1-1996

Family Structure and Institutionalization: Results from Merged Data

James McNally
The University Of Michigan

Douglas Wolf
Syracuse University

ISSN 1084-1695

Papers in Microsimulation Series Paper No. 2

FAMILY STRUCTURE AND INSTITUTIONALIZATION: RESULTS FROM MERGED DATA

James McNally and Douglas Wolf

Maxwell Center for Demography and Economics of Aging

Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
Syracuse, New York 13244-1090

January 1996

This research was supported by grant numbers R01 AG11815 and T32 AG00238 from the National Institute on Aging.

Abstract

Research on the patterns and behavioral consequences of kin networks among the older population is limited due to the shortcomings of most available survey data. Often, household surveys obtain little information on the number and characteristics of nonresident kin. Moreover, surveys are often confined to the noninstitutionalized population. One possible solution is to merge information from multiple sources, in order to achieve the requisite coverage of populations and data content.

This paper reports on the development of a hybrid data base containing observations from the 1987-88 National Survey of Families and Households (NSFH) and the 1989 National Long-Term Care Survey (NLTCS). One population group—disabled noninstitutionalized elders—is represented in both data sources. However, there is insufficient detail with which to identify such persons in the NSFH. Instead, we develop a probabilistic model for identifying which NSFH cases are drawn from the same population as is the NLTCS, and randomly discard them from the pooled sample using a multiple imputation approach.

A multivariate analysis of the prevalence of nursing home residence based on the pooled sample reveals that the numbers of sons and daughters have different effects on the risk of nursing home residence among older men and women.

INTRODUCTION

Research on kinship patterns in the United States often uses data that provide an incomplete representation of the population of interest. Studies of intergenerational coresidence, for example, rarely consider the institutionalized elderly population, not because such individuals are unimportant potential coresidents, but because most survey sampling frames exclude the institutionalized. Similarly, studies of the institutionalized elderly are rarely able to include contrasts to an otherwise equivalent population of noninstitutionalized elderly. Unified samples containing both institutionalized and noninstitutionalized elderly would enhance the capacity for research on their residential status and intergenerational linkages.

Another central element in understanding kinship support patterns is information on adult children. The role of adult children in intergenerational transfers involving older parents is poorly understood, as research from the adult child's perspective has been limited by a lack of direct information on the availability of living parents. Failure to control for the existence of living parents—in other words, to properly specify the population at risk for parent-child transfers—can result in seriously understated estimates of exchange or support (McNally 1994). A related issue is the advantage to the child from participating in intergenerational resource flows. Although generally lacking controls for kin availability, prior work suggests that children are often the net gainers in resource flows with parents in terms of financial support, low-cost housing, and child care. Again, to study the advantages and costs of intergenerational kinship support and exchange we need to have both adult children and elderly parents in the same sample. Unfortunately, no existing data set adequately addresses these research needs in a systematic manner.

Some previous researchers have dealt with the problem of incomplete coverage of a population in survey data by pooling data from different samples. Relevant examples are Weissert and Scanlon (1983) and McBride (1989), both of whom combined data from institutional and noninstitutional samples. McBride (1989), for example, pooled observations from the 1984 Supplement on Aging (containing only noninstitutionalized respondents) and the 1985 National Nursing Home Survey (containing information only on institutionalized individuals). Apart from differences in timing, these data sources clearly represent nonoverlapping populations.

In this paper we propose a methodology for pooling and analyzing data from two *overlapping* samples, the 1987 National Survey of Families and Households (NSFH) and the 1989 wave of the National Long-Term Care Survey (NLTCS). For our purposes a key strength of both data sources is their extensive detail on kin networks. However, these two samples represent overlapping populations: the NSFH represents the *noninstitutionalized* population aged 19 and over, while the NLTCS contains detailed data on the *disabled* population 65 and older, whether or not institutionalized. Therefore, both data sources represent the following segment of the U.S. population: noninstitutionalized, disabled persons aged 65 and older.

A sample representing the entire elderly population, containing detailed covariates, could be created by pooling the NSFH with only the *institutionalized* portion of the NLTCS. However, for several reasons, we prefer to include *all* disabled respondents from the NLTCS: (a) the sample in the NLTCS is much larger than in the NSFH; (b) a much more extensive array of information pertaining to disability and long-term care service use is available in the NLTCS than in the NSFH; and (c) it is possible to construct parallel measures of disability for the institutionalized and noninstitutionalized portions of the NLTCS sample, but this is not possible for disabled respondents in the NSFH. Therefore, when pooling the data we must take account of the overlapping portions of these two samples. Our goal is to eliminate from the pooled sample the NSFH observations that would be classified as disabled using NLTCS criteria, if those criteria were available in the NSFH.

The NLTCS uses a brief "screener" interview, administered to a large sample of elderly individuals, to identify probable disabled respondents for detailed follow-up interviewing. Refined determination of disability status uses information obtained in these follow-up interviews (Manton 1988; Manton, Corder, and Stallard 1993). Screener instruments are not typically intended for use in secondary analyses, but rather to establish and delineate a target sample. A particularly difficult problem is that a screener rarely obtains complete demographic information for all members of the screened population. Those who lack the desired attributes are screened out with only a minimal record of basic demographic information. For this reason screener data are rarely used for analysis except under special circumstances. McNally (1989), for example, employed the screener information from the National Survey of Family Growth to link it back to its sampling frame, the National Health Interview Survey, in order to analyze nonresponse behavior. Hogan, Hao, and Parish (1990) employed the screener from the National Longitudinal Survey of Youth Labor Force Behavior to develop measures of kin availability drawn from the unpublished lists of follow-up contact persons gathered at the point of interview by field staff. The success of both projects resulted from access to confidential information in the screener, such as addresses and individual identifiers, which are rarely available to secondary data users. While both studies emphasized the importance of screener information, both also emphasized the difficulty of employing this information systematically. Manton, Corder, and Stallard (1993) also employed screener information from the NLTCS to provide summary measures of the prevalence of activity limitations among the elderly in the United States, but could not perform analysis on subgroups due to a lack of consistent covariates across the screened-in and screened-out samples.

Our basic approach is to apply the NLTCS screening criteria to relevant NSFH records (i.e., those 65 and older) in order to identify and discard from the NSFH sample those respondents who appear to "screen in" to the NLTCS. Unfortunately, the NLTCS screener data present us with some of the problems discussed in the preceding paragraph. Moreover, the NSFH does not contain many of the data elements with which to determine whether an NSFH respondent would be classified as "disabled" according to NLTCS criteria. Consequently, we develop a probabilistic model for the likelihood that a screened respondent would be classified as disabled according to NLTCS follow-up criteria. Because of the error inherent in this (as in any) probabilistic model, we employ a multiple imputation and analysis strategy with the pooled data.

Multiple imputation is a technique which makes use of repeated rounds of imputation (prediction) of the event of interest, in our case assigning a value to the unobserved variable "screen in to the NLTCS disabled category." Analyses based on the resulting series of independent estimates are then averaged. This final estimate is more reliable than one resulting from a single imputed value of the unobserved or missing data since it averages over imputation error. Inferences can be made using multiply imputed data that take into account both model variation and the additional variation resulting from the imputation process itself (Rubin 1987; Freedman and Wolf 1995).

In our analysis we multiply impute to NSFH observations an indicator that they would screen in to the NLTCS sample of disabled *non*institutionalized individuals. NSFH respondents so indicated are then discarded from the NSFH, and the remaining cases are pooled with NLTCS observations on disabled persons—both institutionalized and noninstitutionalized—for analysis. This results in a working data file representing all individuals 65 and older in the United States, regardless of their institutional status, and all noninstitutionalized individuals under the age of 65.[1]

One obvious problem with our approach is the timing of the two surveys. The average interview date for NSFH respondents occurred about 19.5 months earlier than the average interview date for the NLTCS screener. There is clearly some chance that the population represented by our modified NSFH sample includes some individuals who, during a 19.5-month period, became disabled and thus entered the population from which our NLTCS observations are drawn. We interpret our pooled analysis sample as one representing a hypothetical or "synthetic" cohort; however, it can be taken as representative of the U.S. population during this period as long as the distributions of characteristics within the disabled and nondisabled elderly populations were essentially stable during this period. We also note that the NSFH weights are calibrated to 1988 Current Population Survey control figures, while the list sample from which the NLTCS observations were drawn represents the 1988 Medicare rolls. These facts further support the use of the pooled data for population inferences, while not resolving the timing issue fully.

