NRRI-84-14

A METHODOLOGY FOR TELEPHONE USAGE STUDIES RELATING USAGE TO DEMOGRAPHIC OR OTHER VARIABLES

Clark A. Mount-Campbell
Associate Professor of Industrial and Systems Engineering

John B. Neuhardt
Professor of Industrial and Systems Engineering and Statistics

Baumin Lee
Graduate Research Associate

THE NATIONAL REGULATORY RESEARCH INSTITUTE
2130 Neil Avenue
Columbus, Ohio 43210

March 1985

This report was prepared by The National Regulatory Research Institute (NRRI) with funding provided by participating member commissions of the National Association of Regulatory Utility Commissioners (NARUC). The views and opinions of the authors do not necessarily state or reflect the views, opinions, or policies of the NRRI, the NARUC, or NARUC member commissions.

EXECUTIVE SUMMARY

This study is a methodology for planning telephone usage studies that will relate measures of usage to demographic characteristics of the subscriber population. The methods studied are also applicable to other factors that may affect telephone usage, as long as a readily observable variable, like the location of households, is correlated with the factors. The purpose of a study of this type is to identify ways that such studies can be done scientifically at a reasonable cost. By keeping the cost reasonable, an incentive for doing usage studies is provided. It is our hope that, as a result of this study of methods, telephone usage studies will increase in number so that the regulatory community will have the data needed to understand more broadly the telephone industry's role in society and the regulatory commissions' participation in effectively overseeing that role.

The first chapter of this report is an introduction to the problem and it documents the need for usage studies. It contains a discussion of the cost of doing usage studies and putting this report to use. In chapter 2 is provided a general discussion of sample design and analysis considerations for the technically minded reader. Also for the technical reader is appendix B that presents a simulation model of household telephone usage. The simulation model was developed to produce artificial usage data that possess the same statistical qualities that have been attributed to telephone usage in the literature. Chapter 3 contains a step-by-step working guide to the design of telephone usage studies including an analysis of certain critical design decisions. This analysis uses the simulation program in appendix B. A pilot test of the working guide was performed for West Virginia and is discussed in chapter 4. The West Virginia pilot also produced a model data request found in appendix D. Appendix A is a summary of seven of the most prominent telephone usage studies found in the published literature, while appendix C is a list of all the known state agencies whose function is to disseminate U.S. Census Bureau information.

With the emerging changes in the telecommunications industry comes the need for telephone usage studies that can shed light on issues like local measured service, extended area service, and cost-based rates for both local and long-distance service. The determination of the distributional effects of changes in policy, such as the new access charge policy, is dependent on knowing the extent to which different levels of usage occur among different demographic groups. In addition to policy and pricing issues, better long term strategic planning of capacity and financial requirements can occur from a clearer understanding of telephone usage.

On the other hand, the cost of telephone usage studies can be quite high. The costs of the data collection phase of usage studies can be divided into three components: (1) administration of the study, (2) measurement of the population characteristics, and (3) measurement of usage. Items (2) and (3) can account for the expenditure of substantial resources by either the telephone company or the commission if it is directly involved in the study.

Costs associated with the measurement of usage would most likely be borne by the telephone company and consist of two types of costs. First, there is an office set-up cost of providing central offices with the facilities to measure usage. This involves the use of portable measuring equipment, engineering and equipment changes for remote measurement, new equipment for on-site measurement, or a generic computer program to do measurement. These are expensive alternatives that can cost more than $10,000 per office, but in many parts of the country today this cost can be totally avoided by concentrating the sample of households in central offices already having measurement capability. While this approach can introduce some bias into the results, the widening use of electronic office equipment together with optional local measured service will reduce the amount of the bias over time. Second, for each household in a sample there is a household set-up cost roughly equal to the cost of processing a service order. In West Virginia, this cost was quoted as thirty-five dollars per service order. However, this figure is an average cost that may be greatly reduced if the company is given enough lead time to process the study-related service orders during slack periods.

If office set-up costs can be avoided, the costs associated with the measurement of the demographic characteristics of sampled households are perhaps the largest single remaining expense of the telephone usage study. Indirect assessment of population characteristics offers an opportunity for great cost savings and is based on using the existing summary statistics available from the U.S. Bureau of the Census. Since the use of census data can have a deleterious effect on the quality of results from a usage study, part of the work in this methodology was intended to examine the magnitude of that effect in the pilot state.

A first issue with respect to designing a study that will relate telephone usage to demographic characteristics of households is the question of whether to plan a single-period study or a multiperiod study. The problem of designing a good sampling plan is greatly complicated by the autocorrelation that would be present in the time series of data collected in each household of a multiperiod study. The approach taken in this study was to concentrate on the design of a good single-period study with the idea that a time series of several periods could be collected from the sample of households intended for the single-period study. It is thought that this approach would result in a good multi-period study design that would be somewhat "over-designed" in that the sample size requirements would be smaller in the multiperiod study than in the single-period study.

