated with subsets of each descending population (a model for recruited was done on the full sample, profiled done on the recruited subset, active done on the profiled subset, sample done on the active subset, and final done on the sample subset).

This analysis gives a better picture of what demographic characteristics are significant between any two attrition levels. Those with higher incomes appear to be less willing to be recruited and profiled, but further retention in the panel is not significantly different once they are members. Higher educated people are more likely to agree to be recruited and profiled, but are less likely to remain active or be part of our sample. Older panelists are less likely to be recruited or profiled, and to be part of our sample, but are more likely to remain active panel members. Black and Hispanic panelists have significant attrition at every level except for inclusion in our sample. Large households are more likely to be recruited, but less likely to be profiled, remain active, and inclusion in our sample. Geographically, urban areas are more likely to be recruited and profiled, but less likely to remain active. All regions (compared to the Midwest) are less likely to be recruited, profiled, and remain active (except the Northeast), but more likely to be in our sample.

This analysis also indicates that caution should be used when assuming that a panelist who has been through several attrition levels can be considered a representative of that panelist's census tract.

Table 9: Descending Probits (only coefficients and significance shown)

	Recruit	Profile	Active	Sample	Final
N	1,159,220	260,678	125,549	40,060	5,276
Log (income)	-0.1549 ***	-0.1101 ***	-0.0098	-0.1519 ***	0.0112
Years of education	0.0227 ***	0.0128 ***	-0.0337 ***	-0.0436 ***	0.0218
Age	-0.0056 ***	-0.0049 ***	0.0119 ***	-0.0052 **	0.0011
Race: black	-0.1192 ***	-0.5341 ***	-0.1129 ***	0.2789 ***	-0.3625 ***
Hispanic	-0.5065 ***	-0.3931 ***	-0.2837 ***	0.3992 ***	-0.2973 **
Gender: female	0.5424 ***	0.0238	-0.6988 ***	0.2254	-0.2780
Household size	0.2121 ***	-0.0265 ***	-0.0411 ***	-0.0548 *	-0.0632
Metropolitan Area	0.0884 ***	0.0879 ***	-0.1622 ***	-0.0188	0.0798
Region: Northeast	-0.1772 ***	-0.1534 ***	0.0184	0.1953 ***	-0.0788
Region: South	-0.1109 ***	-0.0765 ***	-0.0452 ***	0.1265 ***	-0.0494
Region: West	-0.0782 ***	-0.0950 ***	-0.0407 ***	0.0943 ***	-0.0628
State lake quality	0.0001 **	-0.0002	0.0008 ***	0.0005 *	-0.0004
Lake acres per State square mile	0.0012 ***	-0.0002	-0.0017 ***	-0.0014	-0.0020
Intercept	0.1104 **	1.3832 ***	0.1812	1.2106 ***	0.6634
Pseudo R^2	0.0061	0. 0090	0. 0076	0. 0156	0. 0089

As the results above show, there are significant demographic differences between the subsets and the total population of contact attempts, as well as between each subsequent attrition level. Since systematic differences seem to be present, it is important to determine if such differences affect survey results, in this case, for our water quality survey.

Sample Enumeration Applied to Tracts at Each Attrition Level

The next step is to use the sample enumeration described in Section 2, applied to the tract demographics model, to predict the water quality value within the final survey participants, then each of the other Knowledge Networks contact sub-samples.

This test shows only very small differences between estimated water quality values:

Table 10: Predicted Water Quality Values for Each Knowledge Networks Attrition Level

Sub-sample:	Mean, Regional Water Quality Value	Standard Deviation
Survey Respondents (Final)	$ 33.00	$ 4.60
Survey Invitees (Sample)	$ 32.84	$ 4.62
Active	$ 33.84	$ 4.64

Profiled	$ 33.72	$ 4.75
Recruited	$ 33.46	$ 4.95
All Data	$ 33.62	$ 5.11

These results are encouraging in that panel effects do not seem to have large effects on our estimation of water quality value (less than 3% difference between largest and smallest estimates). This indicates that though there are significant demographic factors that affect panel attrition, those factors do not translate into large differences in our survey results.