In the remainder of the paper we discuss each data source in more detail, and discuss the modeling and imputation process. We present evidence of the ability of the imputed data to reproduce population counts and prevalence estimates from other data sources. Finally, we illustrate the value of the pooled data in a multivariate analysis of the prevalence of nursing home residence among the older population.

DATA AND METHODS

National Survey of Families and Households

Our information on the noninstitutionalized, nondisabled adult population comes from the National Survey of Households and Families (NSFH). The NSFH is a complex, post-stratified nationally representative probability sample of 13,007 primary respondents age 19 and over, fielded from March of 1987 through May of 1988.[2] As noted above, a weighting scheme is employed to make the final sample representative of the United States based upon 1988 Current Population Survey estimates (Sweet et al. 1988).

The NSFH is attractive for the study of kin relationships for several reasons. From the perspective of the adult child it obtains information on the number and gender of surviving siblings, which have been shown to affect risks and patterns of parental mortality, coresidence, and institutionalization. In addition, the NSFH obtains information on the number of living children and selected demographic characteristics on each child. Unfortunately, the NSFH is limited to noninstitutionalized respondents. Another shortcoming of the NSFH lies in the fact that it represents the general adult population of the United States, so the sample contains only a small number of noninstitutionalized elderly with disabilities. We adopt the common practice of defining as disabled those persons with limitations on their ability to perform Activities of Daily Living (ADLs). This group represents an important contrast to the institutionalized impaired and, optimally, we would desire a data set which oversamples from this important subpopulation.

National Long-Term Care Survey

The 1989 National Long Term Care Survey (NLTCS) is the third in a series of nationally representative surveys of individuals aged 65 and older with health-based functional limitations. Surviving respondents from the earlier 1982 and 1984 interview waves were reinterviewed in 1989, while newly age-eligible respondents were screened for follow-up. Thus with appropriate weighting the data can be used for both longitudinal and cross-sectional analysis. The screening interview is administered to a random sample drawn from a list of individuals enrolled in Medicare, and thus represents virtually the entire population of persons 65 and older. The 1989 sample contains 4,463 noninstitutionalized persons (many of whom do *not* report disabilities) and 1,354 institutionalized individuals for whom detailed information was obtained in the follow-up interviews.

The NLTCS data are useful for the analysis of support networks from the parent's perspective, as they provide detailed information about the respondents' health, demographic and socioeconomic characteristics. The survey also obtains a set of characteristics of each of the individuals' surviving children, including age, sex, marital status, hours worked, presence of children in the child's home, distance from parent, and last date of contact with parent.

Unfortunately, the NLTCS also has limitations. As the data set lacks detailed information on the family networks of individuals without chronic limitations, we also lack a comparison group when our research interest extends further than the characteristics of the impaired aged population in the United States. While useful (and widely used), the NLTCS is limited in its ability to support comparisons between impaired and unimpaired older individuals; we are unable, for example, to analyze the transition from unimpaired to impaired in relation to the composition and behavior of kin networks.

While both of these data sets provide some information of great value for the study of family exchange among the elderly, neither provides all of the information needed for a complete analysis of this process. We argue that, lacking any alternative data set with this level of information, the merging of the NSFH and the NLTCS provides a reasonable means of indirectly obtaining the sample required for the analysis of exchange between the elderly and their kin network. As noted above, a suitable means for pooling the two samples can be developed using multiple imputation techniques.

Multiple Imputation

Although we are unable to determine which NSFH respondents would be categorized as disabled on the basis of NLTCS criteria, we are able to develop a probabilistic model of the chances that they would be, using data elements common to the two data sources. We can then predict, or impute, disability status to NSFH respondents using randomized assignment techniques. Rubin (1987) has shown that the use of multiple imputation improves upon standard single imputation methodologies. Further, the use of multiple imputation allows us to evaluate not only the sampling variance present in any sample survey, but also the additional variance introduced by the use of the imputations themselves.

Multiple imputation employs a multistage methodology in which a predictive model is estimated and used to predict otherwise missing values $K \geq 2$ times (Rubin 1987; Herzog and Rubin

1980; Treiman, Bielby, and Cheng 1988). Each round of prediction randomizes over the parameters of the predictive model. The imputed values are adjoined to observed values to produce K complete data files, each of which is used separately for statistical analysis. Results from the K separate analyses are averaged, with variances reflecting both average within-round and an adjusted between-round variance.

Rubin (1987, p. 160) defines three formal tasks to create a series of imputed values which approximate the posterior distribution of Y_{mis} (the set of observations with missing values of the variable Y) under a Bayesian model: the modeling task, the estimation task, and the imputation task. The *modeling* task consists of postulating a predictive model for Y, often one that conditions on covariates X. The *estimation* task produces the posterior distribution of the parameters of the predictive model, for example multivariate normal with mean vector and covariance produced in estimation based on the complete-data cases. The *imputation* task takes a random draw (Y_i^*) from the posterior distribution of Y_{mis} by first drawing a value from the posterior distribution of model parameters, then calculating Y conditional on that parameter value (and X_i if appropriate). This third step is repeated K times to create K imputations embedded in K complete data sets that are used for the final statistical analysis.

Separate estimates of summary statistics (e.g., means or regression coefficients) from each of the K complete data sets are combined by simple averaging, i.e., if θ_k represents the k^{th} such summary statistic then

$$\theta^* = \sum_{k=1}^{K} \theta_k , \tag{1}$$

represents the final estimate of the parameter of interest. The variance of θ^*, T^*, reflects both average within-round variance (denoted S^*) and adjusted between-round variance, with the latter calculated as

$$V^* = \frac{1}{K-1}\sum_{k=1}^{K} (\theta_k - \theta^*)^2. \tag{2}$$

The total variance of θ^* is then equal to

$$T^* = S^* + (1 + \frac{1}{K})V^* . \tag{3}$$

While little used so far in either demography or sociology, several applications of multiple imputation to social science data do exist. Herzog and Rubin (1980) used this methodology to compare a formal analytical model of imputation to an informal hot deck model to predict social security benefits using the March 1977 CPS. More recently, McNally (1994) has used multiple imputation to predict the likelihood of noncoresident children having living parents, and Freedman and Wolf (1995) use the characteristics of children to predict the marital status of a noncoresident mother or mother-in-law.

DATA POOLING USING MULTIPLE IMPUTATION

Our analysis plan is straightforward. Using information from the full NLTCS respondent pool (both screen-ins and screen-outs), we produce a predictive equation that describes the chances that a noninstitutionalized respondent to the NLTCS screener interview is selected for follow-up *and* determined in that follow-up interview to be disabled. Then, we use the estimated equation to predict the chances than an NSFH respondent would be included in the NLTCS follow-up sample, and classified as disabled, had that respondent received the NLTCS screener interview. This approach obviously requires that the predictive equation contain only variables available in equivalent form in both data files. NSFH respondents predicted to be classified as disabled are then discarded from the NSFH sample, leaving behind a subsample that we will interpret as "nondisabled." This subsample is then pooled with NLTCS observations on disabled respondents—both institutionalized and noninstitutionalized—to create a sample representative of the United States population (with the exception of institutionalized individuals *under* the age of 65).