A second issue with respect to designing a usage study is the question of what design criterion can be usefully applied to the problem. A number of such criteria appear in the literature; only two have gained the greatest prominence. They are D-optimality and A-optimality, where the D and A refer to specific mathematical characteristics of the matrix that specifies the independent variable values for the households in the study. These mathematical characteristics are indicators of how small the confidence intervals will be for the parameters of a model that relates demographic characteristics to telephone usage.

In chapter 2 of this study the theory for a sample design method is presented that consists of selecting geographic areas where samples of households would be concentrated in order to achieve a D-optimal and/or an A-optimal design. These geographic sampling areas (GSA) are defined as the smallest geographic areas for which summary federal census statistics are available and can therefore be used as proxies for the actual statistics of sampled households within them. General procedures are stated that will lead to answers to the following critical design questions:

1. How many GSAs should be sampled?

2. Which GSAs should be sampled?

3. What fraction of the total number of households in the study should be sampled in each GSA?

4. What should be the total sample size?

To help relieve the computational burden associated with D-optimal and A-optimal design theory a simplified approach that transforms the demographic variables into two- and three-level factors is suggested. The approach has the goal of achieving an orthogonal fractional factorial design in which the total sample is split equally among the GSA in the design. Such designs are known to be D-optimal, but they may not exist when one is restricted to the GSAs available in a given study area. In this case, it is suggested that a good practical approach is to find an infeasible, optimal but nearly feasible design and then to consider designs derived from the infeasible one by either deleting the infeasible GSAs or replacing the infeasible GSAs with the most demographically similar GSAs available.

A general framework for planning and conducting a usage study is given in a flowchart format with each step detailed and illustrated in the test. To summarize, we list here the general steps in the planning process:

1. Define study objectives--be specific about what one hopes to learn from the study. Presumably the objective could be stated in terms of estimating the relationship between telephone usage and other observable variables like demographic characteristics

2. Identify candidate dependent and independent variables

3. Determine the method of measuring dependent variables (usage), and independent variables

4. Determine whether to use an existing sample such as one the telephone company has used to estimate the subscriber line usage factor needed in the separations process

5. Collect and analyze population summary data on the independent variables

6. Specify the GSAs where samples of households will be located

7. Evaluate the proposed design and determine the sample size

8. Issue a data request to obtain usage data for the sample of households

9. Analyze the results and estimate the benefit that might be made with improved independent variable data

10. Determine whether or not a survey should be made of the sample households

11. If a survey is to be conducted design a subset of the sampled households to be surveyed

12. Survey the subset of the sampled households and update the data base of the independent variables and return to step (9)

Two other steps pertaining to using an existing sample of households are given in chapter 3, but are not listed here. The two most critical issues in the process listed above are the decision points in steps (4) and (10) where it is decided, respectively, whether to use the telephone company's pre-existing sample of households or to design a new sample, and whether to survey the sample of households to obtain household demographic data rather than use U.S. Census Bureau summary data. These issues are critical because they have the greatest potential effect on both the quality and cost of the study. To examine these issues in the West Virginia case, a computer simulation study was made of four cases derived from the combinations of possible answers to the two issues. These four cases were as follows:

Case 1: Use a new optimal sample but do not survey

Case 2: Use a new optimal sample and do a survey

Case 3: Use a pre-existing sample and do a survey

Case 4: Use a pre-existing sample but do not survey

The best situation from the standpoint of the quality of data is case 2, but the other three cases can produce results of quality equal to case 2 if the sample sizes are sufficiently increased. For example, if case 2 uses a total sample of 100 households, the other cases could produce equally good results if the sample size is increased to 190 for case 3, 425 for case 1, and 4,730 for case 4. The direct household set-up costs for these four cases were roughly estimated and could be as low as $1,275 for case 1, $5,300 for case 2, $10,450 for case 3 and $23,650 for case 4. If the full cost of processing a service order is used to estimate the set-up cost of cases 1 and 2, then their costs would be $16,150 and $8,800 respectively. Since the pre-existing sample in West Virginia consisted of about 1,400 households it is evident that a poor study would result from using that sample if a survey is not also conducted.