Heckman Selection Model

For additional confidence in sample selection effects on survey results, we estimated a Heckman selection model on a non-censored version of our water quality value regression model, with a selection model that includes the census tract demographics.

The regression model looks much like the censored regression results from our previous research (as expected). As for the selection model, while several of the tract demographics are significant (similar to the results we found in the Probits earlier in this section), the model itself is not, so we cannot reject the null hypothesis that there is no selection bias with this analysis.

Table 11: Heckman Selection Model of Regional Water Quality Value using Census Demographics in the Selection Equation

Log of Regional Water Quality Value	Coefficient	Standard Error
Regression Equation:		
Log (income)	0.0981 ***	0.0201
Years of education	0.0397 ***	0.0069
Age	0.0056 ***	0.0011
Race: black	-0.1998 ***	0.0511
Hispanic	0.0506	0.0571
Gender: female	-0.0303	0.0339
Household size	-0.0285 **	0.0133
Metropolitan Area	-0.0227	0.0475
Region: Northeast	0.0026	0.0530
Region: South	-0.0564	0.0480
Region: West	-0.0560	0.0526
State lake quality	0.0003	0.0007
Lake acres per State square mile	0.0032 *	0.0019
Intercept	0.6135	0.7421
Selection equation		
Log (income) (Tract Average)	-0.1660 ***	0.0320
Years of education (Tract Average)	-0.0222 **	0.0102
Age (Tract Average)	-0.0034 **	0.0015
Race: black (Tract Average)	-0.1721 ***	0.0291
Hispanic (Tract Average)	-0.3481 ***	0.0404
Gender: female (Tract Average)	0.1819	0.1808
Household size (Tract Average)	0.0351 *	0.0193
Metropolitan Area (Tract Value)	-0.0040	0.0159
Intercept	-0.5899 ***	0.2291
/athrho	0.1021	0.2468
/lnsigma	0.0457 *	0.0257
rho	0.1018	0.2443
sigma	1.0467	0.0269
lambda	0.1065	0.2581

LR test of indep. eqns. (rho = 0): chi2(1) = 0.16 **Prob > chi2 = 0.6883**

Notes: * significant at .10 level, ** significant at .05 level, *** significant at .01 level, all two-tailed tests.

Section 4: Modeling Tract Demographic Outcomes Using Tract Demographics

The previous analyses demonstrated that survey estimates differ very little when tract values are used as proxies for individual demographic data. We also showed that survey estimates using tract values remain consistent across survey attrition levels.

We next examine how attrition affects the demographic aspects of the census tracts in which panel contacts live. Aside from examining mean values of tract demographics at each of the different attrition levels, we can model how demographic characteristics of tract residents affect demographic outcomes (income, education, marital status, home ownership, and household size) of people in those tracts.

If the relationship between demographics and outcomes differ greatly between attrition levels, that might indicate that attrition of the survey panel is occurring in a systematic way.

Tract Means

Mean values are presented to give context to the regression models using these variables that follow. Setting aside whether the differences are significant, variations between the attrition levels can be instructive.

The tract averages of most of these variables are very similar across each attrition level. Attrition appears to be more common in tracts with higher percentage of blacks, and less common in tracts with higher percentage of whites. Tracts with higher Hispanic percentage experience more attrition. Tracts with higher home ownership rates are more likely to be retained in the panel. Tracts in the Midwest experience the least attrition among regions. However, even these differences are quite modest. The difference between the largest and smallest average incomes is less than $5,000, average education differs less than one third of a year, even the average ages differ by less than half a year across all attrition levels.