The Modeling Task

Data Sources and Sample Selection. Our use of the 1989 NLTCS data is complicated by the fact that this survey employed different procedures for sample members previously followed up (in 1982 or 1984) and those newly screened in 1989. The 1989 NLTCS screener instrument includes questions on a small set of demographic variables, all potentially useable in our predictive model: age, sex, race, marital status, household size, region of residence, number of noncoresident children, and type of ADL or IADL limitation. Ideally we would have complete information on these variables from the screener for all 14,793 individuals in our sample. This, however, does not turn out to the case.

NLTCS respondents previously screened for follow-up were treated in 1989 as “automatic screen-ins” and, rather than receiving the entire screener interview, received only a brief address check and then proceeded directly to follow-up. For these automatic screen-ins, most covariates are recorded only in the follow-up interview. Among those *newly* screened, those screened *out* due to lack of apparent functional status limitations were asked the full battery of screener items. In this group, all covariates are recorded only in the screener interview. Those screened in, however, were skipped past the questions on demographic variables in the screener interview, and instead asked analogous questions in the follow-up interview. This creates two problems. The first problem is that of data loss. By deferring the gathering of basic information until the follow-up interview the risk of nonresponse to these questions is increased. The second is timing. The average date of follow-up interviewing occurred about 2.4 months later than the average date of screening. While some of the characteristics such as race and sex remain constant between the screener and the follow-up, many of the others are subject to change. During even this brief interval respondents can become widowed or divorced, and may experience changes in household size or number of noncoresident children, and, indeed, in functional status. This problem, while recognized, cannot be addressed. By necessity we assume that any changes between the time of the screener and follow-up interviews are negligible, and that any errors so introduced into our model are random.

Of the 16,147 persons screened in the 1989 NLTCS, 1354 were nursing home residents. These nursing home residents are excluded from the predictive model, since there are no NSFH respondents in nursing homes to exclude from our pooled data file. Of the remaining 14,793 individuals, 11,275 individuals are coded as nondisabled (and by definition noninstitutionalized) while the remaining 3,518 individuals are coded as disabled using the criteria of Manton et al. (1993). These various sample components, their role in the estimation of our predictive equation, and the sources of key covariates, are summarized in Table 1.

Variables. Marital status, coded from responses taken from different interview items as indicated in Table 1, is represented with four dummy variables indicating, respectively, that the respondent is *widowed, separated, divorced,* or *never married.* An additional variable indicates those respondents with *marital status unknown*; the omitted group is those currently married. *Household size* and *number of noncoresident children* were obtained in direct questioning in the screener interview. For those not asked those screener questions, however, we constructed the corresponding variables from listings provided by respondents to the follow-up questions. The respondent is excluded from the household size variable. Additional covariates included measures of *age*, sex (a dummy variable indicating a *male* respondent), race (dummy variables for those indicated as *black* or *other races*), and region (dummy variables for those living in the *Northeast*, *North Central* or *West*). For all the preceding variables, exact counterparts for NSFH respondents can be constructed or are provided directly.

An important predictor of disability status is, of course, reported difficulty with basic Activities of Daily Living (ADLs) or Instrumental Activities of Daily Living (IADLs). Unfortunately, the wording of IADL questions in the NSFH and the NLTCS are very different. While information on specific IADLs is obtained individually in the NLTCS screener, and in great detail in the NLTCS follow-up interview, the NSFH obtains IADL information using a single global question. Similarly, the wording of ADL questions differs significantly between the two surveys. The only exception is for the ADL question concerning whether or not the respondent has difficulty walking around in their own home. In this case the wording of the two questions is sufficiently close to allow us to construct a covariate reflecting this aspect of disability.

The NSFH includes a self-administered battery of items that includes the question "Do you have a physical or mental condition that limits your ability to ... move about inside the house?" with response categories "Yes" and "No." In the NLTCS screener interview, respondents are asked "Do [you] have any problem ... walking around inside without help?" NLTCS respondents automatically screened into follow-up were asked a more complex set of questions, including the series "[In the last week] did any person help [you] get around inside (or didn't [you] get around inside at all?)" ; "Did [you] use special equipment like a wheelchair, cane, or other device to help [you] get around inside?" and (if the answers to the preceding questions were "no") "Do [you] NEED help with getting around inside?" The variable *ADL Difficulty* used in our predictive equation codes as one those NLTCS respondents with affirmative responses to any of the preceding questions, using either screener or follow-up items as appropriate. The specific conditions sought in the two surveys are, strictly speaking, not identical: the NSFH directs respondents to consider only problems traceable to a "physical or mental condition," while the NLTCS does not. There are also unknown mode-effect differences, since the NSFH item is self-administered (albeit interspersed with face-to-face questions posed by the interviewer) while the NLTCS screener is administered either by telephone or in person, and the NLTCS follow-up is in person. Despite these differences, we treat the derived NLTCS variable as comparable to the similar NSFH item. The results of our imputations, discussed later, suggest that the model performs well, but this cannot be taken as conclusive evidence that our assumption of cross-survey comparability is entirely warranted.

Finally, the dependent variable in our predictive equation is coded one for NLTCS observations classified as disabled with respect to ADL or IADL functioning, according to criteria developed by Manton et al. (1993), and zero otherwise. These criteria imply not only that any reported difficulties with, or need for assistance with, ADL or IADL tasks are health related, but that they are serious enough to have lasted, or are expected to last, three or more months. Counts of respondents in these categories, further distinguished by sample component, are included in Table 1. About 300 cases (2 percent of the potential sample) are lost from the analysis due to missing information on various covariates.

We complete the modeling task by postulating a logistic regression to represent the relationship between disability status and the set of explanatory variables described above. Our priors for this model include the possibility of interactive effects among these explanatory variables, in addition to the usual main effects.

The Estimation Task

Prior to presenting the estimation results it bears mentioning that our goal is neither to test hypotheses nor to "explain" differentials in disability. Nor is the model intended to represent causal relationships. Rather, we seek to predict, as accurately as possible, whether a respondent with a specified set of traits will in addition be classified as disabled. Table 2 presents the results of two logistic regressions based on the sample and variables just described. The first equation contains the full set of covariates, and the second contains only those variables found to produce statistically significant coefficients.

The results in Table 2 indicate that age has a positive relationship to disability, but that this effect increases at a diminishing rate. A likely explanation for this pattern is the increasing selectivity of the noninstitutional population at older ages. Males are only 69 percent as likely to be included among disabled community residents as are women, while blacks are more than one and a half times more likely to be disabled than whites. While people of other races are also more likely to be classified as disabled than whites, this relationship is not significant. Both the gender and race relationships conform to expectations. Differential mortality between the sexes increases the number of elderly females as compared to elderly males in the ages where ADLs and IADLs are most common.

Significant marital status differentials in disability status are found as well. Widows are 25 percent more likely to be disabled, while those divorced are over two times more likely to be disabled, compared to the currently married. Those never married are also more likely to be disabled. These findings are consistent with the large literature showing that the currently married elderly tend to be healthier and more autonomous than those in a never married or post-married state. The control variable for unknown marital status proved to be insignificant, suggesting that this group does not represent a distinctive subsample of the respondent pool. Additional demographic characteristics, household size and the number of coresident children, are found to be strongly associated with disability status.

Regional variations in the chances of being disabled are also marked. Respondents who lived in the Northeastern United States are only 54 percent as likely to be disabled as those living in the Southern United States (the reference group). Similarly, those living in the North Central United States were only 64 percent as likely, while those living in the Western United States were 70 percent as likely to be disabled as were residents of the South.

The most powerful predictor in the model is the *ADL Difficulty* variable indicating difficultly moving around indoors. An NLTCS respondent who reported this limitation was 175 times more likely to be included in the NLTCS community interview, compared to one who reported that he or she did not have this limitation. While this odds ratio is extremely high, it must be noted that not all respondents reporting difficulty with indoor mobility are classified as disabled. Extensive testing of two- and three-way interactions found only one interaction which attained statistical significance. The interaction of being black and living in the West is positively related to being disabled, raising the odds by over two times.