The chapter 4 applies the first 8 steps listed above to design a sampling plan that would be implemented in the Charleston area and would sample a total of 250 households. Four good plans were found using both the D- and A-optimal criteria. Two of the plans called for sampling in 4 GSAs, one uses 5 GSAs, and one uses 6 GSAs. These designs are expected to provide data to estimate a model that would relate usage to income, age of head of household, and the presence of children. Minutes of use are recorded for local, intrastate, and interstate calls. In addition to the designed demographic/usage study, a request for telephone usage and household location data for the existing, state-wide sample of households' normally used, subscriber-line usage studies was formulated. Such data will be of some value in a demographic/usage study, but when combined with the designed study in Charleston will provide a way of confirming the simulation results achieved in this study. A final item of interest to the staff of the West Virginia commission was the telephone usage on business lines. Thus, the data request formulated in the pilot study also asks for business-line telephone usages for the three services listed above and will use number of lines to a particular facility as the independent variable. As of this writing the data request has yet to be issued to the participating telephone company in West Virginia.

The study concludes with an example of ancillary results that will come as a byproduct of a usage study like the one designed herein; and a discussion of such additional issues as a measured service assumption, multi-period usage data collection, and a technical problem associated with finite populations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

FOREWORD

This report is not a telephone usage study, but rather a report on devising a method for planning and implementing telephone usage studies that will relate measures of usage to demographic characteristics of the subscriber population. The purpose was to describe ways that such studies can be done scientifically at a reasonable cost. Sample design and analysis considerations and a simulation model are presented and discussed. A pilot text of this working guide was performed for the West Virginia Public Service Commission and a sample data request to accomplish all this was prepared.

We feel that in addition to the policy and pricing issues that currently confront regulation in telecommunications better long term strategic planning of capacity and financial requirements can also flow from a clearer understanding of actual telephone usage. This report is presented in that light.

Douglas N. Jones
Director
Columbus, Ohio
April 18, 1985

ACKNOWLEDGMENTS

The authors would like to thank E. Dandridge McDonald, former Chairman of the West Virginia Public Service Commission, for involving his staff in the pilot study of telephone usage methods in West Virginia. We are also indebted to Joel Shifman and Danny Walker, both of the West Virginia Public Service Commission staff, for their interest and support. Both Joel and Danny took many hours out of their heavy schedules to make our several visits to Charleston productive. Their knowledge of West Virginia and the problems of regulating telecommunication services there was invaluable in conducting the pilot study.

The authors are also indebted to the C & P Telephone Company of West Virginia for its cooperation and ability to respond to our many informal requests for information about its operations. One of our major contacts with the company was Susan Lawson who was especially helpful in arranging our visits and answering our questions. Her thorough, professional, and efficient manner helped us get answers to some of our technical questions in a fraction of the time that might otherwise have been required.

Finally, a special appreciation is owed to Katherine Shiflet of the West Virginia Governor's Office of Economic and Community Development who provided us with a complete set of West Virginia census maps and books of census statistics together with a short course on the use of these materials. Her cheerful and knowledgeable approach to our problems was of great value in conducting the pilot study.

CHAPTER 1

INTRODUCTION

This is a study of the means by which one can efficiently and effectively study patterns of telephone usage.[1] The main goal is to develop a method that can lead to a reduced cost of studying the effects of household characteristics on telephone usage. In this way it is thought that a needed incentive will be provided to do more usage studies. Examining the published literature on telephone usage and demand, one finds that there is a great need for the type of information a usage study can yield, but that a major limiting factor in these studies is their cost. This cost factor apparently greatly reduces sample sizes--and hence the accuracy of the results. Just how many potential studies have been avoided because of their high costs is unknown.

In any event, it is clear that well-designed, statistically rigorous studies of telephone usage are needed in the regulatory community as commissions face decisions that will shift costs from heavy toll users to subscribers of local service, and from one type of local user to another. Such studies are scarce. It is our hope that The National Regulatory Research Institute (NRRI), as the research arm of the National Association of Regulatory Utility Commissions (NARUC), can serve as a catalyst to stimulate the establishment in many states of ongoing studies of telephone usage patterns by class of service and

---

[1]Telephone usage refers to a measurement of the amount of time a line is in use as distinguished from the surrogate measure of usage--"peg count," which is the number of calls made on a line. The usual unit of usage is CCS (hundred call seconds). Both usage and peg count could be interesting as dependent variables, and we use them interchangably throughout this report.

characteristics of customers. This study is a first step toward the achievement of that long term goal. We believe that finding ways of conducting usage studies without too large a commitment of either commission or company resources is the best way to encourage states to participate in such studies. To this end chapter 2 includes a general discussion of sample design and analysis considerations for the technically minded reader. Also for the technical reader is appendix B, which presents a simulation model of household telephone usage. The simulation model was developed to produce artifical usage data that possess the statistical qualities that have been attributed to telephone usage in the literature. Such a simulation model is a general purpose tool that allows one to "try out" a sampling plan and evaluate its performance before it is actually implemented. Chapter 3 contains a step-by-step working guide to the design of telephone usage studies, including an analysis of certain critical design decisions using the simulation program in appendix B. A pilot test of the working guide in West Virginia is given in chapter 4. The West Virginia pilot also produced a model data request found in appendix D. The remainder of chapter 1 documents the continuing need for studies of this type and discusses some general cost concepts.

