Title
Estimating the Population of Rural Communities by Age and Gender: A Case Study of the Effectiveness of the Local Expert Procedure

Journal
Small Town, 25(6)

Authors
Swanson, David A
Carlson, John
Roe, Linda
et al.

Publication Date
1995

Peer reviewed

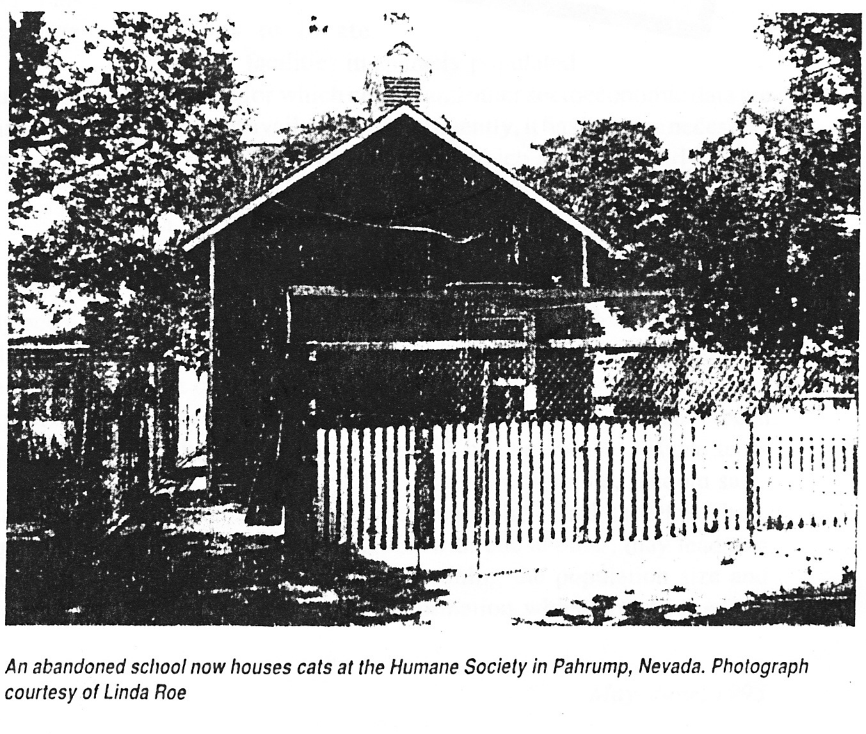
Small Towns Institute
Post Office Box 517 Ellensburg, Washington 98926

David Swanson
Center for Population Research
and Census
Portland State University
P.O. Box 751
Portland OR 97207-0751

NON-PROFIT
ORGANIZATION
U.S. POSTAGE
PAID
ELLENSBURG
WASHINGTON
PERMIT NO. 154

In This Issue...

Volume 25, No. 6 ***May–June, 1995***



An abandoned school now houses cats at the Humane Society in Pahrump, Nevada. Photograph courtesy of Linda Roe

Estimating the Population of Rural Communities by Age and Gender:

A Case Study of the Effectiveness of Local Expert Procedure

by David Swanson, John Carlson, Linda Roe and Christopher Williams

David Swanson was Senior Demographic Specialist at the Institute for Economic Advancement, University of Arkansas at Little Rock when he wrote this article. Effective September 1, 1995, Swanson will be at the Center for Population Research and Census, School of Urban and Public Affairs, Portland State University, P.O. Box 751, Portland, Oregon, 97207-0751. John Carlson and Linda Roe are both Researchers with the Science Applications Corporation, which conducts research on the socioeconomic impacts associated with high-level nuclear waste repositories. Chris Williams, formerly with Science Applications Corporation, is currently a Researcher with the Washington State Department of Social and Health Services. An earlier version of this paper was presented at the 1993 Public Health Conference on Records and Statistics, Intercensal Population Estimates session, held in Washington, D. C.