In order to use the predictive model for imputation it was re-estimated with nonsignificant regressors dropped from the equation.[3] This revised predictive equation is also presented in Table 2. The elimination of nonsignificant regressors produces only slight changes in the point estimates, and their standard errors, for the variables retained in the model.

The second step of the imputation process, the estimation step, is completed by taking the estimated parameters of the second equation in Table 2, and the covariance matrix of those parameters, and treating them as the mean and variance, respectively, of the distribution of parameters to be used in the imputation step. Appealing to the asymptotic properties of maximum-likelihood estimation, we treat those quantities as the mean ($\hat{\beta}$) and variance ($\hat{\Sigma}$) of a multivariate normal distribution.

The Imputation Task

Randomization of Predictive Parameters. Using matrix manipulation routines detailed in Freedman (1990), a series of 16 independent draws from the posterior distribution of logistic regression parameters was performed. This set of 16 parameter vectors is displayed in Table 3. The average values for these 16 parameter vectors is also shown, for purposes of comparison to the original estimates. The averages are quite close to the “true” center of the distribution. It also appears that the variability in these randomized parameter vectors is reasonably small, a consequence of the relatively small standard errors of the original parameters.

Assignment of Imputed Values. Completion of the imputation task is relatively straightforward. We first multiply the parameters from one of the 16 parameter sets (shown in Table 3) by the corresponding variables for an observation in the NSFH. A logistic transformation of the resulting linear combination is then applied, producing a calculated probability of being disabled. These probabilities are compared to random draws from the uniform (0, 1) distribution. If the randomly drawn number is less than the predicted probability, the observation is classified as disabled. Otherwise the observation is determined to be nondisabled. This part of the imputation process reflects the uncertainty present in the predictive model. We know that persons with characteristics x_i have a particular chance of being disabled under the estimated model. We also know, however, that within each stratum of individuals sharing such characteristics not all individuals are in fact disabled. A distribution will exist within each stratum due to the effects of unmeasured variables and due to fundamental randomness. Failure to account for this variability causes imputed values to tend towards the mean value within strata (Rubin 1987). This tendency, even when values are distributed randomly, results in understated standard errors and overstated confidence intervals when statistical tests are employed.

As noted already, the multiple imputation procedures described above produce 16 imputed values of 0 and 1, indicating “nondisabled” and “disabled,” respectively, for each respondent 65 or older in the NSFH. We create 16 alternative pooled data files, in each case consisting of all NSFH cases coded as nondisabled (a number which can vary across imputation rounds) and all NLTCS disabled observations. Each such pooled data file represents an independent complete-data file with which to conduct further statistical analyses.

In order to guard against potential biases attributable to unmodeled correlations between disability status and the attributes used to stratify the NSFH sample, we confine our attention to the NSFH observations in the base sample. Observations from the various oversamples were discarded. Among the 2,017 NSFH respondents aged 65 and older, 231 are from oversample strata and discarded, leaving a useable sample of 1,786. Our decision to eliminate cases from the NSFH oversamples required that we reweight the cases from the base sample. We identified 45 distinct ex post strata (using the weights provided on the public-use file) and recalibrated them so that when used with the base sample alone, they reproduced the population distributions observed when the unadjusted weighting factors are used with the combined base and oversamples.

The random imputation procedures described above classified 361 (20.2 percent) of the NSFH older respondents as disabled, on average; the standard deviation of the number classified as disabled was about 5 percent. The remaining NSFH observations (1,425 on average) were pooled with 4,872 NLTCS respondents in each of 16 analysis files. The very different sampling rates used in the two source files necessitates the use of weights in all analyses of the pooled data: the average weight among NSFH observations is approximately ten times that of NLTCS observations.

Evaluation of the Multiply Imputed Data

The preceding steps produce 16 alterative data files, each representing the full population of older persons regardless of disability or institutional status. Here, we consider how well the imputed data files reproduce other estimates of the size of, and the prevalence of disability within, the population of older persons. We begin with comparisons of the merged, imputed data file to disability prevalence estimates based exclusively on the NLTCS (Manton et al. 1993); recall that the NLTCS is designed to produce prevalence estimates of disability, but contains only very limited sets of covariates for the *non*disabled population.

Figure 1 presents a scattergram which compares estimates of disability prevalence based on each of our 16 rounds of imputation, and their average, to the value presented in Manton et al. (1993). Each point in this scattergram represents the proportion of the total population 65 and older which has *no* reported ADL or IADL disability. Sampling weights are used in the calculation of these prevalence estimates. According to Manton et al. (1993), 77.4 percent of the 65 and older population reported no ADL or IADL limitations in 1989, with a standard error of ±0.3 percentage points. Using this information we have constructed a 90 percent confidence interval for the Manton et al. estimate, to see how well our predictive model replicates this expected proportion. The upper and lower bounds of this confidence interval are represented by the solid horizontal lines in Figure 1. Overall, the various imputations do quite well. Ten out of 16 of the within-round estimates are well within the constructed confidence interval. Of the 6 estimates which lie outside the confidence interval, only 1 represents a pronounced outlier and the other 5 are distributed near the lower boundary of the confidence interval.

The most important estimate is the final one, which averages the 16 independent estimates of the prevalence of disability within the older population. This illustrates the strength of the multiple imputation strategy which employs the uncertainty inherent in sparse predictive models to maximize limited information. The averaged value of the 16 within-round estimates falls within the 90 percent confidence interval associated with the NLTCS estimate.

An additional evaluation of our merged data set is presented in Table 4. Here the data set is stratified by sex and age, and then used to produce weighted estimates of population counts that are compared to analogous estimates based on the NLTCS alone, and to the intercensal population estimates for 1989 provided by the Census Bureau. Overall, the merged sample underestimates the total US population 65 and over by approximately 0.02 percent compared to a .004 percent overestimation for the NLTCS survey alone. This problem of underestimation is greatest for males, where the merged sample underestimates the population by .03 percent compared to .014 underestimation for the NLTCS alone. For females this difference is virtually identical, with .019 for the merged data set compared to .016 for the NLTCS alone. Based upon this comparison the merged data set reasonably reproduces the US population of those 65 and older as of 1989.

While it might be argued that the NLTCS does a marginally better job than our merged data set, the population estimates from the NLTCS are dependent upon the use of the analytic and screener files in addition to the community and institutional interviews. Because these files contain only the most basic demographic information they can only be used to generate prevalence estimates. Without the additional information provided by the merging of information on the nondisabled population in the NSFH with the disabled and institutionalized population in the NLTCS, detailed statistical analysis of exchange behavior and related outcomes for the entire United States elderly population is impossible.

Multivariate Analysis of the Multiply Imputed Data

In this section we use the multiply-imputed data set for an analysis of the cross-sectional risk of residing in a nursing home. We highlight the effects of number of living sons and daughters, effects uniquely revealed through use of our pooled data. As we further develop the pooled data, it will become possible to extend this analysis to include additional covariates such as income sources and amounts.

Differential risks of entry into, and exit from, nursing homes have been analyzed in several studies (Cohen, Tell, and Wallach 1986; Greene and Ondrich 1990; Hanley et al. 1990; Newman et al. 1990; Liu, Coughlin, and McBride 1991). Some studies have investigated associations between family size or composition and risks of nursing home admission or discharge, with somewhat mixed results. Spitze, Logan, and Robinson (1992) found no effect of the number of children on the probability of transitioning from living alone to being institutionalized, while Dwyer, Barton, and

Vogel (1994) found a significant negative effect of the number of children on the risk of nursing home admission, among persons living in cities and urban areas, but not rural residents. Garber and MaCurdy (1990) found significant effects of number of children on the probability of transitions into and out of nursing homes. Freedman (1993) found that women with *any* children returned from nursing homes at higher rates than did childless women, but found no such effect for men.

