September 22, 1995

MEMORANDUM FOR Elizabeth Martin
Chief, Center for Survey Methods Research
Statistical Research Division

From: Yukiko Ellis
Center for Survey Methods Research
Statistical Research Division

Subject: Examination of Census Omission and Erroneous Enumeration Based on 1990 Ethnographic Studies of Census Coverage

Attached is the second of two reports on results of the study based on the Ethnographic Evaluation of the Behavioral Causes of Census Undercount for the 1990 Decennial Census. This report focuses on the results of an exploratory data analysis in which logistic linear regression models were fit to data to assess the relative values of explanatory variables on the prediction of census omissions and erroneous enumerations. This paper was presented at the 1995 Joint Statistical Meetings in Orlando, Florida on August 15, 1995.

I. INTRODUCTION

This paper reports results of an analysis of data from the Ethnographic Evaluation of the Behavioral Causes of Census Undercount for the 1990 Decennial Census (referred to as the 1990 Ethnographic Evaluation henceforth). The purpose of the analysis was to search for factors related to the two components of census coverage errors, omission and erroneous enumeration of persons, in the 1990 census.

Earlier papers (de la Puente, 1992 and 1993b) on the 1990 Ethnographic Evaluation focused on census omissions and examined the effects of demographic variables on the outcome of census enumeration. This paper broadened the scope of the analyses on two fronts. First, not only census omissions but also erroneous enumerations were examined. Second, in addition to demographic variables, factors related to the social aspects of the sample areas were included in the analyses. The results of the analyses in this paper confirmed many of the results of the earlier studies on the census coverage errors but also shed some new light on the possible effects that social and demographic factors might have had on the outcome of the census enumeration.

II. BACKGROUND

The Census Bureau began a series of ethnographic evaluations of census coverage in 1986, culminating in the 1990 Ethnographic Evaluation. The history of the ethnographic evaluations and the study design of the 1990 Ethnographic Evaluation have been documented by Brownrigg and Martin (1989) and Martin, Brownrigg, and Fay (1990). The 1990 Ethnographic Evaluation consisted of intensive studies of twenty-nine small areas conducted by ethnographers. Each principal ethnographer had a close tie with the community and previously worked in and resided near the study area. As part of the evaluation project, each ethnographer conducted an Alternative Enumeration (AE) which was an independent (from the census) listing of the residents in the sample area, using participant observation and ethnographic interviews. The ethnographer collected Census Day residency status of each individual during the period of June to August 1990. The AE person list was later linked to the census person list, and persons missed or erroneously counted by the census were identified in the Resolved Enumeration (RE). Each sample area included about 100 households in one or more census blocks. Twenty-eight of the sample areas were located in the continental U.S. and one in Puerto Rico. This paper will be concerned only with the twenty-eight sample areas in the continental U.S. The sample areas were selected, purposively, representing five groups (Blacks, Hispanics, Asians, American Indians, and recent immigrants) in which undercounts were known or suspected to be high. The sample areas were also selected from three settings: ethnically homogeneous

urban areas, ethnically heterogeneous urban and suburban areas, and ethnically homogeneous rural areas. In all, there were a total of 110 census blocks, 3367 housing units and 8718 individuals in the RE list.

One of the goals of the ethnographic evaluations was to understand and identify causes of differentially high undercount of minority males, especially of Black and Hispanic males. In the Ethnographic Coverage Reports[1/], the ethnographers reported that, in almost all sample areas, a group of factors, rather than one single factor, contributed to census omission and erroneous enumeration. The factors most frequently cited by the ethnographers were:

* Irregular and complex household arrangements;
* Language and illiteracy barriers;
* Concealment of information to protect resources; and
* Missed or erroneously enumerated housing units.

The ethnographers attempted to measure and quantify the above factors and additional information such as residential mobility and presence of violence in the behavioral log that each ethnographer was asked to keep during the project. However, a review of all of the behavioral logs revealed variations in consistency and completeness, raising concerns about the reliability of cross-site comparisons.

Using the demographic data collected by the ethnographers, de la Puente (1992) examined the census omissions among Hispanic individuals in the nine of the twenty-nine sample areas where 50 or more percent of the population were Hispanic. He later studied persons from all race/ethnic groups in all twenty-nine sample areas, and examined analytically the relation between the census omissions and the demographic variables and their interaction terms (1993b).

This paper extended the study of de la Puente (1993b) by examining erroneous enumeration as well as census omission. The effect of sample areas on the census coverage was also investigated. Its importance was discussed but not quantified by de la Puente in his paper. In addition, the paper attempted to corroborate quantitatively some of the findings in the Ethnographic Coverage Reports through use of the census long-form questionnaires as proxies to summarize the social, economic, and educational backgrounds of the persons residing in and around the sample areas. Together with the demographic information from the AE and a variable that attempted to summarize the sample area effects, subsets of these factors that best predicted the outcome of either census omission or erroneous enumeration were obtained.

III. LIMITATIONS

The AE data from the twenty-eight sample areas did not represent a probability sample. Hence, the results from this study

should not be generalized to any population or group beyond the twenty-eight sample areas in the study.

The Race variable in this study was defined as Hispanic, Black and Other. The last group, "Other," consisted of Asians, American Indians and Whites. This definition of "Other" category hence limited what one could learn specifically about Asians and American Indians.

The issue of the comparability and quality of ethnographic data from the twenty-eight sample areas will not be addressed in this paper. Note, however, that a training and orientation conference was held for the principal ethnographers from all sites to train them on census definitions and geography, and uniform procedures for the alternative enumerations.

The data based on the census long-form questionnaires from the sample and surrounding blocks were used in the analyses under an assumption that the socioeconomic landscape summarized by these data was stable over several contiguous census blocks in and around each sample area.[2/]

IV. METHODOLOGY

The data used in this report came from three sources - AE files, census short-form questionnaires and census long-form questionnaires.

The AE file for each of the twenty-eight sample areas was linked at the Census Bureau to the census short-form questionnaires which were delivered to every household identified by the Census Bureau in the sample areas. As the result of the linking process and the follow-up field work, a listing of the Resolved Enumeration was created by the ethnographer for each sample area. The RE included the data on Race, Age, Gender, Marital Status, Relation of an individual to the householder (in whose name the house was owned or rented), Household size, Geography, and Source. Source indicates whether the ethnographer enumerated the individual by direct observation, by information supplied by a household resident, or by other means such as information from neighbor, owner of building or administrative records. All but the Household size variable mentioned above were treated as discrete variables.