Since the key attrition levels are the difference between all contacts and the active panel (since the active panel is the pool from which survey respondents are drawn, and all contacts is the pool from which a random digit dial sample would be drawn), these differences warrant particular attention. Most demographic characteristics are quite similar, though racial percentages (12.78% Black vs. 10.36%, and a corresponding percentage point difference for White) and ethnicity (11.63% vs. 8.76% Hispanic) have differences that are large enough to cause concern. Those remaining in the panel are slightly more likely to be from tracts with higher percentages married, born in their home state, and homeowners. This is not surprising, as all three of these characteristics may be associated with greater stability, and thus a greater ability to keep up with survey obligations of panel membership. Geographically, this panel seems to have more trouble retaining panelists from every region except the Midwest.

Table 12: Mean Values of Tract Demographics Across Attrition Levels

	All Contacts	Recruit	Profile	Active	Sample	Final
N	1,159,220	260,666	125,549	39,062	5,311	4,016

Log (income)	10.9020	10.9001	10.9084	10.8982	10.8191	10.8291
Years of education	13.3526	13.3436	13.3961	13.3695	13.1213	13.1644
Age	45.3698	45.1541	45.1816	45.5645	45.1716	45.2491
Race: black	0.1278	0.1301	0.1114	0.1036	0.1366	0.1277
Race: white	0.7579	0.7604	0.7858	0.8038	0.7563	0.7678
Hispanic	0.1163	0.1084	0.0993	0.0876	0.1121	0.1063
Gender: female	0.5214	0.5216	0.5207	0.5206	0.5216	0.5214
Household size	2.6294	2.6574	2.6356	2.6093	2.6326	2.6186
Metropolitan Area	0.8573	0.8656	0.8660	0.8316	0.8234	0.8252
Born in Home State	0.5855	0.5966	0.5988	0.6084	0.6070	0.6067
Never Married	0.2735	0.2716	0.2673	0.2608	0.2748	0.2725
Married	0.5089	0.5144	0.5213	0.5272	0.5030	0.5075
Home Ownership	0.6828	0.6964	0.7042	0.7168	0.6793	0.6851
Region: Northeast	0.1923	0.1711	0.1615	0.1681	0.1836	0.1830
Region: South	0.3567	0.3450	0.3352	0.3270	0.3668	0.3608
Region: West	0.2301	0.2305	0.2267	0.2148	0.2165	0.2149
Region: Midwest	0.2209	0.2534	0.2767	0.2901	0.2331	0.2413

The question to be explored next relates to how demographics of tract residents affect tract demographic outcomes. For instance, the outcome demographic of household income may be related to education, age, race, and region of the country.

Ideally, one would expect the effect of demographics on outcomes to be consistent across attrition levels. If differences in the effects of tract demographic characteristics are apparent, that could demonstrate that the tracts of panelists who are not retained are different than those who stay in the panel in important ways.

Where such differences are identified, they must be tested for whether and how those differences can affect survey results. If no important differences are found, this would support the idea that a survey administered using a panel developed from random digit dialing, such as the Knowledge Networks Internet panel, is an appropriate substitute for a survey administered to a sample recruited using random digit dialing.

Log (Income)

The first tract demographic outcome we will model is household income.

The model using all of Knowledge Networks random digit dial contacts was significant, with a high r-square (0.8341). All of the mean tract demographic characteristics used as independent variables were significant at the .01 level (in fact, the independent variables would also meet a significance test of .001). These characteristics are also true of the models using those successfully recruited by KN, as well as those who completed KN's profile demographics. The model using active members of the panel loses significance for one independent variable (percentage of people born in the state where they currently live), while the model using the water survey sample loses significance for one additional independent variable (Hispanic percentage). The final survey group looks much like the sample group, though with one additional independent variable falling to the .05 significance level (Region: South).