In our cross-sectional analysis, covariate effects reflect influences on the risk of nursing home entry, or the risk of live discharge from a nursing home, or both. While incidence estimates are preferable for purposes of analyzing lifetime use of nursing homes, prevalence estimates are also of interest, for example with respect to identifying target populations for various types of service delivery. When an elderly individual suffers from chronic limitations, the most common alternative to institutionalization is to obtain assistance from kin networks. Under this alternative the major costs of housing and caring for the impaired elderly relative are borne by family members. Our analysis is suggestive of the net effectiveness of offspring as alternatives to institutionalization.

Data and Variables*.* Our analysis uses as dependent variable a 0,1 indicator of whether an elderly person resides in a nursing home. Based upon the existing literature on the dynamics of institutionalization we introduce as explanatory variables several measures that have been shown to influence patterns of nursing home use.[4] Race is expected to influence the risk of living in a nursing home. Blacks are less likely than those in other races to live in nursing homes. This is due, in part, to their lower life expectancies, but also due to social factors which increase the access of blacks (especially black women) to the support network of kin groups. Marital status is also relevant to the analysis. Numerous studies have shown that married people have the lowest risk of institutionalization, while the never-married and formerly married have the highest risk. Geography is also considered important to the risk of institutionalization. Work by researchers such as Litwak and Longino (1987) describes patterns of elderly migration which start with amenity moves to the

South and West early in retirement, followed by a later return to the North or Midwest when functional limitations begin to require access to kin support networks. The degree of functional limitation is also central to the risk of residing in a nursing home. We expect that the more severe the degree of ADL limitations reported, the greater the risk of institutionalization. Age is included as an additional covariate, reflecting additional health and functional status needs not adequately captured by the ADL variables.

In our merged data file, all observations classified as disabled but noninstitutionalized come from the NLTCS sample. However, among observations classified as *institutionalized*, all of which also come from the NLTCS, not all are coded as disabled in the present analysis. Our model of nursing home residence includes indicators of difficulties with ADLs, but not with IADLs; nor does it include measures of disease conditions or cognitive functioning. Among the 1,354 cases living in nursing homes, 31 (over 2 percent) have no recorded ADL impairments.

The support network available to an older person also plays a central role in the risk of institutionalization. While a support network can be described with very complex models, research such as that presented by Rossi and Rossi (1990) has shown that a primary source of support revolves around the child-parent dyad. Within this set of relationships, the gender of the child moderates both the level of support and the type of support provided to the elderly parent (Wolf and Soldo 1988). Accordingly, we include in our analysis separate measures of the numbers of living sons and daughters. These variables represent information uniquely available in our merged data file, since the NLTCS alone does not provide such information for nondisabled respondents, while the NSFH alone excludes the institutionalized population.

Methods. We estimate a logistic regression equation for nursing home residence 16 times, using each imputed merged data file in turn. This process results in 16 independent sets of regression equations drawn from 16 independent data sets. After this task is completed, the set of 16 regression equations is summarized as a single regression equation modeling the risk of institutionalization, using the formulae presented earlier. As noted earlier, the sampling weights are approximately 10 times larger, on average, among observations contributed to the analysis from the NSFH file than among those from the NLTCS. Use of these weights is essential to the predictive accuracy of the estimated equation. However, standard statistical software fails to produce correct standard errors for weighted data; in this analysis we use an expression for the covariance matrix of parameters estimated using weighted maximum likelihood found in Manski and Lerman (1977). The resulting set of 16 covariance matrices is used to obtain the estimated model variance, as in equation (3).

Results. Table 5 presents estimates of logistic regressions for nursing home residence, estimated for men and women separately, using the merged NSFH/LTCS data set. The column headed "Parameter" contains averages of 16 separate estimates of each equation. The standard errors shown reflect a combination of model (within-round) and imputation (between-round) variance; these two components are also shown individually, illustrating the relative contribution of each source of variability to the standard errors.

As anticipated, age is positively related with the likelihood of residing in a nursing home. The risk of nursing home residence increases by 5 percent for each year of age for women, and by 3 percent per year for males. Blacks have a lower risk of residing in a nursing home, with black females being only 43 percent as likely as whites and others to be in a nursing home, while black males are only 38 percent as likely compared to whites and others. Marital status has a strong association with the likelihood of institutionalization. Female widows are 3 times as likely to be in a nursing home as currently married women while widowed men are 6 times as likely as currently married men. Formerly married (separated and divorced) women are 3 times as likely to be in a nursing home, while formerly married men are 7 times as likely to be in a nursing home. Never married females are 4 times as likely of being in a nursing home while never married males are 5 times as likely. All of these relationships are as expected with the exception of formerly married males, who have a greater likelihood of being in a nursing home than never married males. This finding might be explained by recent work by Goldscheider (1990), who suggests that divorced men are isolated from their children and correspondingly lacking in social support in later life.

Regional variation conforms to theoretical expectation as well. Both men and women who live in the Northeast and North Central regions of the United States are more likely to reside in a nursing home than those living in the South. The risk of institutionalization for those living in the West is not significantly different from those in the South. This finding conforms with the amenity migration hypothesis, as both the South and West of the United States are considered desirable retirement locations for the healthy and active elderly who are largely autonomous and possess portable sources of income.

The reported number of activity limitations proves to be the strongest predictor of residence in a nursing home. Females reporting 1 to 2 ADLs are 8 times more likely to be in a nursing home than women with no ADLs, while females reporting 3 to 4 ADLs are 23 times as likely and those reporting 5 to 6 ADLs are 107 times as likely to be in a nursing home when compared to females reporting no ADLs. A similar pattern is seen for men, with those reporting 1 to 2 ADLs being 12 times as likely to reside in a nursing home, those reporting 3 to 4 ADLs are 24 times as likely, and those with 5 to 6 ADLs are 104 times as likely to be in a nursing home compared to men reporting no ADLs.

The largest differences in the risks of residing in a nursing home between males and females are found in the effect that children play in the relative risks of institutionalization. For women, each additional son reduces her risk of residing in a nursing home by 13 percent, and each additional daughter reduces her risk by 23 percent. Conversely, the risk of residing in a nursing home increases by nearly 40 percent if a female has *no* living children. For men a different picture emerges. While each additional son reduces a man's risk of institutionalization by 29 percent, each additional daughter reduces this risk by only 4 percent, and the relationship lacks statistical significance. The risk of residing in a nursing home for males without living children is 66 percent higher than for those with one or more children of either sex.

These findings suggest that while both sons and daughters will contribute to keeping their mother out of a nursing home, only sons have the effect of reducing a father's chances of living in a nursing home. The results also suggest that women without children are better able to access alternative caregiving and support networks than are men. These findings are intriguing and will be explored in greater detail in future work.

Another important aspect of the model is the fact that the variance due to sampling variability of the model and the variance due to the imputation process can be separately evaluated. This is important, as failure to account for the imputation component of variability in parameter estimates can result in understated sampling errors and overstated confidence intervals. Table 5 indicates that the additional variance being introduced by the imputation process is quite small, in both absolute and relative terms, for all of the regression parameters.

CONCLUSIONS

The study of kinship relationships is assuming greater importance in the family demography literature. This new interest has led to an improved understanding of the role of noncoresident family members in the decision making process of other family members. Over the past 10 years we have seen a growth of new theory and terminology which incorporates kinship constructs into the discipline (Pullum and Wolf 1991; Rossi and Rossi 1990). This is particularly the case in the study of the elderly where the effects of external support networks can be central to autonomous living.