The census long-form questionnaires were delivered to a sample of households (about one in six) throughout the U.S. The following eight continuous variables were defined for this study at the site level, based on the census long-form questionnaires collected within the sample areas and the ring of one block surrounding the sample areas:

%LowEd = Proportion of persons with less than high school education among persons 18 years old or older;
MedInc = Median household income;
%Foreign= Proportion of persons born abroad;
%OthLan = Proportion of persons who spoke a language other

than English at home;
%FemHH = Proportion of female householders with no spouse;
%Owner = Proportion of owner-occupied housing units;
%Vac = Proportion of vacant housing units; and
%Ereturn= Proportion of households enumerated by enumerators and not by mail.

These eight variables were used to group the twenty-eight sample areas into clusters, using the average linkage method of the SAS's CLUSTER procedure. All data were standardized. The purpose of the clustering was twofold: first, to assess whether there was a natural grouping of the sample areas according to their socioeconomic backgrounds; and second, to create and introduce a variable that reflected a sample area effect in a manageable and efficient manner in the subsequent data analyses, thus circumventing the cumbersome use of indicator variables for the twenty-eight sample areas. See Ellis (1995) for clustering of the sample areas. Figure 1 illustrates the five clusters of the twenty-eight sample areas in the framework of the sample design by race/ethnicity and type of setting. The five clusters are:

Cluster 1 (9 sample areas): Hispanic and Asian immigrants with low MedInc and high %LowEd;
Cluster 2 (9 sample areas): Blacks, high %FemHH and high %Vac;
Cluster 3 (5 sample areas): Rural homeowners, Hispanic and American Indian;
Cluster 4 (3 sample areas): Hispanic and Asian immigrants with high MedInc and low %LowEd;
Cluster 5 (2 sample areas): List/Enumerate sample areas.

Three logistic regression models were fit to data derived from the RE, using the SAS's LOGIST procedure with stepwise option: one for census omission with two response categories (missed/correctly enumerated in the census), the second one for census omission with three response categories (correctly counted in the census/partial household (HH) miss/whole HH miss), and the third one for erroneous enumeration (erroneously/correctly enumerated). A person was a "partial HH miss" if the person was missed in a household in which at least one other resident was correctly enumerated in the census. A person was a "whole HH miss" if nobody in the household was correctly enumerated by the census. A parallel lines logistic regression model for an ordinal response was fit to data for the second model, based on the cumulative distribution probabilities of the response categories. For the census omission models, the data included persons who were either correctly enumerated in the census or who were enumerated in the AE but missed in the census. For the erroneous enumeration model, the data included persons who were either correctly or erroneously enumerated in the census. For the latter model, the variable Source was not entered into the model as this information was not available for persons erroneously enumerated in the census.

For each of the eight discrete explanatory variables, a set of design variables was formed to represent the categories of the variable, using the reference cell coding method (p.48, Hosmer and Lemeshow, 1989). The category in which the persons had the lowest odds of being missed in the census (or erroneously enumerated in the census, for the second part of the analysis), given all other explanatory variables in the model, was chosen as the reference cell. For a continuous variable, interpretation of the estimated coefficient depends on the particular units of the variable. Hence, for the eight continuous variables expressed in proportions, the estimated odds ratio was computed for an increase of twenty percent in the variable. For the MedInc variable, the estimated odds ratio was computed for an increase of $1000 in the variable.

V. RESULTS

A. Census Omission

Table 1 shows the maximum likelihood estimates of the coefficients, standard errors, odds ratios and their 95 percent confidence intervals for a multiple logistic regression model in which the census omission was the binary response variable. This was the best model obtained among the models that included the main effects only. The estimated coefficient is statistically significant at a 5 percent significance level if the 95 percent confidence interval of the odds ratio computed from the coefficient does not include the value of 1.

All eight discrete variables were found to contribute significantly in explaining the outcome of census omissions, after controlling for the variables in the model. Among the eight variables, the Cluster and Relation variables were found to have the strongest effects on the response variable. Persons in Cluster 4 (Hispanic and Asian immigrants with high median household income), for example, were twenty-nine times more likely and persons in Cluster 1 (Hispanic and Asian immigrants with low median household income) were eleven times more likely to be missed by the census than persons in Cluster 3 (Rural homeowners, American Indian and Hispanic). Clusters 3 and 5, which included rural sample areas in the study, had relatively low odds of persons being missed in the census in comparison to other clusters. Note, however, that the parameter estimates for the Geography variable show that persons in rural areas were most likely to be missed in the census compared to persons in urban or urban/suburban areas, after controlling for all other variables in the model.[3/]

Persons not related to householders in the Resolved Enumeration were almost four times more likely to be missed by the census than householders. With respect to the race variable, the odds ratio of 1.87 between Hispanic and Other indicates that persons were missed by the census almost twice as often among

Hispanics than among Others, after statistically adjusting for all other variables in the model. Blacks were slightly more likely to be missed than Others. The two odds ratios (1.87 and 1.17) were lower than what one might have expected, probably because the sample in this study included disproportionately large proportions of Asians (1/3) and American Indians (1/3) in the 'Other' race category compared to the general population.

Among the eight census long-form variables and the HHsize variable, all but the MedInc, %Othlan, and %Ereturn variables were found to contribute significantly to the model, given all other variables in the model. Among the six variables, %Foreign had the largest effect on the response. The negative sign on its parameter estimate indicates that the higher the proportion of persons born abroad in a sample area the lower the likelihood of a resident of the sample area being missed in the census, given the specific native-born populations in the study. At first glance, this finding appears to contradict what one might have expected. This will be further discussed later. %FemHH, %Owner and %Vac were all found to have negative signs on their parameter estimates. Again, the negative signs on the parameter estimates for %FemHH and %Vac might not be what one expected. For %LowEd, an increase of 20 percent in the proportion of persons with less than high school education in a sample area would increase the chance of being missed by the census by almost twofold.

When models with two-way interaction terms were fit to data, spurious results were obtained. Upon inspection, it became clear that empty cells were the culprit, which were generated when interaction terms were defined involving some sparsely distributed discrete explanatory variables. Hence, instead of fitting a model with two-way interaction terms, some cross-tabulations were examined. Table 2 contains four tables of the observed proportion of census omissions, conditional on the selected sets of two explanatory variables. The first three tables (Tables 2A, 2B, and 2C) were selected because of the paper's special interest in differential undercount by race. The Race variable was cross-classified with each of the three variables with the large estimated odds ratios in Table 1. Table 2D is a cross-tabulation between the two discrete variables with the largest estimated odds ratios.