Table 13: Regression Models of Tract Log(Income) as a Function of Tract Demographics Across Attrition Levels

Log (Income)	All Contacts	Recruit	Profile	Active	Sample	Final
N	1,159,220	260,666	125,549	39,062	5,311	4,016
Years of education	0.2614 ***	0.2533 ***	0.2507 ***	0.2500 ***	0.2424 ***	0.2435 ***
Age	0.0111 ***	0.0114 ***	0.0111 ***	0.0102 ***	0.0096 ***	0.0095 ***
Race: black	-0.2099 ***	-0.2274 ***	-0.2482 ***	-0.2836 ***	-0.2372 ***	-0.2286 ***
Race: white	-0.1991 ***	-0.2219 ***	-0.2409 ***	-0.2545 ***	-0.2487 ***	-0.2472 ***
Hispanic	0.0472 ***	0.0374 ***	0.0377 ***	0.0285 ***	0.0102	0.0020
Gender: female	-1.2455 ***	-1.1802 ***	-1.1344 ***	-1.0923 ***	-0.9280 ***	-0.9269 ***
Household size	0.2715 ***	0.2672 ***	0.2642 ***	0.2683 ***	0.2635 ***	0.2669 ***
Metropolitan Area	0.0888 ***	0.0952 ***	0.0980 ***	0.0982 ***	0.0848 ***	0.0838 ***
Born in Home State	-0.0199 ***	-0.0247 ***	-0.0087 ***	0.0040	-0.0231	-0.0072
Never Married	-0.3241 ***	-0.3654 ***	-0.3498 ***	-0.2596 ***	-0.4641 ***	-0.4639 ***
Home Ownership	0.2751 ***	0.2962 ***	0.3082 ***	0.3186 ***	0.2882 ***	0.2688 ***
Region: Northeast	0.0534 ***	0.0463 ***	0.0410 ***	0.0344 ***	0.0441 ***	0.0536 ***
Region: South	-0.0174 ***	-0.0234 ***	-0.0243 ***	-0.0154 ***	-0.0221 ***	-0.0156 **
Region: West	-0.0271 ***	-0.0330 ***	-0.0349 ***	-0.0394 ***	-0.0368 ***	-0.0243 ***
Intercept	6.8527 ***	6.9360 ***	6.9616 ***	6.9536 ***	7.0985 ***	7.0765 ***
Adj. R-square	0.8341	0.8363	0.8302	0.8286	0.8047	0.8000

Overall, the models do a very good job of predicting tract household income using mean tract demographic characteristics (significant independent variables, high r-square). The low standard errors for most of the coefficients will be helpful when comparing the models across attrition levels.

The model demonstrates interesting features about tract household income. Tracts with higher education, larger household size, and higher mean age, and greater home ownership predict higher tract incomes, as one would expect. Metropolitan areas and, perhaps related, Hispanic percentage also predict higher income. Both white and black percentages predict lower incomes, suggesting that tracts with more homogenous racial characteristics tend to be poorer (if a squared

term is introduced, white becomes positive with the squared negative, while black becomes negative with the squared positive). Tracts with more females tend to have lower household incomes, though it is unclear whether this is due solely to historic wage differences by gender or related to tracts with retired populations (more widows). Tracts with more unmarried people have lower incomes, which is not surprising since unmarried people have fewer wage earners per household. Regional variables show the Northeast as having the highest income, followed by Midwest (excluded), South, and then West.

Because standard errors in these models were generally so small, most differences between attrition levels tend to fall outside of the 95% confidence interval. For instance, the lower 95% CI for the coefficient for years of education is 0.2611 for all contacts, while the upper 95% CI for the recruit sample is 0.2540. This means there is a small, though significant, change in how education affects household income between these two groups of tracts. So it is clear, with a high level of confidence, that there are significant differences between how tract characteristics affect tract income between attrition levels.

However, when taken together with the previous sections of this report which demonstrate only small effects of the demographic differences between attrition levels on survey results, the differences in the models between attrition levels does not cause a great deal of concern.

Years of Education

The models predicting tract income as an outcome of tract demographics across attrition levels can be repeated for other tract demographic outcomes.

For years of education, the models are again similar across attrition levels. For all of these models, higher income, white percentage, female percentage, never married percentage, and home ownership predict more education in the tract. Higher age, black percentage, Hispanic percentage, household size, and born in home state predict less education in the tract.