Unfortunately, the development of data sets that allow the study of kinship has proceeded more slowly. One response to data limitations is, of course, the collection of appropriate data, a time-consuming and expensive proposition. This paper has presented an alternative, namely the creation of a merged data set which is nationally representative of both institutionalized and noninstitutionalized elderly. The process of merging data is not a simple one and requires special techniques to provide an efficient sample which is representative of the population of interest. We have employed the multiple imputation methodology to overcome many of the statistical and methodological concerns related to the combination of unrelated sampling frames into a single analysis unit. Basic bivariate tests of aggregate weighted data has shown that our merged data accurately reflects the 1989 population of United States elderly 65 and older, and more sophisticated statistical tests have shown that the additional variance introduced into explanatory models employing the merged data set is minimal.

Our application of the data set employed a simple explanatory model for the cross-sectional risk of residing in a nursing home. While preliminary, it shows the value of being able to incorporate all elderly in the analysis of this risk, rather than restricting the contrast group to non-institutional elderly with a preexisting history of ADLs. Our analysis has suggested that, compared to women, men in general are much less likely to benefit from the presence of children as a support network. Women can expect their children to reduce their risk of institutionalization regardless of the child's gender, while men appear to benefit only from the existence of sons. Formerly married men have an even lower likelihood of having their children reducing this risk. These findings are informative, and support inadequately tested assumptions concerning the differential support of children to their mother and fathers.

The analysis presented only begins to exploit the potential richness of the merged data. As more variables from the NSFH and the NLTCS are standardized and added to the file, far more detailed analysis of kinship patterns and their consequences can be explored. It is our hope that this work will not only add to the expanding literature on family networks, but also encourage the use of similar methodological approaches to data limitations. Demographers have a tradition of maximizing information found in limited and deficient sources through the use of their understanding of population processes and their underlying stability. Our analysis represents an application of that understanding, which has allowed us not only to overcome a serious data limitation, but to take the vital step of assuming the limitation could be overcome in the first place.

TABLE 1

COMPONENTS OF 1989 NLTCS SAMPLE AND ROLE IN PREDICTIVE EQUATION

				Source of Covariates	
	Used in Predictive Model?	Number of Cases	Coding of Dependent Variable	Marital Status, Household Size, Number of Children	Whether ADL Difficulty
Automatically Screened In		3890		Follow-up	Follow-up
Nondisabled	Yes	668	0		
Disabled					
Noninstitutionalized	Yes	2210	1		
Institutionalized	No	1012	--		
Newly Screened In		1950		Follow-up	Screener
Nondisabled	Yes	300	0		
Disabled					
Noninstitutionalized	Yes	1308	1		
Institutionalized	No	342	--		
Screened Out	Yes	10,307	0	Screener	Screener

TABLE 2

ESTIMATED MODEL OF DISABILITY STATUS: NONINSTITUTIONALIZED NLTCS RESPONDENTS, 1989

	Full Model		Reduced Model	
Variable	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Intercept	-3.2148	0.1060	-3.1999	0.1057
Age	1.4489	0.1296	1.4494	0.1295
Age Squared	-0.1570	0.0483	-0.1574	0.0483
Male	-0.3710	0.0620	-0.3740	0.0620
Black	0.4877	0.0992	0.5002	0.0981
Other Race	0.1818	0.1305		
Widowed	0.2505	0.0673	0.2387	0.0665
Divorced	0.7661	0.1207	0.7519	0.1203
Separated	0.3882	0.2507		
Never Married	0.4709	0.1331	0.4610	0.1328
Marital Status Unknown	0.2728	0.4461		
Household Size	0.1682	0.0255	0.1679	0.0255
Noncoresident Children	0.0709	0.0135	0.0718	0.0135
ADL Difficulty	5.1643	0.1425	5.1637	0.1424
Northeast	-0.6140	0.0767	-0.6101	0.0767
North Central	-0.4503	0.0699	-0.4506	0.0699
West	-0.3543	0.0788	-0.3417	0.0784
Black*West	0.7693	0.3103	0.7515	0.3101

TABLE 3

COMPARISON OF BASELINE PREDICTIVE MODEL TO 16 ROUNDS OF MULTIPLE IMPUTATION REGRESSORS

Round	Intercept	Age	Age Squared	Male	Black	Widow	Divorced	Never Married
Observed	-3.1999	1.4494	-0.1574	-0.3740	0.5002	0.2387	0.7519	0.4610
1	-3.2805	1.3872	-0.2060	-0.4199	0.3086	0.2195	0.6355	0.3762
2	-3.1254	1.3516	-0.1304	-0.2405	0.5528	0.2054	0.6021	0.6312
3	-3.1624	1.3507	-0.1729	-0.3658	0.4613	0.2014	0.8700	0.3615
4	-3.3123	1.5605	-0.1209	-0.3906	0.4258	0.3047	0.9415	0.4558
5	-3.3842	1.6091	-0.1962	-0.3260	0.4614	0.3666	0.9047	0.5099
6	-3.2162	1.4553	-0.1620	-0.4004	0.6043	0.1911	0.7461	0.3436
7	-3.3266	1.5065	-0.1764	-0.3199	0.6437	0.2168	0.6515	0.7034
8	-3.1921	1.5226	-0.0852	-0.3453	0.4981	0.2763	0.6575	0.4860
9	-3.2849	1.5206	-0.0752	-0.3230	0.4550	0.1159	0.7180	0.3448
10	-3.4731	1.5553	-0.1695	-0.3223	0.3845	0.2512	0.7390	0.5480
11	-3.1615	1.5502	-0.1121	-0.4458	0.7405	0.2361	0.7474	0.2759
12	-3.1495	1.4962	-0.1322	-0.3587	0.4509	0.1736	0.6365	0.3885
13	-3.1035	1.4901	-0.1373	-0.3039	0.4004	0.2474	0.7904	0.5170
14	-2.9422	1.6164	-0.1298	-0.4079	0.4066	0.2869	0.7657	0.5379
15	-3.0257	1.3156	-0.1596	-0.4261	0.3487	0.2035	0.6842	0.4121
16	-3.3793	1.5976	-0.0946	-0.4676	0.5893	0.0958	0.5468	0.2274
Average	-3.2200	1.4928	-0.1413	-0.3665	0.4832	0.2245	0.7273	0.4449

TABLE 3 (CONT.)

Round	People in Home	Children Out of Home	Can’t Walk	Northeast	North Central	West	Cross 54
Observed	0.1679	0.0718	5.1637	-0.6101	-0.4506	-0.3417	0.7515
1	0.0699	-0.0350	5.1154	-0.5040	-0.4071	-0.2565	0.9602
2	0.1767	0.1275	5.0495	-0.6046	-0.4522	-0.4725	0.8262
3	0.1135	0.0263	5.2458	-0.4923	-0.3906	-0.4041	0.7997
4	0.1976	0.1471	5.2668	-0.7559	-0.4886	-0.3691	0.7366
5	0.2185	0.0895	5.1402	-0.6331	-0.4654	-0.5318	1.2738
6	0.1624	0.0557	5.1158	-0.5519	-0.4602	-0.1995	0.2578
7	0.1838	0.0896	5.0892	-0.6864	-0.4499	-0.1809	0.0522
8	0.2053	0.1278	5.3219	-0.7625	-0.5228	-0.3227	0.8424
9	0.1308	0.0621	5.3023	-0.5664	-0.4085	-0.5004	1.0639
10	0.1662	0.0776	5.1778	-0.6357	-0.5011	-0.3185	0.9275
11	0.1469	0.0826	5.1190	-0.6459	-0.5340	-0.2346	0.7417
12	0.1912	0.0457	5.0188	-0.5844	-0.4145	-0.4053	0.9396
13	0.2048	0.1156	5.3021	-0.6285	-0.5422	-0.4336	0.9472
14	0.2283	0.1190	5.2864	-0.6514	-0.5133	-0.4044	0.5322
15	0.1434	0.0397	5.2110	-0.4349	-0.5007	-0.3882	1.1840
16	0.1328	-0.0101	5.2893	-0.5891	-0.3123	-0.3156	0.8117
Average	0.1670	0.0725	5.1907	-0.6079	-0.4602	-0.3586	0.8060