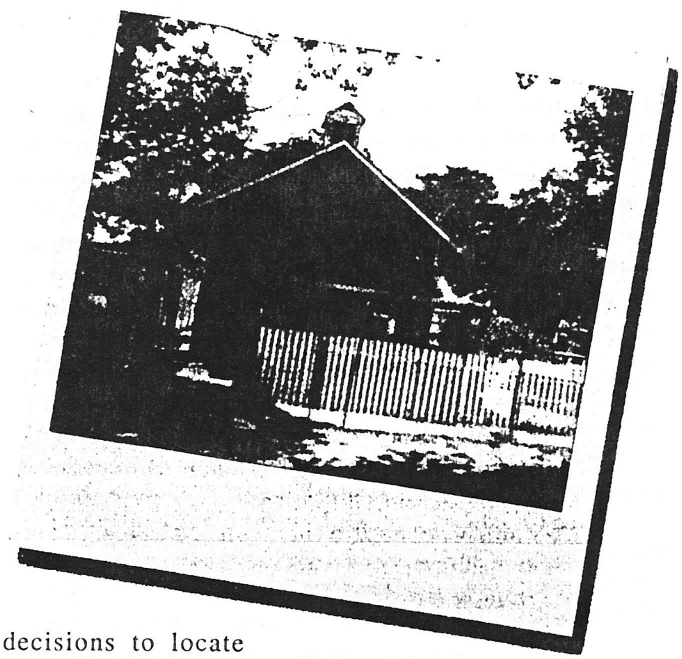
This article examines the accuracy of a survey-based technique called the Local Expert Procedure (LEP), for estimating selected demographic characteristics of small, rural areas. The procedure is a variation of the Housing Unit Method (HUM) and employs local citizens to provide demographic information about households which were randomly selected from a residential sample extracted from utility records.

The procedure is nonintrusive and less costly than traditional survey data collection efforts. And, because it is based on random sampling, confidence intervals can be constructed around the estimates generated by this technique.

Introduction

The Local Expert Procedure (LEP) provided highly accurate estimates of the total population in three small, rural Nevada communities: Beatty, Amargosa Valley and Pahrump. However, the LEP had not been tested for its accuracy when estimating the population *by age and gender* in these same three communities.[1]

Several reasons exist for wanting to develop a cost-effective, nonintrusive method for estimating the demographic characteristics of small, rural communities. For example, concerns over legal and environmental issues have resulted in



decisions to locate unpopular facilities in sparsely populated rural areas for which census and other socioeconomic data are usually not available.[2] Consequently, it has become necessary to develop methods of inquiry which are particularly suited for small, rural areas, and which fully exploit the available data, are less costly and, in many cases, less intrusive than area, telephone and mail surveys. We believe that the LEP contributes to this type of methodological development.

One of the strong features of the LEP is that it is based on the housing unit method, which allows it to be adapted to a range of data environments.[3]

The LEP combines analytical techniques that have been largely developed in isolation of one another. Those techniques are the HUM of population estimation, random sampling and key informant ethnography (local experts). These procedures, in combination with one another, may lead to a very powerful means of obtaining the population size and household composition information which are required for

TABLE 1
Sample Characteristics and Results of the Accuracy Test For The Estimated Total Population

1990

Community	Households	Sample Estimates PPH	SE	Population	1990 Census Count	Confidence Interval Low	High
Amargosa Valley	326	2.58	0.11	841	853	771	911
Beatty	672	2.43	0.10	1,633	1,623	1,501	1,765
Pahrump	3,224	2.23	0.06	7,190	7,424	6,810	7,569

most impact assessment projects. Since the LEP relies primarily on people (the local experts) and administrative records (active utility records) that are available for any geographic area, it can, in principle, be applied to any small, rural area for the development of population estimates.

One strong feature of the LEP is its connection with sampling theory, which allows a person to make probabilistic statements about population estimates made using the LEP. This feature is virtually unique among techniques used for estimating population. With it, one can construct confidence intervals around an LEP population estimate (something we, in fact, do in this article). This gives a user some idea of the precision inherent in a given estimate.