The Need for Usage Studies

In commenting on studies of MTS markets conducted separately by Auray and Doherty, Chessler said "The fundamental difficulty everyone faces is lack of data."[2] At a 1979 telecommunications industry workshop

---

[2]Robert R. Auray, "Customer Response to Changes in Interstate MTS Rates," Assessing New Pricing Concepts in Public Utilities (East Lansing: Michigan State University, 1978), pp. 47-81; Noel A. Doherty, "Econometric Estimation of Local Telephone Price Elasticities," Assessing New Pricing Concepts in Public Utilities (East Lansing: Michigan State University, 1978), pp. 101-129; David Chessler, "Comments," Assessing New Pricing Concepts in Public Utilities (East Lansing: Michigan State University, 1978), p. 135.

a member of the Wisconsin Public Service Commission, spoke of the regulator's perspective on local measured service (LMS) and said: "... I believe that, given the appropriate data about cost and elasticity of usage, difficult questions of equity and economic efficiency can be satisfactorily resolved."[3] In a joint industry paper on local measured service given at the same conference, Beauvois et al. suggested that "usage studies should be made in exchanges where implementation of LMS is planned."[4] In perhaps the most extensive study to date of demographic effects on telephone usage, Brandon states: "... one may be interested in comparing the calling behavior of two divisions of a single demographic characteristic.... This kind of comparison is generally what is desired when one investigates the impact of pricing changes among demographic groups."[5] Brandon et al. also indicate that

> A project that could make the above kind of analysis more robust would be the pooling of the Chicago data with similar data from other cities. Hopefully, the rate structure in some of the other cities would be flat rate (i.e. no charge for local calls); and the cities sizes, density, weather, etc., would vary. It might be possible to estimate from such pooled data how the number of local calls is affected by demographics, by price over a wider range of variation than is observed in Chicago and its suburbs, and by other variables such as the characteristics of the cities themselves.[6]

---

[3]Edward M. Parson, Jr., "Local Measured Service: A Regulator's Perspective," Perspectives on Local Measured Service (Kansas City, Mo.: The Telecommunications Industry Workshop Organizing Committee, 1979), pp. 177-182.

[4]Edward C. Beauvais et al., "The Financial Effects of Local Measured Service," Perspectives on Local Measured Service (Kansas City, Mo.: The Telecommunications Industry Workshop Organizing Committee, 1979), p. 81-117.

[5]Belinda B. Brandon et al., The Effects of Demographics of Individual Households On Their Telephone Usage (Cambridge, Mass.: Ballinger Publishing Co., 1981), p. 14.

[6]Ibid., p. 366.

The authors go on to say that their study of usage in Chicago has partly answered the question of how demographic characteristics of households relate to telephone usage, but they conclude with the observation that "there is much room left for further research in this area."[7] In another study of the relationships between demand for local telephone calls and household characteristics, Infosino concludes that

> the evidence presented here suggests that household characteristics may be a fruitful area of research to obtain a more general understanding of the demand for telephone services. More data and more analysis are needed to study alternative model specifications and to precisely determine the relationships between household characteristics, the demand for telephone services, and the prices of telephone services.[8]

Infosino went on to point out the need for a pooled analysis of data from many different areas and a need to investigate the correlations among household variables (multicollinearity) as well as the interactive effects of household variables on calling patterns.

Usage Study Costs

Several of the authors cited above, as well as others, have mentioned cost as an important factor governing the design of usage studies. In general, when usage studies are designed to obtain data that would allow analysts to relate telephone usage to characteristics of the population, the cost of the data collection phase of such studies can be separated into component parts corresponding to (1) administration of the study, (2) measurement of the population characteristics, and (3) measurement of usage.

---

[7]Ibid.

[8]W. J. Infosino, "Relationships Between Demand for Local Telephone Calls and Household Characteristics," The Bell System Technical Journal (July-August 1980): 951.

The administration costs are shared by the telephone company and the state regulatory commission if the commission seeks to exercise some measure of control over the study design. The credibility of a study is enhanced in appearance, and presumably in fact, by having a study designed by objective researchers who have no vested interest in the outcome of the study, in pricing, or in other regulatory decisions that may be influenced by the study. This suggests that significant commission staff involvement in the study design is desirable. Thus, rather than a commission issuing a data request asking the company to "do a usage study," the data request should completely specify the sampling plan. Only then is it truly a data request and not a "study" request. However, such a data request should specify a plan that is feasible at a reasonable cost if the public is to be well served by it. For this reason meetings with company officials, preliminary data requests, and the like are necessary and create administrative costs for both the company and the commission. Developing the data request containing the final sampling plan is, however, the major part of administrative costs.