However, there is one significant variable that changes signs between models. Urban tracts predict higher education in tracts for all contacts, but less education at every attrition level. Such a reversal of effects deserves attention, but the overall effect of metropolitan status is associated with very small differences in average years of education in a tract. The difference between the greatest and least effects corresponds to less than 0.06 of one year of education average for a tract.

Table 14: Regression Models of Tract Education as a Function of Tract Demographics Across Attrition Levels

Years of Education	All Contacts	Recruit	Profile	Active	Sample	Final
N	1,159,220	260,666	125,549	39,062	5,311	4,016
Log (Income)	2.4818 ***	2.5429 ***	2.5687 ***	2.6051 ***	2.4843 ***	2.4861 ***
Age	-0.0522 ***	-0.0524 ***	-0.0526 ***	-0.0498 ***	-0.0554 ***	-0.0557 ***
Race: black	-0.0637 ***	-0.0390 ***	-0.0312	0.0360	-0.0476	-0.1454
Race: white	0.3155 ***	0.3304 ***	0.3238 ***	0.3046 ***	0.3362 ***	0.2230 **
Hispanic	-0.9378 ***	-0.9566 ***	-0.9830 ***	-0.9610 ***	-0.8536 ***	-0.8629 ***
Gender: female	4.8124 ***	4.6881 ***	4.5766 ***	4.3824 ***	4.1191 ***	4.1527 ***
Household size	-1.0459 ***	-1.0305 ***	-1.0307 ***	-1.0470 ***	-1.0866 ***	-1.1144 ***
Metropolitan Area	0.0058 ***	-0.0215 ***	-0.0384 ***	-0.0491 ***	-0.0104	-0.0113
Born in Home State	-0.8292 ***	-0.8243 ***	-0.8691 ***	-0.8468 ***	-0.8249 ***	-0.8509 ***
Never Married	1.4620 ***	1.5231 ***	1.5215 ***	1.3340 ***	1.6803 ***	1.7069 ***
Home Ownership	0.3940 ***	0.3260 ***	0.3053 ***	0.2457 ***	0.4379 ***	0.5315 ***
Region: Northeast	-0.0472 ***	-0.0340 ***	-0.0179 ***	-0.0058	-0.0329	-0.0541 **
Region: South	-0.0136 ***	0.0022	0.0059	-0.0222 ***	-0.0058	-0.0391 *
Region: West	0.1139 ***	0.1278 ***	0.1302 ***	0.1421 ***	0.1306 ***	0.0875 ***
Intercept	-11.4183 ***	-12.0118 ***	-12.1662 ***	-12.4552 ***	-10.9638 ***	-10.8440 ***
Adj. R-square	0.8498	0.8387	0.8297	0.8262	0.8146	0.8132

Never Married Percentage

For the percentage of people in a tract who have never been married, the models are once again similar across attrition levels. Higher education, black percentage, Hispanic percentage, female percentage, metropolitan status, and percentage born in their home state are all associated with a higher percentage of people in the tract who have never been married. Higher income, age, white percentage, household size, and home ownership are all associated with a lower percentage of people who have never been married.

Table 15: Regression Models of Tract Non-Marital Status as a Function of Tract Demographics Across Attrition Levels