Take, for example, the 95 percent confidence interval for the estimated number of persons ages 0 to 18 years in Amargosa Valley, which provides a low of 198 and a high of 324, a range of 126 persons (see Table 3). This is a much wider range than the 95 percent confidence interval for the estimated number of persons aged 19-39 in Amargosa Valley, which has a low of 171 and a high of 255—a range of only 84 persons. In this comparison, we can see that the estimate for the persons aged 19-39 is more "precise" than the estimate for those aged 0 to 18 years. In other words, we have evidence that our estimate for the number aged 19-39 is probably closer to the true number than is the estimate for those aged 0 to 18.

Study Area

The research reported here is part of a comprehensive program to assess the socioeconomic characteristics of communities located near Yucca Mountain, Nevada—the site currently being evaluated for the nation's first geological repository for spent nuclear fuel and high-level radioactive waste. During 1989, a socioeconomic plan for the Yucca Mountain Project was developed to ensure that adverse socio-economic impacts (which may result in Southern Nevada from the scientific investigations program) are avoided or minimized to the maximum extent possible.[4] The data used in this study are required in order to monitor the communities nearest Yucca Mountain for any changes in the size or composition of the population as the scientific investigations continue. The data reported here were collected in the late spring of 1990, which makes them comparable in time to data from the 1990 Census. The study area data are also consistent with 1990 census geography, concepts and definitions.

The communities included in this study are located in southern Rye County, Nevada. Each is distinct, both geographically and demographically.[5] For example, the bulk of Beatty's 1,600 citizens is contained within a 1-square mile area. By contrast, Amargosa Valley is a community of nearly 900 people scattered over 500 square miles. Pahrump has the largest population of the three communities, with over 7,400 residents living in a 280-square mile area.[6]

Differences in age composition also exist. According to the 1990 census, nearly a quarter of Pahrump's population (21.5 percent) is 65 years of age and older. By contrast, only 6.3 percent of Beatty's population and only 10.2 percent of Amargosa Valley's residents fall into this age category.[7] Pahrump is the only one of three communities believed to have experienced significant growth since 1990. From 1990 to 1992, Pahrump experienced an increase of 702 households.[8]

Using the standard nonlinear method for calculating average annual rates of change and the 1990 household count of 3,555, the increase of households in Pahrump represents a 9.5 percent average annual rate of change. Beatty, on the other hand, experienced an increase of only 85 households—a 5.5 percent average annual rate of change. Amargosa Valley showed an increase of only 2 occupied units (0.3 percent average annual rate of change) during the same period.[9]

TABLE 2
Estimated Number of Males by Community, 95% Confidence Intervals for the Estimates, and 1990 Census Counts of Males

Community	Households	1990 Sample Estimates Male PPH	SE	Males	1990 Census Count	Confidence Interval Low	High
Amargosa Valley	326	1.32	0.06	430	436	389	475
Beatty	672	1.23	0.05	827	914	756	908
Pahrump	3,224	1.09	0.03	3,514	3,700	3,316	3,713

Methodology

The sample selection and data collection techniques used in this research were designed to ensure the development of a demographic data base that accurately represents each community. A random selection of households, based on the coding system that identified each element by location and housing type, was made separately for each community. Sample sizes were based on the total number of households in the community.

The formula used to calculate each sample assumed a 5 percent margin of error, an alpha level of .05 (this corresponds to a 5 percent change of having the true value of a number we are estimating fall outside the 95 percent confidence interval) and a dichotomous variable (a variable with only two values) with an estimated proportion of 0.5. This strategy, of course, is conservative since it accepts the "cost" of a sample size that is likely to be larger than required for the precision we desire. Once the sampling size was determined, an additional 15 percent was added to allow for the possibility of having to exclude households from the sample for which data are unknown.

During a preliminary phase of the research, contacts were made with elected officials and other community leaders. This move resulted in the development of a network through which local experts were initially identified on the basis of their experience in community activities. Each potential expert was interviewed by project researchers and asked to complete a form designed to assess their interests and qualifications. A written explanation of the project and specific instructions regarding the data collection were provided to the candidates. Those selected then received information regarding the conditions of their temporary employment (pay, working hours, etc.) and instructions on confidentiality. For this project, we found that the meter readers employed by local utilities constituted an interested and knowledgeable source of local experts.