TABLE 4

WEIGHTED COUNTS OF ELDERLY POPULATION: NLTCS, MERGED NLTCS/NSFH, AND CENSUS BUREAU ESTIMATES FOR 1989 (in thousands)

				Absolute Difference		Percentage Difference	
Age	**1989 Census Estimates**	**LTCS Data Frequency**	**Merged Data Frequency**	**LTCS versus Census**	**Merged versus Census**	**LTCS versus Census**	**Merged versus Census**
				Females			
65 to 74	10,040	9,279	9,946	-761	-94	-0.076	-0.009
75 to 84	6,171	6,893	6,124	722	-47	0.117	-0.008
85 and above	2,137	2,469	2,222	332	85	0.155	0.040
	18,348	18,642	18,292	294	-56	0.016	-0.019
				Males			
65 to 74	7,824	7,182	7,399	-642	-425	-0.082	-0.054
75 to 84	3,678	4,143	3,486	465	-192	0.126	-0.052
85 and above	831	835	882	4	51	0.004	0.061
	12,333	12,160	11,767	-173	-566	-0.014	-0.032
				Total			
65 to 74	17,864	16,461	17,345	-1,403	-519	-0.079	-0.029
75 to 84	9,849	11,036	9,610	1,187	-239	0.121	-0.024
85 and above	2,968	3,304	3,104	336	136	0.113	0.046
	30,681	30,801	30,059	120	-622	0.004	-0.024

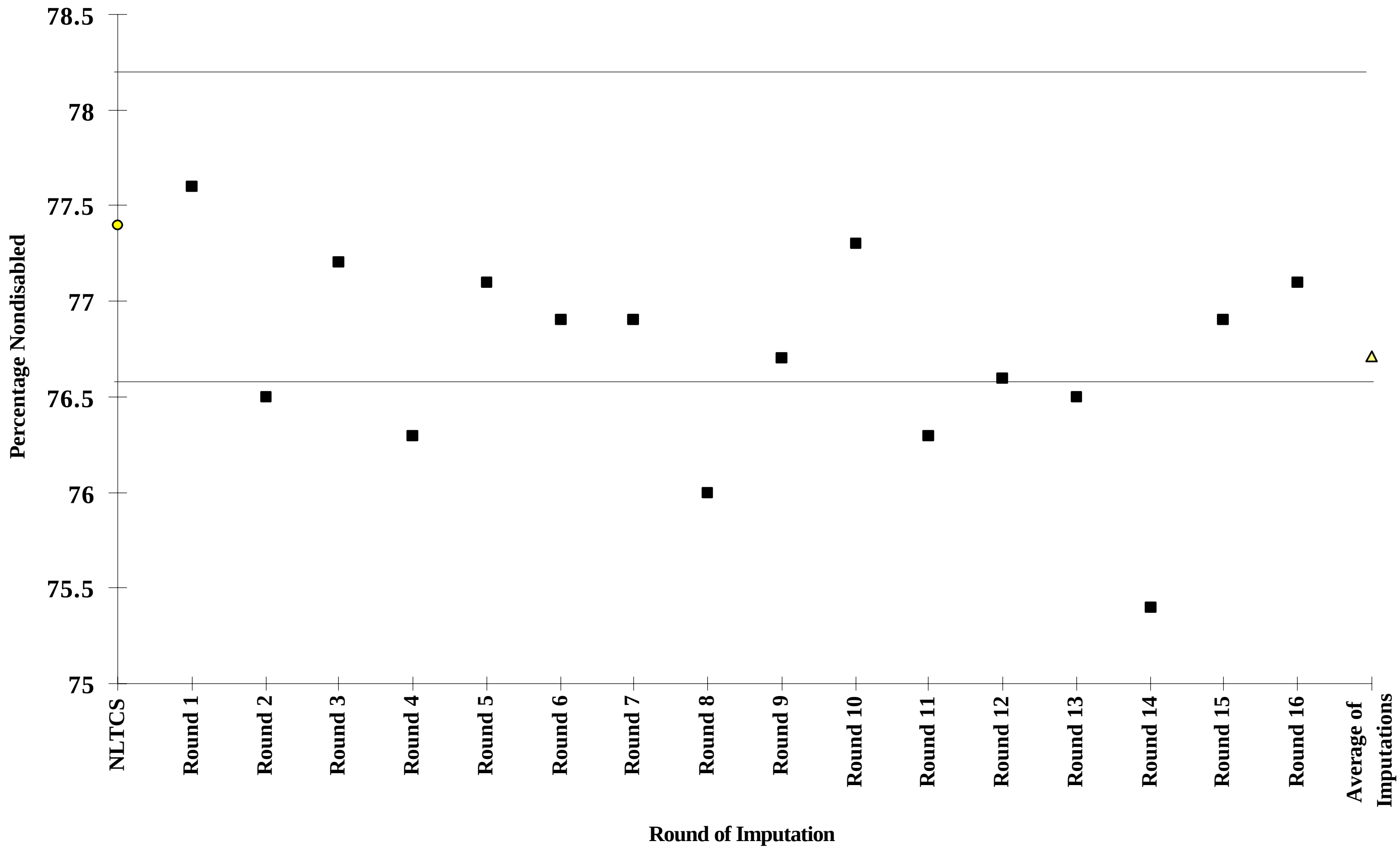
TABLE 5

LOGISTIC MODEL FOR THE LIKELIHOOD OF RESIDING IN A NURSING HOME: MERGED NSFH/LTCS SAMPLE OF ADULTS AGED 65 AND OLDER

Variable	Parameter	SE	Parameter ÷ SE	Imputation Variance (% of total)	Odds Ratio
			Females		
Intercept	-10.668	0.556	-19.198	0.376	
Age	0.048	0.007	7.165	0.156	1.049
Black	-0.817	0.169	-4.834	0.008	0.442
Widowed	1.044	0.139	7.493	0.006	2.841
Never Married	1.366	0.233	5.861	0.059	3.921
Divorced	1.098	0.215	5.107	0.022	2.999
West	0.064	0.141	0.455	0.028	1.066
Northeast	0.280	0.129	2.171	0.032	1.323
North Central	0.339	0.119	2.840	0.020	1.403
1-2 ADLs	3.887	0.261	14.892	3.862	48.763
3-4 ADLs	4.891	0.258	18.927	3.983	133.080
5-6 ADLs	6.476	0.254	25.519	4.140	649.176
Number of Sons	-0.137	0.050	-2.725	0.010	0.872
Number of Daughters	-0.258	0.048	-5.374	0.011	0.773
No Living Children	0.333	0.140	2.374	0.013	1.395
			Males		
Intercept	-9.716	0.926	-10.498	0.079	
Age	0.031	0.012	2.683	0.073	1.031
Black	-0.969	0.319	-3.032	0.008	0.380
Widowed	1.801	0.182	9.895	0.025	6.057
Never Married	1.611	0.367	4.385	0.074	5.007
Divorced	1.978	0.323	6.117	0.041	7.230
West	-0.090	0.237	-0.382	0.007	0.914
Northeast	0.250	0.217	1.153	0.021	1.284
North Central	0.673	0.208	3.230	0.039	1.961
1-2 ADLs	4.373	0.353	12.403	1.213	79.275
3-4 ADLs	5.065	0.355	14.266	1.212	158.372
5-6 ADLs	6.595	0.343	19.214	1.250	731.647
Number of Sons	-0.331	0.074	-4.505	0.019	0.718
Number of Daughters	-0.036	0.071	-0.504	0.018	0.965
No Living Children	0.472	0.251	1.879	0.082	1.603

FIGURE 1

**PERCENTAGE NONDISABLED AMONG PERSONS 65 AND OLDER, 1989:
16 INDEPENDENT REPLICATIONS OF MERGED NSFH/NLTCS FILE**



ENDNOTES

* James McNally is a Post-Doctoral Fellow associated with the Maxwell Center for the Demography and Economics of Aging, and Douglas A. Wolf is Professor of Public Administration and Director of the Maxwell Center for the Demography and Economics of Aging at Syracuse University. An earlier version of this paper was presented at the annual meetings of the Population Association of America held in San Francisco April 6-8, 1995. Helpful comments on this research have been received from Daniel Hill and J.S. Butler.