Never Married %	All Contacts	Recruit	Profile	Active	Sample	Final
N	1,159,220	260,666	125,549	39,062	5,311	4,016
Log (Income)	-0.0307 ***	-0.0362 ***	-0.0349 ***	-0.0250 ***	-0.0449 ***	-0.0443 ***
Years of education	0.0146 ***	0.0150 ***	0.0148 ***	0.0123 ***	0.0159 ***	0.0160 ***
Age	-0.0090 ***	-0.0090 ***	-0.0092 ***	-0.0090 ***	-0.0092 ***	-0.0096 ***
Race: black	0.0707 ***	0.0652 ***	0.0614 ***	0.0626 ***	0.0499 ***	0.0569 ***
Race: white	-0.0922 ***	-0.0946 ***	-0.0998 ***	-0.1008 ***	-0.1140 ***	-0.1008 ***
Hispanic	0.0722 ***	0.0640 ***	0.0718 ***	0.0722 ***	0.0608 ***	0.0637 ***
Gender: female	0.0149 ***	0.0242 ***	0.0162 ***	0.0374 ***	0.0443 *	0.1025 ***
Household size	-0.0527 ***	-0.0475 ***	-0.0548 ***	-0.0587 ***	-0.0486 ***	-0.0501 ***
Metropolitan Area	0.0089 ***	0.0100 ***	0.0095 ***	0.0080 ***	0.0112 ***	0.0097 ***
Born in Home State	0.0260 ***	0.0286 ***	0.0294 ***	0.0276 ***	0.0317 ***	0.0296 ***
Home Ownership	-0.1342 ***	-0.1387 ***	-0.1329 ***	-0.1362 ***	-0.1274 ***	-0.1311 ***
Region: Northeast	0.0152 ***	0.0142 ***	0.0152 ***	0.0152 ***	0.0176 ***	0.0170 ***
Region: South	-0.0340 ***	-0.0363 ***	-0.0364 ***	-0.0355 ***	-0.0331 ***	-0.0337 ***
Region: West	-0.0097 ***	-0.0113 ***	-0.0110 ***	-0.0124 ***	-0.0094 ***	-0.0101 ***
Intercept	1.0864 ***	1.1276 ***	1.1437 ***	1.0678 ***	1.2152 ***	1.1924 ***
Adj. R-square	0.7565	0.7529	0.7468	0.7484	0.7701	0.7703

Married Percentage

The percentage of people in a tract who are married is approximately the opposite of the never been married demographic, but the two demographics are not exhaustive, as they do not include widowed, separated, or divorced people. Because of this, not all independent demographic tract averages have opposite signs. Higher education and female percentage also predict higher percentage of married people, while the rest of the signs are opposite of what was found above.

Again, all models predict this percentage similarly, with consistent signs and magnitudes very close to one another.

Table 15: Regression Models of Tract Marital Status as a Function of Tract Demographics Across Attrition Levels

Married %	All Contacts	Recruit	Profile	Active	Sample	Final
N	1,159,220	260,666	125,549	39,062	5,311	4,016
Log (Income)	0.0398 ***	0.0469 ***	0.0443 ***	0.0388 ***	0.0550 ***	0.0566 ***
Years of education	0.0088 ***	0.0087 ***	0.0099 ***	0.0104 ***	0.0089 ***	0.0084 ***
Age	0.0040 ***	0.0042 ***	0.0046 ***	0.0049 ***	0.0046 ***	0.0051 ***
Race: black	-0.1579 ***	-0.1399 ***	-0.1341 ***	-0.1327 ***	-0.1088 ***	-0.1178 ***
Race: white	0.1073 ***	0.1207 ***	0.1275 ***	0.1324 ***	0.1522 ***	0.1403 ***
Hispanic	-0.1251 ***	-0.1120 ***	-0.1194 ***	-0.1195 ***	-0.0961 ***	-0.1019 ***
Gender: female	0.1535 ***	0.1406 ***	0.1095 ***	0.0897 ***	0.1092 ***	0.0310
Household size	0.1203 ***	0.1161 ***	0.1262 ***	0.1305 ***	0.1176 ***	0.1207 ***
Metropolitan Area	-0.0203 ***	-0.0201 ***	-0.0186 ***	-0.0171 ***	-0.0178 ***	-0.0163 ***
Born in Home State	-0.0247 ***	-0.0245 ***	-0.0246 ***	-0.0254 ***	-0.0254 ***	-0.0263 ***
Home Ownership	0.2083 ***	0.2116 ***	0.2043 ***	0.2114 ***	0.1992 ***	0.1979 ***
Region: Northeast	-0.0155 ***	-0.0146 ***	-0.0160 ***	-0.0167 ***	-0.0157 ***	-0.0160 ***
Region: South	0.0285 ***	0.0307 ***	0.0308 ***	0.0295 ***	0.0286 ***	0.0293 ***
Region: West	-0.0056 ***	-0.0032 ***	-0.0043 ***	-0.0027 ***	-0.0014	-0.0033
Intercept	-0.7851 ***	-0.8691 ***	-0.8865 ***	-0.8558 ***	-0.9834 ***	-0.9734 ***
Adj. R-square	0.8138	0.8143	0.8080	0.8046	0.8076	0.8075