When one of the explanatory variables was Relation (as in Tables 2A and 2D), the proportions of census omissions for householders and spouses were found to be close in magnitude. The proportion was slightly higher for "other relative" than for householders and spouses. For persons not related to householders, the proportions increased greatly to the same level across all categories of the second explanatory variable. For example, in the Relation by Race table (Table 2A), the observed proportion of census omissions for persons not related to householders ranged from 43 percent to 46 percent across the Race categories. In the

Relation by Cluster table (Table 2D), the observed proportion of census omissions in Cluster 4 stood out at 67 percent among persons not related to householders. For each cluster, the observed proportion of census omissions for persons not related was at least twice as big as the proportion for other relatives and it hovered in the range of 37 - 47 percent in Clusters 1, 2 and 3. Hence, these proportions in Tables 2A and 2D indicated two things. First, there appeared to be a two-way interaction in each table, namely, between the Race and Relation variables in Table 2A and between the Cluster and Relation variables in Table 2D. Second, given that a person was not related to the householder, the proportion of census omissions appeared to be statistically independent of the second explanatory variable.

Table 2B shows the proportion of census omissions in each cell defined by Cluster and Race. The number of sample persons in the denominator of the proportion varies widely from one cell to another because not all race/Hispanic groups were equally represented in each cluster. For example, there were very few (9) Hispanic persons in Cluster 5 (L/E sample areas) and all of them were correctly enumerated in the census. The table indicates that Cluster 5 had very low proportions of census omissions. It also demonstrates that the high proportion of census omissions in Cluster 4 (Hispanic and Asian immigrants with high median household income) was largely attributable to the high proportion of census omissions among Hispanic persons in the cluster. Cluster 4 had the sample persons consisting of 46 percent Hispanic, 44 percent Other (mainly Asian), and 10 percent Black.

The Source by Race table in Table 2C shows that the proportions of census omissions were high among Black and Hispanic persons whose enumerations in the Alternative Enumeration were based on the information either by 'other means' (a neighbor, administrative records, or owner of building) or through the ethnographers' participant observations. Hence, same people who were not report in the census were not reported to ethnographers by household members in the AE. The results point to the effectiveness of the ethnographic and administrative approaches to the census taking.

For the census omission response variable with three categories, the three response groups (correctly enumerated, partial HH miss, whole HH miss) were found to have different relations to the explanatory variables. That is, the parallel lines assumption was not met. Hence, a proportional odds model was not appropriate for the data.

Instead of fitting a logistic model to the data with the response variable with three categories, two three-way frequency tables in Table 3 were examined between the type of census omission and the selected explanatory variables. The Geography variable was selected so that the author could verify one of the findings of an earlier study by Childers (1993). Childers found in his 1990

Housing Unit Coverage Study (HUCS) that the proportion of whole household omissions where the census also missed their housing units was higher in rural areas than large and other urban areas. The Geography variable was cross-classified by the Race variable in Table 3A. Table 3B shows the relation by type of census omissions.

Table 3A shows that each race group had a distinct pattern of census omission by geography. Among persons in the "Other" race category, persons living in urban areas were most likely to be missed in the census compared to those living elsewhere. For each type of geography, the proportion of whole HH misses was almost twice as big as the proportion of partial HH misses for this race group. In the Black race category, persons living in rural and urban areas were more likely to be missed than those living in urban/suburban areas. In rural areas, a Black person was five times more likely to be a whole HH miss than to be a partial HH miss. These whole household misses might reflect the difficulty in finding housing units, and consequently, missing everybody living in the units, in sparsely populated rural areas where, for example, addresses were not marked clearly or units were hidden from public view down rural roads. The whole HH misses were also predominant among Blacks in urban areas as well. According to Hamid (1992), in the Harlem, NY, sample area where crimes were widespread, run-down buildings appeared abandoned but were not, and some brownstones seemed to be one family dwelling, but in fact, contained numerous housing units, all missed by the census. In the Hispanic race category, persons living in urban/suburban and urban areas were more likely to be missed than those living in rural areas. Partial HH misses were just as numerous as whole HH misses at each level of Geography among Hispanics. This could be partially explained by the ethnographers' observations that, in many urban sample areas, irregular housing went hand in hand with complex or irregular household arrangements. Because of a shortage of affordable housing, families and unrelated individuals doubled up in single housing units, some of which might have been illegally converted. Also, the proportion of homeowners was high among Hispanics in rural areas, based on the long-form questionnaires.

Unlike in the HUCS, the whole household person misses in this study were not further subclassified by whether their housing units were enumerated in the census or not. Hence, the comparison between the findings from this study and Childers' findings would not be valid. However, Table 3A shows that whole HH omissions (regardless of their housing unit enumeration status in the census) were more frequent in rural areas than urban or suburban areas only among Blacks.

Table 3B shows the frequency tabulation of the Relation variable by Type of Census Omission. Among persons not related to householders, only 56 percent of them were correctly enumerated in the census. The number of partial household omissions among persons not related to householders was twice as many as the number

of whole household omissions. This was in a stark contrast to the proportions for householders, spouses, and other relatives where the whole household omissions were predominant over the partial household omissions.

B. Erroneous Enumeration

Table 4 shows the maximum likelihood estimates of the coefficients, standard errors, odds ratios and their 95 percent confidence intervals for a multiple logistic regression model in which the erroneous enumeration was the binary response variable.

Among the eight discrete variables, the Gender, Race, and Geography variables were found not to contribute significantly in predicting the erroneous enumeration, after statistically adjusting for all other variables in the model. Among the remaining four discrete variables, the Cluster variable was again found to play a major role in predicting the response. This time, however, persons in Cluster 1 (Hispanic and Asian immigrants with low median household income) were found to be seventeen times more likely to be erroneously enumerated than persons in Cluster 4 (Hispanic and Asian immigrants with high median household income). Persons in Cluster 4, who had the highest odds of being missed, had the lowest odds of being erroneously enumerated by the census, after controlling for all other variables in the model. In the Miami, FL, sample area in Cluster 1, Stepick and Stepick (1992) reported duplicate enumeration of households that were visited more than once by the census workers. In the San Diego, CA, sample area, also in Cluster 1, Velasco (1992) mentioned irregular housing as being responsible for multiple enumeration and other erroneous enumerations of households. Persons in Cluster 3 (Rural homeowners, American Indian and Hispanic) had the next highest odds of being erroneously enumerated. More than half of the erroneous enumeration in Cluster 3 took place in the Marion County, OR, sample area where migrant workers in a migrant worker camp were erroneously enumerated by the census (Montoya, 1992).[4/]

Next to the Cluster variable, the Relation variable had a strong effect on the response variable. Persons not related to householders and "other relatives" had higher odds of being erroneously enumerated than householders and spouses. One of the anecdotes from the Ethnographic Coverage Reports that related to this finding included residential mobility among persons not related to householders as one of the contributing factors of erroneous enumeration in the sample areas with sizable immigrant populations. In another case, an erroneous inclusion of adult children was cited as one explanation of erroneous enumeration in the two sample areas with a sizable Chinese population.