1. The merged file remains incomplete for the *entire* United States population as institutionalized persons under 65 are not represented.

2. The data include a base sample of over 9,500 respondents, as well as several oversamples selected on the basis of minority group membership or the presence of selected family attributes (including single-parent status, families with step-children, cohabitors, and the recently married).

3. Prior work with multiple imputation using the SAS statistical package suggests that the inclusion of nonsignificant regressors with large standard errors can significantly distort multiple imputation results.

4. Most of the covariates are defined as in the preceding analysis of disability status.

REFERENCES

Cohen, Marc A., Eileen J. Tell, and Stanley S. Wallach. 1986. "Client-Related Risk Factors of Nursing Home Entry among Elderly Adults," *Journal of Gerontology*, 41: 785-792.

Dwyer, Jeffrey W., Amy J. Barton, and W. Bruce Vogel. 1994. "Area of Residence and the Risk of Institutionalization," *Journal of Gerontology: Social Sciences*, 49: S75-84.

Freedman, Vicki. 1990. "Using SAS to Perform Multiple Imputation," UI-PSC-6. Washington, DC: Urban Institute, November.

_____. 1993. "Kin and Nursing Home Lengths of Stay: A Backward Recurrence Time Approach," *Journal of Health and Social Behavior,* 34: 138-152.

Freedman, Vicki and Douglas Wolf. 1995. "A Case Study in Multiple Imputation," *Demography*, 32: 459-470.

Garber, Alan M. and Thomas MaCurdy. 1990. "Predicting Nursing Home Utilization among the High-Risk Elderly." In D.A. Wise (ed.), *Issues in the Economics of Aging*. Chicago: The University of Chicago Press, pp. 173-200.

Goldscheider, Frances. 1990. "The Aging of the Gender Revolution: What Do We Know and What Do We Need to Know?" *Research on Aging*, 12(4): 531-545.

Greene, Vernon L. and Jan I. Ondrich. 1990. "Risk Factors for Nursing Home Admissions and Exits: A Discrete-Time Hazard Function Approach," *Journal of Gerontology: Social Sciences*, 45: S250-258.

Hanley, Raymond J., Lisa Maria B. Alecxih, Joshua M. Wiener, and David L. Kennell. 1990. "Predicting Nursing Home Admissions: Results from the 1982-1984 National Long-Term Care Survey," *Research in Aging*, 12: 199-228.

Herzog, Thomas N. and Donald B. Rubin. 1980. "Using Multiple Imputation to Handle Nonresponse." In William G. Madow, Ingram Olkin, and Donald B. Rubin (eds.), *Incomplete Data in Sample Surveys*, Vol 2. New York: Academic Press, pp: 210-248.

Hogan, Dennis, Ling-Xin Hao, and William Parish. 1990. "Race, Kin Networks, and Assistance to Mother-Headed Families," *Social Forces*, 86(3): 797-812.

Litwak, Eugene and Charles F. Longino. 1987. "Migration Patterns among the Elderly: A Developmental Perspective," *The Gerontologist*, 27: 266-272.

Liu, Korbin, Teresa Coughlin, and Timothy McBride. 1991. "Predicting Nursing-Home Admission and Length of Stay," *Medical Care*, 29: 125-141.

Manski, Charles F. and Steven R. Lerman. 1977. "The Estimation of Choice Probabilities from Choice Based Samples," *Econometrica*, 45: 1977-1988.

Manton, Kenneth A. 1988. " A Longitudinal Study of Functional Change and Mortality in the Untied States," *Journal of Gerontology,* 43: S153-161.

Manton, Kenneth G., Larry S. Corder, and Eric Stallard. 1993. "Estimates of Change in Chronic Disability and Institutional Incidence and Prevalence Rates in the U.S. Elderly Population from the 1982, 1984 and 1989 National Long Term Care Survey," *Journal of Gerontology: Social Sciences*, 48(4). S153-S166.

McBride, Timothy D. 1989. " Measuring the Disability of the Elderly: Empirical Analysis and Projections into the 21st Century." Presented at the Annual Meetings of the Population Association of America, Baltimore, Maryland, March 31, 1989.

McNally, James W. 1989. "The Characteristics of Nonrespondents: A Preliminary Model Based on Examination of the National Survey of Family Growth, Cycle IV." Unpublished Master's Thesis, Georgetown University, Washington DC.

____. 1994. "The Indirect Estimation of Parental Mortality: An Extension of Orphanhood Techniques to Maximum Likelihood Estimates with Bayesian Applications." Unpublished Ph.D. dissertation, Brown University.

Newman, Sandra J. and Raymond Struyk, with Paul Weight and Michelle Rice. 1990. " Overwhelming Odds: Caregiving and the Risk of Institutionalization," *Journal of Gerontology: Social Sciences*, 45: S173-183.

Pullum, Thomas and Douglas Wolf. 1991. "Correlations between Frequencies of Kin," *Demography*. August 28(3): 391-409.

Rossi, Alice and Peter Rossi. 1990. *On Human Bonding: Parent-Child Relations across the Life Course*. New York: A. Deruyter Press.

Rubin, Donald R. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley and Sons.

Spitze, Glenna, John R. Logan, and Joyce Robinson. 1992. "Family Structure and Changes in Living Arrangements among Elderly Nonmarried Parents," *Journal of Gerontology: Social Sciences*, 47: S289-296.

Sweet, James, Larry Bumpass, and Vaughn Call. 1988. "The Design and Content of the National Survey of Families and Households." Working Paper NSFH-1, Center for Demography and Ecology, University of Wisconsin-Madison.

Treiman, Donald, William Bielby, and Man-tsun Cheng. 1988. "Evaluating a Multiple-Imputation Method for Recalibrating 1970 U.S. Census Detailed Industry Codes to the

1980 Standard." In *Sociological Methodology*, Vol. 18. American Sociological Publications.

Weissert, William and William Scanlon. 1983. "Determinants of the Institutionalization of the Aged." Working Paper Number 1466-21 (revised). Washington, D.C.: The Urban Institute.

Wolf, Douglas and Beth J. Soldo. 1988. "Household Composition Choices of Older Unmarried Women," Demography, August 25(3): 387-404.

Center for Policy Research
Tim Smeeding, Director
Richard V. Burkhauser, Associate Director for Aging Studies

Center for Demography and Economics of Aging
Douglas A. Wolf, Director

The Maxwell Center for Demography and Economics of Aging was established in September 1994 with core funding from the National Institute on Aging (Grant No. P20-AG12837). The goals of the Center are to coordinate and support research activites in aging, publish findings and policy analyses, and convene workshops, seminars, and conferences intended to train data users and disseminate research findings and their implications to the research and policy communities.

The Center produces two paper series, the Aging Studies Paper Series and Papers in Microsimulation.

Many Center publications are available via anonymous ftp. To access these resources attach to our World-Wide Web home page at http://gerosun.syr.edu.

For more information contact: Martha Bonney (mbonney@maxwell.syr.edu).

Center for Policy Research
426 Eggers Hall
Syracuse University
Syracuse, New York 13244

(315) 443 3114

FAX: (315) 443 1081

PAPERS ON MICROSIMULATION SERIES

Paper No.	Title	Authors	Date
1	A Model for Simulating Life Histories of the Elderly: Model Design and Implementation Plans	Wolf, Ondrich, Manton, Stallard, Woodbury, and Coder	August 1995
2	Family Structure and Institutionalization: Results from Merged Data	McNally and Wolf	January 1996