During the actual data collection process, the local experts were instructed not to identify households by name in order to ensure the anonymity of community members (even though the local experts know from the utility records the householder's name under which the sampled unit's account was maintained). For each household, local experts were asked to record, or to communicate to the researcher, the total number of persons residing in the household as of July 15, 1990, the age category and gender of each household member and the retirement status of persons 50 years and older.

Eight age categories were used. The first two distinguished preschool (0-4) from school aged children (5-18). The next three accounted for persons normally considered active in the work force. These categories include ages 19-29, 20-39 and 40-49. Retirement status was added as a variable for each of the last three categories, 50-59, 60-64 and 65 and over.

Data were collected on a form that listed and identified the sample households' meter number assigned to the unit and the type of housing. All structures, including those identified as "burned down" or otherwise destroyed, unoccupied or "removed from the pad," were considered as part of the final sample. Structures identified as "not a residence" or structures for which data were unknown were not included in the final sample. Again, detailed information on this part of local expert procedure is found elsewhere.

The data representing the age and gender distribution for each household were compiled from the original data collection forms. Data for age were then collapsed into four categories: 0-18 (preschool and school age), 19-39 (labor force and primary child-bearing years), 40-64 (labor force and preretirement) and 65 and over (retirement). These age groups are comparable to the age groupings reported in the 1990 census data.

The actual estimation is based on the Housing Unit Method equation, but modified for a given age or gender segment as follows:

Pi = [H*O*PPHi] + GQi

where

Pi = population in age group (or gender group).

H = total housing units.

O = proportion of occupied units.

PPHi = mean number of persons per household in age group i (or gender group i) as found through the IEP sample survey.

GQi = number of persons in age group i (or gender group i) in group quarters as found through LEP procedure.

An important aspect of the evaluation was to construct confidence intervals as illustrated below. Age is used an example.

First, the arithmetic mean and standard error were computed for a given age category. Second, once the standard error was computed, lower and upper confidence intervals were estimated using the following formulas:

upper limit = (N)*[x+(tn-2, a/2)]*(s.e.)

lower limit = (N)*[x+(tn-2, a/2)]*(s.e.)

where:

N = total households in the community.

x = mean number per sampled household for the age category.

s.e. = standard error.

tn-2 = (a/2)th percentile of the t distribution, with (n-2) degrees of freedom.

Findings

Table 1 provides a review of the results from the previous study which found the LEP capable of providing accurate estimates of the total household population. Included in the table are the total number of households in each community, the estimated persons per household (PPH) generated from the sample data, the standard error of the estimated PPH, the estimated total household population, the 1990 Census count of population, and the confidence intervals for the total household population estimates. These data show that the population counted in the 1990 census for all three communities fit within the 95 percent confidence intervals generated around the estimates. This suggests that the LEP does a good job in terms of accuracy for estimating total population.

Table 2 shows the estimated average number of males per household by community and the confidence interval for each estimate. These estimates are then compared with the census data in terms of the number of males in the total household population.[10] The data show that in both Amargosa Valley and Pahrump the census count for males is within the confidence intervals. In Beatty, the census count is slightly higher (6 males) than the high end of the confidence interval. Again, these results suggest that the LEP is accurate for estimating gender.

The computed confidence limits of the estimates by age category are presented in Table 3 for each community. The actual 1990 census count is indicated to the right of the confidence limits. Overall, the figures show that the estimates are more accurate for the lower age groups than the higher age groups.