The measurement of population characteristics can be done directly by several means, with either the company or the commission bearing the cost. These methods include questionnaire inserts in monthly bills, separately mailed questionnaires, telephone interviews, and personal interviews. All of these are expensive approaches with typically less than 50 percent response rates, and if the survey is to be used to aid sample design in order to help control multicollinearity, then the number of households or businesses surveyed would probably have to be six or seven times the number in the final usage sample.

Indirect assessment of population characteristics offers great cost savings and would be based on the already existing summary statistics available from the U.S. Bureau of Census. Census data has the advantage of containing information on a large enough population that a study design can be made that at least partially controls the household characteristics in a sample by concentrating the samples in certain census block groups. The disadvantages of census data are that they can be obtained only at a summary level and they soon become out of date.[9]

The cost of these methods of measuring population characteristics is functionally related to the number of households being measured. The cost of indirect methods is expected to be one or two orders of magnitude less than any of the direct measurement methods.

A general but simple model for the third category of costs--measurement of usage--is given for flat rate customers by the sum of terms that compute set-up costs incurred to give wire centers the ability to measure flat rate customers; terms that compute the set-up costs incurred to establish measurement on each flat rate line; and terms that compute monitoring costs.

The set-up costs associated with wire centers depend upon the particular configuration of the equipment at the center where some flat rate customers are to be included in the study. If the switching equipment in a wire center is electromechanical, then special measurement equipment must be acquired and installed at a cost that can be as much as $30,000 per office. An alternative in these cases would involve remote measurement, but again, engineering changes and additional traffic-sensitive equipment would be needed. This approach is used infrequently and can have a cost approximating with the $30,000 figure mentioned above. In wire centers with electronic switching equipment special customer class codes are needed to identify the measured flat rate sample customer separately from the ordinary measured rate customer. This requires (if it is not already present) a special generic program for the switch's central processing unit in order to assign, recognize, and utilize the special customer codes. It was indicated in the West Virginia pilot study that license and installation fees for the generic program would cost somewhere between $5,000 and $10,000 per office needing it.

---

[9]Census data are periodically updated in some areas.

Uses of the Report

In the previous section, costs of making usage measurement possible in some of the older offices were cited as posing a major restriction in study design. However, the newest technology in central office equipment is replacing older equipment at a rate that appears to have been quickened by the recent divestiture of the AT&T operating companies and the threat of competition for local service. Thus, this major restriction is being relaxed. There are also trends toward measured pricing for local services and cost-based pricing for all services. As a result, more and more state commissions will want a better understanding of the basic characteristics of telephone usage in order to gauge the implications of their policies. We hope that this report can be a further catalyst by being useful in any one of several ways. First, for commissions with sufficient staff to design and oversee a usage study, this report provides a step-by-step guide, a model data request, and some analyses that will help in evaluating the tradeoff between cost and quality. Second, some commissions may ask the telephone company to design and conduct a usage study, in which case this report could be suggested to the company as a guide and it could be used by the commission staff to evaluate the company's study plan. While this report does not contain a model study request such as might be issued in such a case, the first three steps given in chapter 3 should be carried out by the commission in consultation with the company so that a specific study request can be framed as to what should be learned by a usage study. Third, a commission may wish to contract for the study design work. In this case, this study should be helpful in preparing the request for proposals and in evaluating those proposals submitted by competing contractors.

CHAPTER 2

ANALYSIS MODELS AND OPTIMAL SAMPLING PLANS

This chapter is concerned with methods of planning a cross-sectional experiment that consists of sampling, for one period, the telephone usage of geographic areas with known demographic characteristics. The sampling areas are described, as are the analysis models needed to prescribe the planning process. A planning process is then proposed that reasonably adapts to the restriction that one cannot "manufacture" a geographic area to suit one's needs, but must take whatever exists. The sample design method becomes one of selecting certain regions that collectively have "good" characteristics. After an example, a case in which usage measures are binary-valued is discussed.