Among the nine continuous variables, the %Foreign, %FemHH, %Vac, and HHsize variables did not contribute significantly to the model, after controlling for all other variables in the model.

Among the remaining five continuous variables, the %OthLan variable had the strongest effect on the response. The %OthLan variable, as with the %Foreign variable for census omission, had a negative sign on its parameter estimate, indicating that the lower the proportion of persons speaking a language other than English at home the higher the likelihood of erroneous enumeration.

The model also indicated that the higher the proportion of persons returning their census long-form questionnaires through enumerators, rather than by mail, the higher the odds of persons living in the sample area being erroneously enumerated.

As in the case of census omission, the greater the proportion of persons with less than high school education and the lower the proportion of owner-occupied housing units in a sample area the greater the likelihood of erroneous enumeration.

The MedInc variable, with a positive sign on its parameter estimate, indicated that the higher the median household income of a sample area the more likely it was for persons living in the sample area to be erroneously enumerated. Note that the median household income at site level in the study was low, ranging anywhere from $5,000 to $30,000.

Table 5A shows the observed proportion of erroneous enumerations conditional on the Cluster and Relation variables, the two variables with the largest odds ratios among the discrete variables. The numbers for Cluster 5 were too small to enable any valid observation of a trend to be made. The proportion of erroneous enumerations stayed uniform among relatives in each of the remaining four clusters. The proportion increased somewhat for persons not related to householders in Clusters 1, 2 and 4.

Table 5B shows the observed proportion of erroneous enumerations conditional on the Relation and Age variables. In each of Clusters 1, 2, and 4, the proportion of erroneous enumerations peaked in the 18-29 year old age group. In rural Cluster 3, however, the proportion remained stable in the range of 13-15 percent in all age groups. One possible explanation might be that the type of erroneous enumerations observed in rural areas might have often affected whole households (e.g., geocoding errors where a person was enumerated at the correct address but the housing unit was coded to the incorrect census geography).

VI. DISCUSSION

This paper has attempted to describe patterns of undercount and overcount within the selected sample areas, and to examine possible behavioral causes through indirect measurements of the social, economic, and educational backgrounds of the study areas.

The comparison of the multiple logistic regression models for census omission (Table 1) and erroneous enumeration (Table 4) indicates that persons who had high odds of being missed in the census had both similarities and differences from persons who had

high odds of being erroneously enumerated. (See Table 6 for the summary of the comparison.) A person who was either in the age group of 18-29 or who was not related to the householder had high odds of not only being missed, but also being erroneously enumerated in the census, controlling for all other variables in the model. Also, the higher the proportions of persons with less than high school education or in renter-occupied units in a sample area the higher the odds of census omission and erroneous enumeration in the sample area. The result on the educational variable was consistent with the ethnographers' observations that illiteracy among recent immigrants was one of the contributing factors to census coverage errors. The result on the tenure variable might be a reflection of irregular and complex household arrangements being more prevalent among renter-occupied units where the tenants might be unwilling to reveal their living arrangements to an outsider, as often reported by the ethnographers. Also, one would expect persons in renter-occupied units to be more mobile than persons in owner-occupied units, leading to more coverage errors.

For marital status, single persons who had never been married were most likely to be missed in the census while married persons were most likely to be erroneously enumerated in the census. This might be an indication that, overall, erroneous enumeration happened more at the household level than at individual level. The type of erroneous enumeration that affected whole households included a geocoding error or a fictitious enumeration of a household.

The household size variable contributed significantly to the logistic model for census omission but not for erroneous enumeration. A closer look at the variable showed that the proportion of census omissions hovered between 14 percent and 18 percent for household sizes of 1 to 6. The proportion was 20 percent for household size of 7 persons and 25 percent for household size of 8 persons or more.

For both census omission and erroneous enumeration, the Cluster variable played a major role. Albeit the sample in the study was not a probability sample of the country, the result of the analyses poses a question about the validity of the assumption of geographic homogeneity with respect to census coverage, at least in the subpopulations that included disproportionately large proportions of racial/ethnic minorities. In 1990 the mechanism of undercount was assumed to be different by region. Hence, post-strata were defined within the Census Division. Under this assumption, one might have expected large variability in coverage errors within clusters in this study since each cluster consisted of the sample areas from different Census Regions.[5/] Yet, the estimated standard errors for the clusters were found to be fairly stable and small in comparison to their estimated coefficients both for census omission and erroneous enumeration.

Another question of interest is: What is the implication of the study result on the Cluster variable for the post-stratification in coverage surveys? The study showed that a stratification scheme that incorporated variables on socioeconomic/ educational backgrounds of the sample areas might help fine-tune the scheme that only utilized the race/Hispanic origin and urban/ rural variables.

Both multiple logistic regression models included a few parameter estimates whose signs were contrary to the conventional wisdom. The %Foreign and %OthLan variables, two highly correlated variables, were good indicators of language barriers that might have existed among the residents in a sample area. The ethnographers often cited a language barrier to be one of the important contributing factors to census coverage errors, especially in the sample areas heavily populated by Hispanic and Asian persons. The parameter estimate for %Foreign was negative in the model for census omission. Similarly, the parameter estimate for %OthLan was negative for erroneous enumeration. One possible explanation for these results is that it was not the lack of knowledge of English per se that caused a person to be missed or erroneously enumerated in the census in this study group. The sample areas with large proportions of recent immigrants included larger proportions of persons with the characteristics that influenced census coverage errors (such as being 18-29 years old, unrelated to the householder and living in a renter-occupied housing unit) than the remaining sample areas. When we statistically adjusted for these characteristics that directly affected census coverage errors, we found that the %Foreign and %OthLan variables had an inverse relationship with census coverage errors.