The census counts for Amargosa Valley and Beatty, the

TABLE 3
Estimated Population by Age and Community, 95% Confidence Intervals for the Estimates, and 1990 Census Counts by Age

	Amargosa Valley			Beatty			Pahrump		
	Confidence Interval of Estimate			Confidence Interval of Estimate			Confidence Interval of Estimate		
Age Group	Low	High	1990 Census Count	Low	High	1990 Census Count	Low	High	1990 Census Count
0-18	198	324	267	367	571	424	789	1,409	1,663
19-39	171	255	245	601	762	609	1,277	1,811	1,571
40-64	293	378	252	360	512	487	3,690	4,190	2,597
65+	17	53	89	18	73	103	459	812	1,593

TABLE 4
Summary Results of the 1990 Age and Male (Gender) Estimates and Number of Local Experts by Community

	AGE GROUP (Census count is within 95% confidence interval of the estimate)		MALES (Census count is within 95% confidence interval of the estimate)	Size of the Community (1990 Census count)	Number of Local Experts
Amargosa	0-18	Yes			
Valley	19-39	Yes	Yes	853	2
	40-64	No			
	65+	No			
Beatty	0-18	Yes			
	19-39	Yes	Yes	1,623	4
	40-64	Yes			
	65+	No			
Pahrump	0-18	No			
	19-39	Yes	Yes	7,424	1
	40-64	No			
	65+	No			

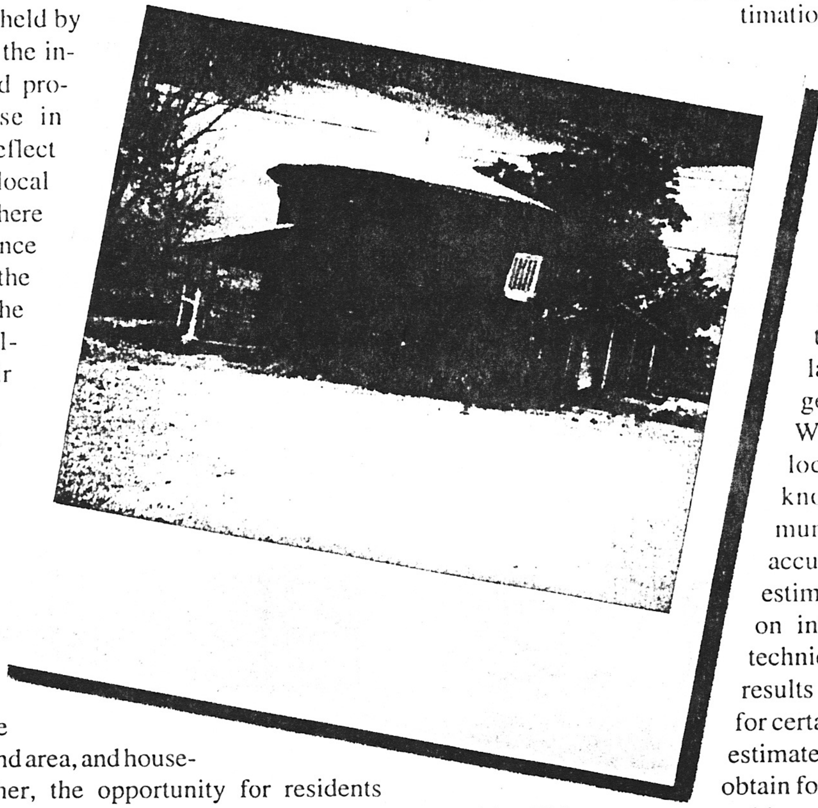
two smallest communities, are within the confidence intervals for the two lowest age groups (0-18 and 19-39). In the highest age group (65 and over), the estimates for both communities are too low for the census to count in order to fall within confidence intervals. In Pahrump, the confidence interval for the highest age group (65 and over) is also too low to include the census count and too high for the age group 40-64. Unlike Amargosa Valley and Beatty, the census count for the lowest age group (0-18) in Pahrump does not fall within the confidence intervals. The only instance where census count falls within the confidence interval in Pahrump is for age group 19-39. Here, the findings are mixed in terms of accuracy. For younger age groups, the LEP seems to be more accurate than for older age groups.

Summary and Discussion of the Findings

Table 4 provides a summary of the findings along with the number of local experts used in each community and the size of the community. When compared to the total population and number of local experts, the findings point to a number of factors that could potentially influence the accuracy of the LEP for estimating age and gender.