Home Ownership Percentage

For the percentage of households in a tract who own their homes, the models are again similar across attrition levels. Higher income, education, age, black percentage, white percentage, household size, and percentage born in their home state all predict higher tract home ownership rates. Higher Hispanic percentage, female percentage, metropolitan status, and never married percentage predict lower home ownership percentage.

Table 16: Regression Models of Tract Home Ownership as a Function of Tract Demographics Across Attrition Levels

Home Ownership %	All Contacts	Recruit	Profile	Active	Sample	Final
N	1,159,220	260,666	125,549	39,062	5,311	4,016
Log (Income)	0.1154 ***	0.1280 ***	0.1308 ***	0.1315 ***	0.1343 ***	0.1195 ***
Years of education	0.0174 ***	0.0140 ***	0.0126 ***	0.0097 ***	0.0199 ***	0.0231 ***
Age	0.0109 ***	0.0106 ***	0.0109 ***	0.0106 ***	0.0119 ***	0.0120 ***
Race: black	0.3026 ***	0.2860 ***	0.2837 ***	0.3164 ***	0.3209 ***	0.3359 ***
Race: white	0.4248 ***	0.3990 ***	0.4092 ***	0.4327 ***	0.4146 ***	0.4371 ***
Hispanic	-0.1823 ***	-0.1725 ***	-0.1720 ***	-0.1735 ***	-0.1104 ***	-0.1113 ***
Gender: female	-0.3052 ***	-0.3087 ***	-0.2846 ***	-0.3020 ***	-0.4234 ***	-0.4628 ***
Household size	0.2221 ***	0.2086 ***	0.2113 ***	0.2129 ***	0.2023 ***	0.2136 ***
Metropolitan Area	-0.0102 ***	-0.0077 ***	-0.0062 ***	-0.0074 ***	-0.0074 *	-0.0061
Born in Home State	0.1662 ***	0.1632 ***	0.1556 ***	0.1248 ***	0.1825 ***	0.1673 ***
Never Married	-0.5956 ***	-0.6054 ***	-0.5652 ***	-0.5839 ***	-0.6130 ***	-0.6104 ***
Region: Northeast	-0.0457 ***	-0.0448 ***	-0.0451 ***	-0.0378 ***	-0.0372 ***	-0.0360 ***
Region: South	0.0173 ***	0.0143 ***	0.0134 ***	0.0066 ***	0.0117 ***	0.0100 **
Region: West	-0.0127 ***	-0.0150 ***	-0.0163 ***	-0.0187 ***	-0.0057	-0.0071
Intercept	-1.9856 ***	-1.9999 ***	-2.0598 ***	-1.9999 ***	-2.1629 ***	-2.0689 ***
Adj. R-square	0.7731	0.7632	0.7622	0.7578	0.7490	0.7559

Household Size

Finally, for the average household size in a tract, the models are consistent across attrition levels. Higher income, Hispanic percentage, female percentage, and home ownership rate all predict larger household size. Higher education, age, black percentage, white percentage, and percentage never married all predict lower household size. However, both metropolitan status and percentage of people born in their home state change signs across attrition levels, though metro status is not significant in the regression including all contacts (the only model with a positive sign for this independent variable). Of more concern is the percent born in home state, which has its largest difference and sign change between all contacts and active panel members. Again, as with the years of education models, the effect of this difference on predicted household size is very small.