In ethnically homogeneous urban areas, the ethnographers observed two different ways in which vacancy rate affected census omission. In the Black urban sample areas, especially in the Harlem, N.Y., sample area, the ethnographers reported that the census incorrectly reported many housing units to be vacant because of their dilapidated appearances. In this instance, the high vacancy proportion would be associated with a high omission proportion. In the Hispanic urban sample areas, on the other hand, the ethnographers often reported lack of affordable housing, leading to overcrowded conditions or addition of illegally converted units in the areas. In this case, the low vacancy proportion would be associated with a high omission proportion. The balance of these opposing influences of the %Vac on census omission was manifested in the negative sign of the %Vac variable.

In order to estimate numbers and characteristics of people missed and erroneously enumerated in the 1990 Census, the Census Bureau conducted the Post-Enumeration Survey (PES), an independent coverage survey with a probability sample, a few months after the census. Refer to Hogan (1993) for a complete background discussion

of the 1990 PES. Based on the PES, Griffin and Moriarity (1992) investigated the characteristics of erroneous enumeration while Moriarity and Childers (1993) and Ellis (1994) investigated the characteristics of census omission. Griffin and Moriarity found that fewer coverage errors were detected on census questionnaires that were completed by household members who returned their questionnaires by mail. They also reported that the most frequent types of error appeared to be due to residence rule violations and duplication. Moriarity and Childers concluded in their study that prompt self-enumeration by a household member would ensure the highest quality in the census data. In the 1990 Ethnographic Evaluation, the %Ereturn variable contributed significantly to the logistic model for erroneous enumeration but not for census omission. The %Ereturn variable in this study was based on the census long-form questionnaires while Moriarity and Childers based their result on the census short-form questionnaires. The %Ereturn variable based on the short-form questionnaires might have been a more sensitive measurement of self-enumeration by a household member than the long-form questionnaires.

The 1990 Ethnographic Evaluation results agreed with most of the findings from the 1990 PES and de la Puente (1993b). There were various factors that contributed to persons being missed or erroneously enumerated in the census. One factor that surfaced as having an especially important role in predicting within-household census coverage errors in this study as well as in Moriarity (1993), Moriarity and Childers (1993), and Ellis (1994) was the Relation variable. If a person was not related to the householder, then the person was found to have a very high risk of being either missed or erroneously enumerated in the census. The reasons were varied. In irregular and complex households, some members could not be easily related to the householder on the census form and they may have been listed in error or missed on the census roster. Some of the examples of those persons at risk that the ethnographers described in the 1990 Ethnographic Evaluation included:

* unrelated individuals living together for the sole purpose of sharing the rent;
* individuals in households that contained two or more "nuclear" families; and
* mobile or ambiguous household members.

Based on the Living Situation Survey (LSS), a coverage survey for roster research, Sweet (1994) reported that incorrect rostering due to residence rule violations were observed more frequently among persons who had high mobility, association with two or more places of residence, and who were considered to be non-household members by the LSS questionnaire respondents.

The PES study (Ellis, 1994) reported that, nationwide, Blacks and Hispanics had larger proportions of persons not related to householders (48 percent and 49 percent, respectively) compared to

White and Other (33 percent). The proportion of being missed in the census among persons not related to householders was uniform across tenure groups and household sizes. In this ethnographic study, it was again found that the proportion of being missed in the census was uniform across race groups among persons not related to householders. In other words, given that a person was not related to the householder, the probability of being missed in the census seemed to be independent of other factors. Hence, the differences in household composition may be the main contributing factor of the differential undercount within households. This possibility was also suggested by Fay (1989) in his analysis of the Current Population Survey data.

Table 3C shows that persons who were enumerated in the AE by household members were most likely to be correctly counted in the census. The high proportions of census omissions among Black and Hispanic persons who were enumerated either by 'other means' or by participant observation in the AE might indicate, on one hand, that the AE was more effective than the census in enumerating persons in these particular racial/ethnic groups. On the other hand, it could also mean that these persons were enumerated in error in the AE and the error was not detected in the review process. In the 1990 Ethnographic Evaluation, the quality assurance procedures were not as vigorous and uniform across the sample areas as in the 1990 PES.

VII. SUGGESTIONS FOR FUTURE RESEARCH

The ethnographic studies have proved useful not only in elucidating various causes of undercount but also in identifying persons who were incorrectly enumerated in the census. More research is recommended to see whether a post-census survey that is used for coverage evaluation can incorporate the ethnographic evaluation to improve census coverage, especially in the areas suspected of high undercount. Wright (1995) suggested this approach in the case where Census Plus was the post-census survey. The consolidation of an ethnographic evaluation with a post-census survey would require a much earlier selection of the sample areas for the post-census survey than was done for the 1990 PES. This early planning should give ample time for the selection of ethnographers and enable the ethnographers to establish rapport with residents in the sample areas. The resolution enumeration between the ethnographic results and the census should be subjected to the same quality assurance put in place for the census and the post-census survey.

More research is recommended on finding ways to enumerate correctly, and estimate with acceptable accuracy, persons not related to householders. The research could be done at different levels: at the census questionnaire level (revise wording and add probes, for example), at the sample area level (target areas with an overcrowding problem), or at the estimation stage (poststratify

by the Relation variable). These issues address the within-household errors. As for the issues addressing the whole-household errors, a better method of address listing needs to be devised, especially for irregular and hidden housing units. Once again, a possible solution is the utilization of participant observations in ethnographic studies and having somebody familiar with each sample area canvass and list housing units.

NOTES

1. Each ethnographer submitted the final Ethnographic Coverage Report to the Census Bureau and the reports were summarized in de la Puente (1993a).
2. The population studied in the Ethnographic Evaluation was compared with the population enumerated by the census long-form questionnaires in the surrounding ring blocks by Ellis (1995) though comparisons of demographic variables in the two populations. The overall frequency distribution of each of the five demographic variables (age, gender, Hispanic Origin, relation to householder, and marital status) in the sample blocks was similar to the distribution of the corresponding demographic variable in the ring blocks. At site level, however, larger differences were observed between the sample areas and their ring blocks in regard to the demographic variables. Nevertheless, the site-level variables from the census long-forms were used in the analyses in this paper under an assumption that, in most cases, the socioeconomic landscape of an area would not drastically change over several contiguous census blocks.
3. This seemingly contradictory result can be explained easily if one compares the proportion of census omissions by geography within each cluster. In the 1990 Ethnographic Evaluation, the persons in rural areas were found to be most likely to be missed in the census within each cluster. Cluster 3 was the only cluster that included the sample areas from all three types of setting, and the persons in rural areas had the highest proportions of census omissions. What one observed here, therefore, resulted from the particular set of data used in the study. The result cannot be applied to the general population in the U.S. without further studies based on probability samples.
4. When the proportions of erroneous enumerations were compared by cluster, without controlling for any other explanatory variables, Cluster 3 had the highest proportion. If the 390 persons in Marion County, OR, in Cluster 3 were excluded from the computations, Cluster 3 was found to have the fourth highest proportion of erroneous enumerations, after Clusters 1, 4, and 2. In spite of this finding, the Marion County, OR, sample area was kept in the study, because the author believes that such procedural errors are not so rare in coverage surveys. Retaining the Marion County, OR, sample area in the analyses might help us identify the characteristics that make some sample areas more susceptible than others to erroneous enumerations.
5. Census Divisions are subsets of Census Regions. Fifty states and the District of Columbia are subdivided into four Census Regions as follows:
   Northeast Region: CT, MA, ME, NH, NJ, NY, PA, RI, VT.
   South Region: AL, AR, DE, DC, FL, GA, KY, LA, MD, MS, NC, OK, SC, TE, TX, VA, WV.
   Midwest Region: IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, WI.
   West Region: AK, AZ, CA, CO, HI, ID, MT, NV, NM, OK, UT, WA, WY.