The first factor presumed to have affected the accuracy of the estimates in this study is the size of the local expert team relative to the size of the community. In the smallest communities, where more than one local expert was employed (Amargosa Valley and Beatty), the estimates were generally more accurate than the estimates generated for Pahrump, where only one local expert was ultimately used (the others originally hired had to discontinue). The highest number of local experts (four) was employed in Beatty, which had accurate age estimates for three of the four age categories. The estimated number of males in Beatty was also very close to the census which reported 6 more males than the upper limit of the confidence interval. In Amargosa Valley, data collected from two local experts produced accurate estimates for two of the four age groups and for the distribution of males. In Pahrump, only one age group (19-39) and the distribution of males was estimated accurately.

Research on decision making in groups indicates that accuracy increases when there is a variance in the information held by group members and all the information is shared and processed.[11] This increase in accuracy is believed to reflect the experience of the local expert group in Beatty. There appeared to be less variance in information held by the two local experts in the Amargosa Valley, although they shared their information openly.



A second factor that could explain why the estimates are generally more accurate in Beatty is that, compared to Amargosa Valley and Pahrump, Beatty appears to be a more tightly-knit community. Since the community is small in land area, and households are closer together, the opportunity for residents to interact is greater. This factor is presumed to influence the extent and nature of each local expert's knowledge of households in the community. It fits within a long tradition of sociological research and theory regarding social integration.[12]

The presence of identifiers associated with preschool and school-age children appears to have been an important factor in estimating the number of persons in the household 0-4 and 5-18 years of age. In all three communities, the estimates of the age distribution show a pattern of greater accuracy in the lower age categories. The absence of a strong identifier to differentiate the number of persons 65 years of age and older from those 40-64 may account for the fact that the census counts for the oldest age fall outside the confidence intervals.