Definition and Discussion of Demographic Sampling Areas

A geographic sampling area (GSA), is defined as a geographic area in which household samples are located and for which summary federal census statistics may be used as a proxy for the actual statistics of the sampled households. By "a proxy for the actual statistics" we mean that if a household is part of a sample of households, then the average (or median) demographic characteristics of the GSA in which that household is located will be recorded in the data set in place of the actual statistics of the household. Further, this means that all sampled households that reside in the same GSA will appear in the data as having identical demographics even though they actually differ one from another. A question that arises in defining GSAs is: how large should they be? If one considers only statistical properties and no practical or cost aspects of a sampling plan, then the best possible GSA size would be a single household. If this was the case, then GSA statistics would not be a proxy for household statistics--instead they would be exactly the same as household statistics. However, individual household statistics cannot be obtained from census statistics. Thus if the census data are to be used, then a GSA larger than one household must be used. On the other hand, very large GSAs, such as entire cities, would be expected to be less homogeneous than smaller GSAs like neighborhoods. A second disadvantage of large GSAs is that the larger a group of households the more the group average approaches the population average. This means that if GSAs are cities, it may be difficult to find cities that differ substantially in their demographic averages. For example, if income is a study variable, a good sample design would call for some households that are wealthy and some that are poor. In fact, it is important to achieve the greatest possible separation between the wealth of the two groups. However, if GSAs are as large as cities of, say, sixty-thousand people, it will be difficult (or most likely impossible) to find two cities with widely divergent median incomes. On the other hand each of the cities could very well contain wealthier neighborhoods of a couple of hundred households and a poverty stricken neighborhood of a couple of hundred households. Sampling from these neighborhoods and using neighborhood median income as proxies for actual household income would be better than doing the same with much larger areas than the neighborhoods.

Thus, for these reasons, and the reason that federal census data will be used to design the sample even if not finally used as a proxy to analyze the data, we would want to use the smallest possible GSA consistent with the availability of federal census statistics.

It is generally possible to obtain summary statistics for areas defined by political boundaries.[1] Examples are states, counties,

---

[1]For a complete discussion of federal census data see: U.S., Bureau of the Census, Users Guide, PHC80-RI-A (1982).

cities, boroughs, magisterial districts, and townships. In more rural areas there are places, rural places, or farms. In addition to these there are smaller statistical units, such as tracts, which are subdivided into block groups that in turn consist of blocks. Only major population areas are tracted. The statistical units in untracted areas are called enumeration districts, which are not more finely divided, except if there is a political boundary cutting through them. These statistical units are hierarchical and do not overlap each other. As a general rule the lowest level population aggregate for which federal census summary statistics can be obtained is the block group. There are many exceptions to this general rule that result when political boundaries cut through a block group. In these cases, summary statistics can be obtained for households that are located on each side of the political boundary and that are within the block group.

For purposes of this chapter, a GSA will be a block group or, whenever available, a part of a block group as separated by a political boundary. Such areas in the West Virginia pilot study may contain from 10 to 1,000 households, with the most typical size being about 250 households.

A Usage and Analysis Model

In this section we propose two analysis models that closely resemble telephone usage models found in the literature. One model assumes perfect information about sampled households while the other assumes only summary information about neighborhoods of households (i.e., GSAs). These models are used as tools for designing a sampling plan.

Assumptions are needed in order to analyze the data collected from any sampling plan. Additional assumptions make it possible to prospectively quantify the amount of "information" that will be gained from the data collected according to a plan. One can then choose the plan that maximizes the information.

In this study a basic assumption, which is evident in the choices below of the households usage models used to design a sample, is that the telephone usage in a household is related to the demographic characteristics of that household. Although several different usage measurements are to be taken for each household (e.g., local calls, toll calls) we will treat each usage measurement in separate models rather than assume a far more complicated multivariate structure. The implicit assumption in this approach is that in any given household the usages of different services are statistically independent. In the event this assumption is not true, the estimates of the parameters of models described below would be unaffected, but the statistical power of any hypothesis tests about the parameters would be affected. The effect on the power of such tests would depend upon the size and sign of the correlations among the various services. In addition, a linear model relating usage or a transformation of usage to demographic characteristics is assumed for the purpose of studying the quality of different sampling plans.[2] Finally, a one-period or cross-sectional study is assumed. Implications for design and analysis of the "n" period, or longitudinal study, are discussed later in this report.

Given these assumptions, consider the following household usage model (HUM):

$$U_{i(j)} = M + X_{i(j)}\beta + e_{i(j)}, \qquad (2.1)$$

where

$U_{i(j)}$ = usage (or transformed usage) of household i in GSA j, during one period,

M = intercept term,

[2]A transformation of the usage measurement may be needed to aid in the analysis of the data. For example, Park et al. (see appendix A) use local calls raised to the .27 power, rather than local calls, in their model. Fortunately transformations need not be specified in the planning effort as long as a linear model is postulated.

$X_{i(j)}$ = (1 x s) row vector of demographic characteristics of household i in GSA j where s is the number of different characteristics.

β = (s x 1) column vector of rates of change in usage with respect to demographic characteristics,

$e_{i(j)}$ = a residual term accounting unexplained sources of variation in usage.