REFERENCE

Brownrigg, L. and Martin, E. (1989). "Proposed Study Plan for Ethnographic Evaluation of the Behavioral Causes of Undercount." Paper presented to the Census Advisory Committees of the American Statistical Association and on Population Statistics at the Joint Advisory Committee Meeting, Alexandria, VA, April 13-14, 1989.

Childers, D. (1993). "Coverage of Housing in the 1990 Decennial Census." Preliminary Research Evaluation Memorandum #236. Washington, DC: Bureau of the Census.

de la Puente, M. (1992). "An Analysis of the Underenumeration of Hispanics: Evidence from Small Area Ethnographic Studies." Proceedings of the Bureau of the Census 1992 Annual Research Conference, pp.45-69. Washington, DC: Bureau of the Census.

de la Puente, M. (1993a). "Why Are People Missed or Erroneously Included by the Census: A Summary of Findings from Ethnographic Coverage Reports." Paper prepared for the Advisory Committee for the Design of the Year 2000 Census Meeting, Suitland, MD, March 3, 1993.

de la Puente, M. (1993b). "A Multivariate Analysis of the Census Omission of Hispanics and Non-Hispanic Whites, Blacks, Asians and American Indians: Evidence from Small Area Ethnographic Studies." Proceedings of the Survey Research Methods Section, American Statistical Association, 641-646.

Ellis, Y. (1994). "Categorical Data Analysis of Census Omissions." DSSD 1990 REX Memorandum Series #PP-10. Washington, DC: Bureau of the Census.

Ellis, Y. (1995). "Clustering of Twenty-Eight Sample Areas in the Ethnographic Evaluation of the Behavioral Causes of Census Undercount for the 1990 Decennial Census." Unpublished document. Washington, DC: Bureau of the Census.

Fay, R. (1989). "An Analysis of Within-household Undercoverage in the Current Population Survey." Proceedings of the Bureau of the Census Fifth Annual Research Conference, pp.156-175. Washington, DC: Bureau of the Census.

Griffin, D. and Moriarity, C. (1992). "Characteristics of Census Errors." Proceedings of the Survey Research Methods Section, American Statistical Association, 512-517.

Hamid, A. (1992). Ethnographic Follow-Up of a Predominately African American Population in a Sample Area in Central Harlem, New York City: Behavioral Causes of the Undercount of the 1990 Census. Ethnographic Evaluation of the 1990 Decennial Census Report #11. Prepared under Joint Statistical Agreement 89-28 with John Jay College of Criminal Justice in New York City. Washington, DC: Bureau of the Census.

Hogan, H. (1993). "The 1990 Post-Enumeration Survey: Operations and Results." Journal of the American Statistical Association, 88, 1047-1060.

Hosmer, D. and Lemeshow, S. (1989). Applied Logistic Regression. New York: Wiley.

Martin, E., Brownrigg, L. and Fay, R. (1990). "Results of 1988 Ethnographic Studies of Census Coverage and Plans for 1990." Paper presented to the Census Advisory Committees of the American Statistical Association and on Population Statistics at the Joint Advisory Committee Meeting, Alexandria, VA, October 18-19, 1990.

Montoya, D. (1992). Ethnographic Evaluation of the Behavioral Causes of Undercount: Woodburn, Oregon. Ethnographic Evaluation of the 1990 Decennial Census Report #25. Prepared under Joint Statistical Agreement 90-06 with the University of Oregon. Washington, DC: Bureau of the Census.

Moriarity, C. (1993). "Characteristics of Census Error-Additional Results." DSSD 1990 REX Memorandum Series #PP-7. Washington, DC: Bureau of the Census.

Moriarity, C. and Childers, D. (1993). "Analysis of Census Omissions: Preliminary Results." Proceedings of the Survey Research Methods Section, American Statistical Association, 629-634.

Stepick, A. and Stepick C. (1992). Alternative Enumeration of Haitians in Miami,

Florida. Ethnographic Evaluation of the 1990 Decennial Census Report #8. Prepared under Joint Statistical Agreement 90-08 with Florida International University. Washington, DC: Bureau of the Census.

Sweet, E. (1994). "Coverage Implications of Using the Term 'Usual Residence'." DSSD 2000 Census Memorandum Series #H-17. Washington, DC: Bureau of the Census.

Velasco, A. (1992). Ethnographic Evaluation of the Behavioral Causes of Undercount in the Community of Sherman Heights, California. Ethnographic Evaluation of the 1990 Decennial Census Report #22. Prepared under Joint Statistical Agreement 89-42 with the Chican Federation of San Diego County. Washington, DC: Bureau of the Census.

Wright, T. (1995). "Census Plus: A Sampling and Prediction Approach for the 2000 Census of the United States." Paper presented at the Bureau of the Census 1995 Annual Research Conference, Arlington, VA, March 20, 1995.

FIGURE 1.

FIVE CLUSTERS OF TWENTY EIGHT SAMPLE AREAS IN THE FRAMEWORK OF SAMPLE DESIGN BY RACE/ETHNICITY AND TYPE OF SETTING

(The Puerto Rico sample area is excluded.)