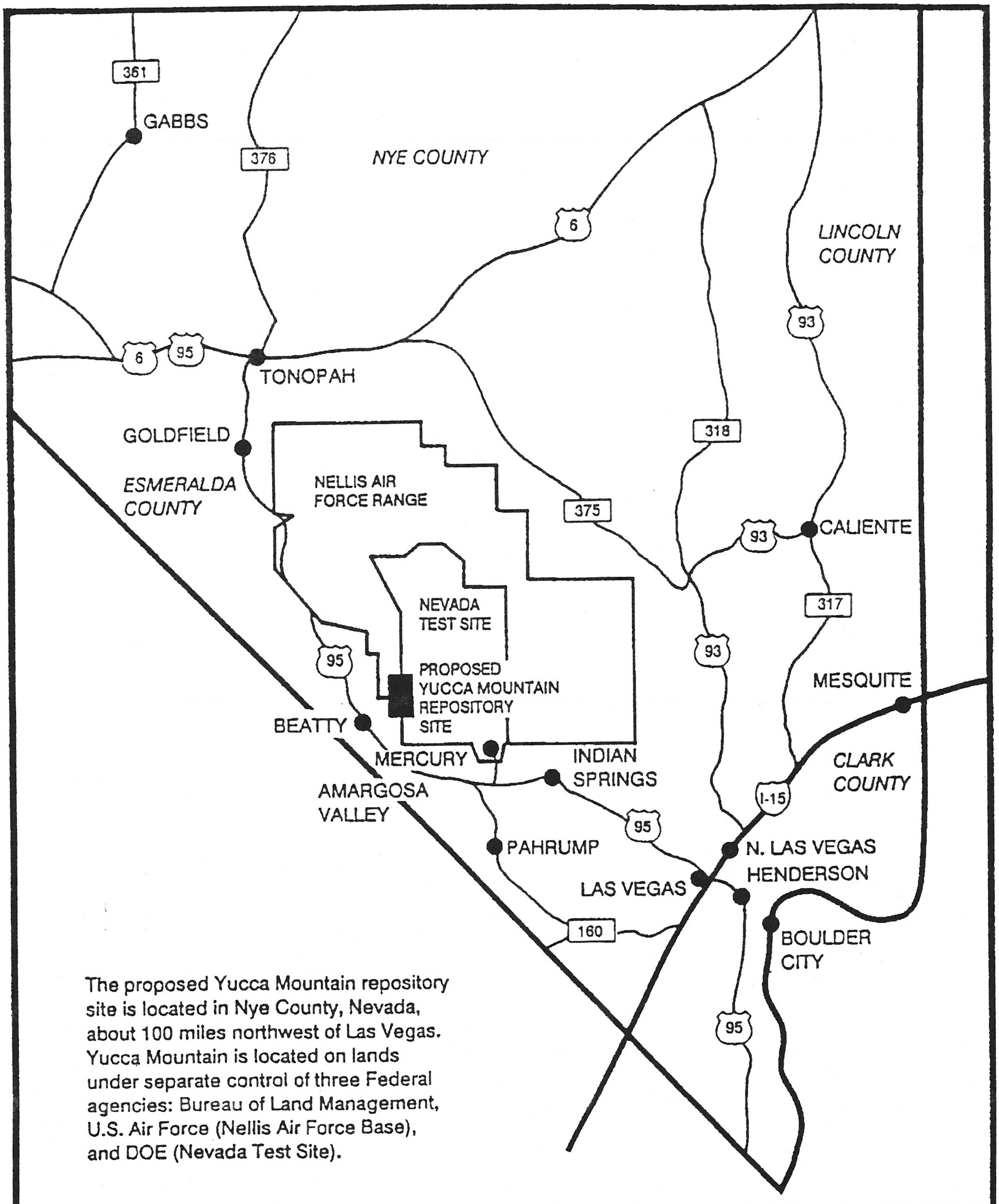
At this point, judgment was used by the local experts to develop the demographic characteristics of each sampled household. Research indicates that the development of "rule-based procedures," or a predefined, rigid set of criteria in the form of a decision tree that leads to a conclusion, may improve accuracy.[13] In this regard, follow-up work with the local experts used in this study would be useful. Accuracy may also improve if the subjective level of uncertainty held by the local experts is better understood, particularly with regard to estimating the age of those who are 65 and over. Research suggests that desirability increases positively with uncertainty.[14] Since there was a tendency to underestimate the age of those 65 and over, and at the same time overestimate those aged 40-64, it may be that social values concerning aging may be leading to the underestimation of the age of those in the initial years of their retirement age since "youthfulness" is more socially desirable than "agedness."[15]

The results of this study suggest that the extension of the LEP to the estimation of population composition (age, gender) holds promise. With a careful selection of local experts and good knowledge of the community in question, both accurate and cost-effective estimates which are not based on intrusive data-collection techniques are possible. The results also show that, at least for certain age groups, accurate estimates are more difficult to obtain for rapidly growing communities. This problem occurs because many people move into a community who are not known to the local experts. Further analysis of the factors believed to affect the accuracy of the estimates is suggested.

The experience of working with a range of local expert group sizes, coupled with research on the accuracy of group decision-making, suggests that local experts be recruited with an eye toward maintaining variance. That would mean assembling a group in which each member has a high level of community knowledge, but where their individual knowledge is based on different community experiences and overlaps only partially with that of others in the group. Once such a group is assembled, the research suggests that the members receive training in the open sharing and processing of information, as well as in the rules developed for decision-making, once a workable set of rules is created.[16]

It may even be the case that these extensions would lead to the accurate application of the local expert procedure to larger communities. For example, an exhaustive and mutually-exclusive neighborhood-by-neighborhood application could lead to accurate estimates for a large city.

PROPOSED YUCCA MOUNTAIN REPOSITORY SITE AND SURROUNDING AREA



The proposed Yucca Mountain repository site is located in Nye County, Nevada, about 100 miles northwest of Las Vegas. Yucca Mountain is located on lands under separate control of three Federal agencies: Bureau of Land Management, U.S. Air Force (Nellis Air Force Base), and DOE (Nevada Test Site).