Since demographic data may not be available at the household level but only at the GSA level, a geographic analysis model (GAM) is derived from HUM:

Let

$R_{\cdot j}$ = (1 x s) row vector of census records of s demographic characteristics of GSA j, (2.2)

$D_{i(j)}$ = (1 x s) row vector of s demographic characteristic differences between household i in (GSA j) and the corresponding s census records of the demographic characteristics of GSA j, (2.3)

so that

$$X_{i(j)} = R_{\cdot j} + D_{i(j)}, \qquad (2.4)$$

and the GAM model becomes

$$U_{i(j)} = M + R_{\cdot j}\beta + f_{i(j)}, \text{ where} \qquad (2.5)$$

$$f_{i(j)} = D_{i(j)}\beta + e_{i(j)}. \qquad (2.6)$$

Notice that the "error" term, $f_{i(j)}$, in the GAM model includes the quantity $D_{i(j)}$. This term represents the demographic error that occurs in a model where the summary census information is substituted in place of individual household characteristics. Thus, it may include two sources of error. First, an error can occur when the household selected as a sample point does not have exactly the characteristics recorded in the census records for the GSA and second, there can be census errors. The latter could be due to a one-sixth sampling rate used for some federal census statistics or it could be due to clerical or other errors; in any case it is assumed not to be significant when compared to the first source of error.

Armed with these two models of usage we proceed to explore the problem of designing the best stratified sampling plan where each stratum is composed of one or more GSAs. For purpose of the discussions that follow, we assume that for reasons of achieving the greatest homogeneity of population of each stratum that the GSAs are the same as strata; hence there would be only one GSA in a stratum.

Design Questions and Principles of Selecting GSAs

Below we formally define the problems of optimizing a sampling plan. We do this by focusing on three preliminary questions and known principles for answering the questions in very simple univariate situations. These principles are then expanded to cover the typically more complex multivariate situations.

Before a stratified sampling plan can be fully developed certain design questions must be answered. These questions are: (1) How many GSAs (strata) should be sampled, (2) Which GSAs (strata) should be sampled, (3) What fraction of the total number of households in the study should be sampled in each GSA (stratum).

In question 3 we assume that the total number of households, N, is given. The problem of determining a value for N can be addressed independently of the above three questions and is answered in a later section.

Given GAM, and a goal of estimating the effects on usage of demographic variables, two basic principles having to do with design questions 1 and 2 may be stated:

Principle One: Suppose we have "s" demographic characteristics for which you wish to estimate the effects on usage. We should select at least s + 1 GSAs, no two of which should have identical demographic characteristics.

Principle Two: If s = 1, then we must select the GSAs as "far apart" as possible.[3]

Principle Two is easily implemented because it applies only when one demographic variable is being considered in the study. For example, consider six hypothetical GSAs with demographic data as shown in table 2-1.

TABLE 2-1

DEMOGRAPHIC CHARACTERISTICS OF SIX GSAS

GSA	Percentage of Households with Head Greater Than 64 Years (Factor A)	Median Income (Hundreds) (Factor B)
1	.1	32
2	.2	29
3	.6	17
4	.4	18
5	.8	9
6	.8	9

Source: Authors' example

We label the "age" variable as factor "A," and suppose that this is the only variable of interest. Then Principle Two suggests that GSA 1 and 5 or 1 and 6 should be selected. The same selection is suggested if factor B (income) is the only variable of interest in the study.

If in fact both factors A and B affect telephone usage, then this example also illustrates the dangers inherent in not including all relevant (and significant) variables when there is correlation among the variables. In this example, if factor A (age) was the only

---

[3]If the goal of the study is to estimate the usage of an entire population, then the systematic selection of GSAs suggested in the second principle would not be appropriate. The assumption here is that the study goal is to estimate the relationship between the demographic variable and usage.

variable considered so that a plan using, say, GSAs 1 and 5 was implemented, then the resulting usage data would systematically contain both the effects of age and the effects of income, but one could not separate those two effects in analysis.

Suppose, however, that we wish to include both factors A and B. Then Principle One suggests that at least three GSAs would be needed. We can choose GSA 1 and 5, or 1 and 6 again, but which should be the third GSA? Unfortunately, no simple rule can be stated for two or more demographic variables; however there are techniques that lead to selections of GSAs in such instances. These are discussed in the next section.