Race/ Ethnicity	Ethnically Homogeneous Urban Site		Ethnically Heterogeneous Urban/Suburban Site		Ethnically Homogeneous Rural Site	
	Clst No.	Sample Area	Clst No.	Sample Area	Clst No.	Sample Area
ASIAN	1 1 4	Chinatown,CA Koreatown,CA Queens,NY	1 1 2 2	Long Beach,CA N.Beach,CA S.St.Louis,MO Chicago,IL		
Undoc. Hispanic Residents	1 1	Bronx,NY Miami,FL	1 4 4	San Francisco,CA Long Island,NY Houston,TX		
HISPANIC	1	San Diego,CA	1 2	New Orleans,LA Hartford,CT	3 3	Santa Barbara,CA Marion County,OR
BLACK	2 2 2 2	Flint,MI Orleans Parish,LA Harlem,NY Ft.Lauderdale,FL	2 3	N.St.Louis,MO Carbondale,IL	2 5	Holmes County,MS Logan County,OK
AMERICAN INDIAN					3 3 5	Little Branch,NC Okfuskee County,OK Isleta Pueblo,NM

Cluster Labels (A number in parentheses indicates the number of sample areas.):

Cluster 1: Hispanic and Asian immigrants with low median household income (9);
Cluster 2: Blacks (9);
Cluster 3: Rural American Indian/Hispanic homeowners (5);
Cluster 4: Hispanic and Asian immigrants with high median household income (3); and
Cluster 5: List/Enumerate sample areas (2).

TABLE 1

MAXIMUM LIKELIHOOD ESTIMATES OF THE MULTIPLE LOGISTIC REGRESSION MODEL OF THE PROBABILITIES OF BEING MISSED IN THE CENSUS

Variable	Est. Coeff	Est. Std.Err	Coeff/SE	Est. Odds Ratio	95% C.I. of Odds Ratio
RACE					
Black vs & Other	0.1541	0.1053	1.46	1.17	(0.95, 1.43)
Hisp vs & Other	0.6270	0.1008	6.22	1.87	(1.54, 2.28)
AGE					
0-17 vs 50+	0.2208	0.1378	1.60	1.25	(0.95, 1.63)
18-29 vs 50+	0.5810	0.1223	4.75	1.79	(1.41, 2.27)
30-49 vs 50+	0.2969	0.1117	2.66	1.35	(1.08, 1.68)
GENDER					
Male vs Female	0.2021	0.0704	2.87	1.22	(1.06, 1.41)
RELATION					
Spouse vs Hsehldr	0.1444	0.1224	1.18	1.16	(0.91, 1.47)
Oth Rel vs Hsehldr	0.2770	0.1157	2.39	1.32	(1.05, 1.66)
Non-Rel vs Hsehldr	1.3323	0.1467	9.08	3.79	(2.84, 5.05)
MARITAL STATUS					
Marrd vs Sp/Dv/Wd	0.1352	0.1336	1.01	1.14	(0.88, 1.49)
Single vs Sp/Dv/Wd	0.3428	0.1248	2.75	1.41	(1.10, 1.80)
CLUSTER					
Clstr1 vs Clstr3	2.3731	0.2542	9.34	10.73	(6.52, 17.66)
Clstr2 vs Clstr3	1.7942	0.2654	6.76	6.01	(3.58, 10.12)
Clstr4 vs Clstr3	3.3662	0.2745	12.26	28.97	(16.91, 49.61)
Clstr5 vs Clstr3	0.3006	0.2479	1.21	1.35	(0.83, 2.20)
GEOGRAPHY					
Urban vs Urb/Subrb	0.2625	0.1078	2.44	1.30	(1.05, 1.61)
Rural vs Urb/Subrb	0.9570	0.1852	5.17	2.60	(1.81, 3.74)
SOURCE					
Obser vs HHmem	0.5072	0.0831	6.10	1.66	(1.41, 1.95)
Other vs HHmem	0.6137	0.1385	4.43	1.85	(1.41, 2.42)
HHsize	0.0546	0.0163	3.35	1.06	(1.02, 1.10)
%LowEd	2.5600	0.4783	5.35	1.67*	(1.38, 2.01)
%Foreign	-5.3612	0.4341	-12.35	0.34*	(0.29, 0.41)
%FemHH	-3.6582	0.5198	-7.04	0.48*	(0.39, 0.59)
%Owner	-1.2160	0.2932	-4.15	0.78*	(0.70, 0.88)
%Vac	-2.2777	0.5227	-4.36	0.63*	(0.52, 0.78)
Constant	-2.7399	0.3872	-7.08		

N = 7292 persons. Log-likelihood=-2923.40.
Association of predicted probabilities and observed responses:
Concordant=72.7%, Discordant=26.8%, Tied=0.5%,
(7,412,691 pairs)

* The estimated odds ratio was computed for an increase of 20% in this variable, assuming that the logit was linear in the variable. For example, the estimated odds ratio for an increase of 20% in %LOWED was: exp(2.5600 * 0.20)=1.67.

TABLE 2

OBSERVED PROPORTION (AND NUMBER) OF CENSUS OMISSIONS CONDITIONAL ON TWO EXPLANATORY VARIABLES

	R A C E		
A. RELATION	Other	Black	Hispanic
Householder	10% (111)	19% (124)	12% (71)
Spouse	11% (64)	20% (40)	13% (44)
Other Relative	14% (197)	20% (178)	20% (239)
Not Related	43% (31)	46% (42)	45% (80)
B. CLUSTER			
Cluster 1	15% (161)	18% (54)	19% (202)
Cluster 2	14% (77)	24% (288)	28% (27)
Cluster 3	13% (105)	10% (7)	9% (72)
Cluster 4	12% (38)	22% (17)	40% (133)
Cluster 5	6% (22)	10% (18)	0% (0)
C. SOURCE			
By Observation	15% (257)	20% (190)	24% (208)
Household Member	10% (115)	16% (106)	15% (206)
By Other Means	13% (31)	35% (88)	23% (20)

	C L U S T E R				
D. RELATION	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Householder	12% (90)	18% (127)	7% (32)	17% (45)	7% (12)
Spouse	15% (52)	20% (45)	8% (25)	17% (23)	4% (3)
Other Relative	17% (211)	21% (180)	13% (117)	30% (84)	8% (22)
Not Related	39% (64)	47% (40)	37% (10)	67% (36)	25% (3)