Table 17: Regression Models of Tract Household Size as a Function of Tract Demographics Across Attrition Levels

Household Size	All Contacts	Recruit	Profile	Active	Sample	Final
N	1,159,220	260,666	125,549	39,062	5,311	4,016
Log (Income)	0.5109 ***	0.5283 ***	0.5060 ***	0.4910 ***	0.5305 ***	0.5058 ***
Years of education	-0.2073 ***	-0.2030 ***	-0.1927 ***	-0.1839 ***	-0.2134 ***	-0.2068 ***
Age	-0.0492 ***	-0.0487 ***	-0.0487 ***	-0.0478 ***	-0.0504 ***	-0.0498 ***
Race: black	-0.3045 ***	-0.2615 ***	-0.2393 ***	-0.2696 ***	-0.3205 ***	-0.3039 ***
Race: white	-0.8848 ***	-0.8462 ***	-0.8372 ***	-0.8677 ***	-0.8639 ***	-0.8460 ***
Hispanic	0.6774 ***	0.6688 ***	0.6756 ***	0.6720 ***	0.5930 ***	0.6009 ***
Gender: female	1.2494 ***	0.9800 ***	0.8495 ***	0.8486 ***	1.0580 ***	1.1508 ***
Metropolitan Area	0.0002	-0.0055 ***	-0.0109 ***	-0.0151 ***	-0.0039	-0.0016
Born in Home State	-0.0103 ***	-0.0088 ***	-0.0076 *	0.0238 ***	-0.0131	0.0019
Never Married	-1.0491 ***	-0.9495 ***	-1.0514 ***	-1.1158 ***	-1.0104 ***	-0.9926 ***
Home Ownership	0.9959 ***	0.9543 ***	0.9534 ***	0.9446 ***	0.8735 ***	0.9100 ***
Region: Northeast	0.0631 ***	0.0638 ***	0.0722 ***	0.0704 ***	0.0551 ***	0.0492 ***
Region: South	-0.0825 ***	-0.0782 ***	-0.0739 ***	-0.0677 ***	-0.0727 ***	-0.0731 ***
Region: West	0.0538 ***	0.0602 ***	0.0647 ***	0.0629 ***	0.0616 ***	0.0542 ***
Intercept	1.6578 ***	1.5057 ***	1.6940 ***	1.7290 ***	1.7482 ***	1.7939 ***
Adj. R-square	0.7823	0.7625	0.7588	0.7600	0.7658	0.7634

Conclusions

This report has conducted several analyses to determine how tract demographic data can be used to apply survey results to a broader context and to test a survey panel for attrition that might affect whether samples drawn from such a panel are biased.

Overall, results were quite encouraging.

- Significant correlation between survey respondents and their home census tracts gives confidence that survey respondents can be considered as representatives of their census tract.

- Regression analyses of individual and tract demographics produce coefficients with similar sign and magnitude.

- Sample enumeration using coefficients from individual regression models of three survey values produces nearly identical results whether individual or tract demographic variables are used.

- Panel attrition reveals significant differences between panel attrition levels. However, sample enumeration using coefficients from the tract regression model for water quality values produces nearly identical results at every attrition level.

- Models of tract demographic outcomes based upon tract demographics are quite similar at every panel attrition level.

Further research will use full individual demographics from the Knowledge Networks survey panel data to replicate much of the above attrition analysis using individual demographics instead of tract factors. This could provide confirmation of previous analyses and test the validity of the assumption that survey respondents are representatives of their tracts. Viscusi, Huber, and Bell are currently operating under an EPA grant that will examine these individual demographic data and conduct this research, as well as other tests of panels for surveys (PI-83359201-0).

Further research under direct NCEE funding will include examination of response differences associated with different survey administration modes, including panel-based internet, central location, mall-intercept, and phone-mail. The water quality project fielded surveys in each of these modes, and differences in responses and demographic characteristics between those surveys will serve as the units of comparison.