We conclude by noting that low costs and a nonexistent response burden are inherent features of the LEP. Coupled with its ability to produce accurate estimates, as well as measures of estimation uncertainty, these features make the LEP particularly attractive.

[1]L. Roe, J. Carlson and D. Swanson, "A Variation of the Housing Unit Method for Estimating the Population of Small, Rural Areas: A Case Study of the Local Expert Procedure," *Survey Methodology*, Vol. 18, June, 1992, pp. 155-163. This document contains the original case study of the LEP conducted in the three Arizona communities discussed in this article and serves as an original reference document for the article.
[2]W. Freudenberg, "Social Impact Assessment," *Rural Society in the U. S.: Issues for the 1980s*, edited by D. Dillman and D. Hobbs. Boulder, Colorado: Westview Press, 1982, pp. 296-303; R. Brown, H. Geersten and R. Krannich, "Community Satisfaction and Social Integration in a Boomtown: A Longitudinal Analysis," *Rural Sociology*, Vol. 54, No. 4, 1989, pp. 568-586; Kenneth Munsell, "Towns Cope with Hazardous Wastes," *Small Town*, Vol. 18, No. 5, March-April, 1988, p. 30 (book review of Raymond D. Scanlon, editor, *Hazardous Materials, Hazardous Waste: Local Management Options*, Washington, D.C.: International City Management Association, 1987).
[3]J. Martin and W. Serow, "Virginia's State-Local Cooperation Program: Conflict and Cooperation in Producing Population Estimates," *State Government*, Vol. 52, No. 4, 1979; D. Swanson, B. Baker and J. Van Patten, "Conceptual and Practical Features of the Housing Unit Method," paper presented at the Annual Meeting of the Population Association of America, 1983; N. Rives and W. Serow, *Introduction to Applied Demography Data Sources and Estimation Techniques*, Beverly Hills, California: Sage Publishers, 1984; Roe, Carlson and Swanson, 1992.
[4]United States Department of Energy, *Yucca Mountain Site Characterization Project Socioeconomic Plan: Revision O*, Washington, D.C.: United States Department of Energy, 1991.
[5]J. Carlson, D. Swanson and C. Williams, "The Development of Small Area Socioeconomic Data to be Utilized for Impact Analysis: Rural, Southern Nevada," *High-Level Radioactive Waste Management Proceedings of the 1990 International Conference*, LaGrange Park, Illinois: American Nuclear Society, 1990, pp. 985-990.
[6]United States Department of Energy, *Summary of Socioeconomic Data Analysis Conducted in Support of the Radiological Monitoring Program During the Calendar Year 1990*, Washington, D.C.: U. S. Department of Energy, 1991.
[7]United States Bureau of the Census, *1990 Census of Population and Housing, General Population Characteristics: Nevada*, Washington, D.C.: U. S. Bureau of the Census, 1991.
[8]United States Department of Energy, *Summary of Socioeconomic Data Analysis Conducted in Support of the Radiological Monitoring Program During Calendar Year 1992*, Washington, D.C.: United States Department of Energy, 1993.
[9]United States Department of Energy, 1993.
[10]United States Bureau of the Census, 1991.
[11]J. Sniezek, "Groups Under Uncertainty: An Examination of Confidence in Group Decision Making," *Organizational Behavior and Human Decision Processes*, Vol. 52, No. 1, 1992, pp. 124-155.
[12]Brown, Geersten and Krannich, 1989.
[13]R. Dawes, D. Faust and P. Meehl, "Clinical Versus Actuarial Judgment," *Science*, Vol. 234, March, 1989, pp. 1688-1674.
[14]N. Weinstein, "Unrealistic Optimism About Future Life Events," *Journal of Personality and Social Psychology*, Vol. 39, No. 4, 1980, pp. 806-820.
[15]T. Buckley and J. Sniezek, "Passion, Preference and Predictability in Judgment Forecasting," *Psychological Reports*, Vol. 70, 1992, p. 1022.
[16]Sniezek, 1992; Dawes, Faust and Meehl, 1989.