TABLE 3. TYPE OF CENSUS OMISSIONS BY RACE/GEOGRAPHY AND BY RELATION

A.RACE/ GEOGRAPHY	TYPE OF CENSUS OMISSION				
	Correct Enumeration	Partial HH Omission	Whole HH Omission	Row Total	Column Percent
Other					
Urban	696 (83%)	54 (6%)	88 (11%)	838 (100%)	(27%)
Urb/Surb	1082 (89%)	45 (4%)	92 (7%)	1219 (100%)	(39%)
Rural	970 (90%)	37 (3%)	74 (7%)	1081 (100%)	(34%)
Black					
Urban	682 (78%)	57 (7%)	133 (15%)	872 (100%)	(48%)
Urb/Surb	473 (83%)	38 (7%)	59 (10%)	570 (100%)	(31%)
Rural	288 (75%)	16 (4%)	81 (21%)	385 (100%)	(21%)
Hispanic					
Urban	532 (77%)	74 (11%)	82 (12%)	688 (100%)	(30%)
Urb/Surb	621 (75%)	80 (10%)	126 (15%)	827 (100%)	(36%)
Rural	722 (91%)	42 (5%)	30 (4%)	794 (100%)	(34%)
B. RELATION					
Householder	2029 (87%)	39 (2%)	263 (11%)	2331 (100%)	(32%)
Spouse	971 (87%)	38 (3%)	107 (11%)	1116 (100%)	(15%)
Other Relat	2877 (83%)	265 (7%)	345 (10%)	3487 (100%)	(48%)
Not Related	189 (56%)	101 (30%)	50 (15%)	340 (100%)	(4%)

TABLE 4

MAXIMUM LIKELIHOOD ESTIMATES OF THE MULTIPLE LOGISTIC REGRESSION MODEL OF THE PROBABILITIES OF BEING ERRONEOUSLY ENUMERATED IN THE CENSUS

Variable	Est. Coeff	Est. Std.Err	Coeff/SE	Est. Odds Ratio	95% C.I. of Odds Ratio
AGE					
18-29 vs 0-17	0.6325	0.1178	5.37	1.88	(1.49, 2.37)
30-49 vs 0-17	0.5402	0.1473	3.67	1.72	(1.29, 2.29)
50+ vs 0-17	0.3355	0.1534	2.19	1.40	(1.04, 1.89)
RELATION					
Hsehldr vs Spouse	0.1158	0.1311	0.88	1.12	(0.87, 1.45)
Oth Rel vs Spouse	0.6175	0.1572	3.93	1.85	(1.36, 2.52)
Non-Rel vs Spouse	0.6245	0.2121	2.94	1.87	(1.23, 2.83)
MARITAL STATUS					
Marrd vs Sp/Dv/Wd	0.3570	0.1450	2.46	1.43	(0.97, 2.12)
Single vs Sp/Dv/Wd	0.2002	0.1516	1.32	1.22	(0.91, 1.64)
CLUSTER					
Clstr1 vs Clstr4	2.8222	0.3159	8.93	16.81	(9.05, 31.23)
Clstr2 vs Clstr4	1.7959	0.3091	5.81	6.02	(3.29, 11.04)
Clstr3 vs Clstr4	2.3183	0.2546	9.11	10.12	(6.17, 16.73)
Clstr5 vs Clstr4	0.4241	0.4691	0.90	1.52	(0.61, 3.83)
%LowEd	1.2864	0.4442	2.90	1.29*	(1.08, 1.54)
MedInc	1.98E-4	0.17E-4	11.47	1.22**	(1.18, 1.26)
%OthLan	-2.9151	0.2989	-7.20	0.56*	(0.50, 0.63)
%Owner	-1.9124	0.2927	-6.53	0.68*	(0.61, 0.77)
%Ereturn	2.5909	0.4976	5.21	1.68*	(1.38, 2.04)
Constant	-8.0719	0.5293	-15.25		

N = 7154 persons. Log-likelihood=-2452.95.

Association of predicted probabilities and observed responses:
Concordant=66.1%
Discordant=32.5%
Tied = 1.4%
(5,088,753 pairs)

* The estimated odds ratio was computed for an increase of 20% in this variable, assuming that the logit was linear in the variable. For example, the estimated odds ratio for an increase of 20% in %LOWED was: exp(1.2864 * 0.20)=1.29.

** The estimated odds ratio was computed for an increase of $1000 in the median household income.

TABLE 5. OBSERVED PROPORTION (AND NUMBER) OF ERRONEOUS ENUMERATIONS CONDITIONAL ON TWO EXPLANATORY VARIABLES

	C L U S T E R				
A. RELATION	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Householder	10% (72)	9% (64)	14% (71)	8% (20)	3% (5)
Spouse	9% (30)	8% (15)	15% (55)	11% (14)	3% (2)
Other Relat	13% (153)	10% (87)	14% (124)	12% (29)	4% (13)
Not Related	17% (23)	17% (13)	8% (2)	21% (8)	14% (1)
	C L U S T E R				
B. AGE	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
0-17	10% (76)	8% (46)	14% (96)	7% (9)	3% (6)
18-29	18% (88)	15% (46)	13% (38)	14% (25)	6% (7)
30-49	11% (66)	12% (55)	15% (70)	8% (15)	5% (7)
50+	9% (48)	7% (32)	14% (48)	15% (22)	1% (1)

TABLE 6. SUMMARY OF LOGISTIC MODELS FOR CENSUS OMISSION (TABLE 1) AND ERRONEOUS ENUMERATION (TABLE 4)

For each discrete variable, its categories are listed in the descending order of the magnitude of their estimated odds. For each continuous variable, the estimated odds ratio is shown for an increase of 20 percent in the variable.

	Statistically Significant Influence on:	
Variable	Omission	EE
Discrete Variables:		
CLUSTER	Cluster 4	Cluster 1
	Cluster 1	Cluster 3
	Cluster 2	Cluster 2
	Cluster 5	Cluster 5
	Cluster 3	Cluster 4
RELATION	Non-Relative	Non-Relative
	Other Relative	Other Relative
	Spouse	Householder
	Householder	Spouse
SOURCE	Other(Neighbor,ect)	(Not Applicable)
	By Observation	
	Household Member	
GEOGRAPHY	Rural	---
	Urban	
	Urban/Suburban	
AGE	18-29	18-29
	30-49	30-49
	0-17	50+
	50+	0-17
RACE	Hispanic	---
	Black	
	Other	
MARITAL	Single	Married
	Married	Single
	Sep/Div/Wid	Sep/Div/Wid
GENDER	Male	---
	Female	
Continuous Variables:		
%LowEd	1.67	1.29
%Owner	0.78	0.68
%Foreign	0.34	---
%FemHH	0.48	---
%Vac	0.63	---
HHsize	1.06	---
MedInc	---	1.22
%OthLan	---	0.56
%Ereturn	---	1.68