Selecting GSAs with Two or More Demographic Variables

A simplifying assumption applied to the GAM model is made to accommodate the case of two or more demographic variables. This assumption holds that the variance of $f_{i(j)}$ is constant and equal to

$$\sigma_e^2 + \Sigma\beta_\kappa^2 \sigma_\kappa^2 , \qquad (2.7)$$

where σ_e^2 is the constant error variance in the HUM model and σ_κ^2 is the variance of the κ^{th} component of $D_{i(j)}$.[4]

Implicit in this assumption is that σ_κ^2 is also constant and independent of i and j.[5] A simple representation of variances of the estimates of the elements of β in terms of the demographic characteristics of any proposed GSA sample can now be made. Suppose "n" GSAs are to be selected, and "s" demographic effects are to be estimated. Let $R_{\cdot 1},\ldots,R_{\cdot n}$ be the vector values of the demographic characteristics of

[4]It is assumed that the covariance of these errors is negligible.

[5]There are some cases where this assumption is not true. For example, when age is a study variable expressed as a fraction above some particular age in a given GSA, then a sample of households would result in a binomial distribution the variance of which depends upon the value of the fraction above the age in the area. This problem will not be considered in the design stage of the study.

the "n" GSAs, and let $Q_{\cdot i} = (1,R_{\cdot i})$ for $i = 1,..,n$. That is, the vectors "Q" are the demographic vectors "R" with an additional value of one in the leading elements. If p_i is defined as the fraction of the total samples, N, taken in the ith GSA, then a matrix, W, can be computed that is directly proportional to the variance-covariance matrix for the estimators of β. The constant of proportionality is $1/N^2$. The formula for computing W is as follows:

$$W = [\sum_{i=1}^{n} p_i \, Q_i \, Q_i']^{-1} \qquad (2.8)$$

Two common criteria for selecting the GSAs are Minimize det (W), or equivalently Maximize det (W^{-1}) and Minimize trace (W).[6] The expression "det" and "trace" denote the determinant and trace of the matrix W and are often referred to as the "D" and "A" criterion, respectively. Both criteria seek to impart desirable characteristics to the variance-covariance matrix of the estimates of effects. In particular, the D-criterion seeks to minimize the volume of a joint confidence region while the A-criterion is intended to minimize the sum of the variances of all the estimators. Unfortunately, the task of finding an optimum design is often computationally difficult, and the degree of collinearity in the independent variables of a selected design cannot be directly assessed using either criterion. These difficulties are alleviated to some degree if the quantitative variables are transformed to indicator variables as follows:

> For each recorded quantitative demographic variable in the GAM model, let $R_{\cdot j} = (r_{1j},..,r_{sj})$ and define the intervals for each variable $(l_i,u_i;\ i = 1,2,..,s)$ with the indicator variables $(t_{1j},....,t_{sj})$ as follows:

$$t_{ij} = \begin{cases} -1 & \text{if } r_{ij} \le l_i \\ 1 & \text{if } r_{ij} \ge u_i \\ 0 & \text{otherwise} \end{cases} \qquad i = 1,..,s;\ j = 1,2,..n. \qquad (2.9)$$

---

[6]See for example, V. V. Fedorov, Theory of Optimal Experiments, (New York: Academic Press, 1972), pp. 51-53.

For example, consider the GSA demographic data given in table 2-1 with $l_1 = .3$, $U_1 = .7$, $l_2 = 12$, and $u_2 = 25$. Then the original demographic values are transformed as shown in table 2-2.

TABLE 2-2

EXAMPLE OF INDICATOR VARIABLE CODING SCHEME[7]

GSA	Age (Factor A)	t_{1j}	Income (Factor B)	t_{2j}
1	.1	-1	32	1
2	.2	-1	29	1
3	.6	0	17	0
4	.4	0	18	0
5	.8	1	9	-1
6	.8	1	9	-1

Source: Authors' calculation from table 2-1 using the indicator variable transformation

Notice that the values t_{ij} translate the demographic values into -1, 0, or 1, according to whether the original values are less than the lower limit, l, between the lower limit and upper limit, u, or greater than the upper limit, respectively. However, there is an advantage if the three levels can be further restricted to two levels by eliminating those GSAs with one or more factors having value 0. The advantage is that there is an extensive set of easily applied methods available to derive some classes of optimal two-level sample designs.

We now define general steps for answering both the question of how many and which GSAs should be selected. A particularly helpful feature of the approach in these steps is that they assume p_i (the proportion of households sampled in GSA i) is equal to 1/n if i is one of the n GSAs selected, and 0 otherwise. This means that all three design questions are answered at once, and the calculation of W is simplified to

---

[7]Nonquantitative variables having two or three levels may be similarly coded with -1; +1; or -1, 0, +1 respectively (e.g. male; female; or urban, suburban, rural).

$$n\left[\sum_{i=1}^{n} Q_i Q_i'\right]^{-1} \tag{2.10}$$

where n is the number of GSAs in which samples are taken. Thus for each fixed n the problem is reduced to finding the set of n GSAs minimizing

$$\text{Det}\left\{\left[\sum_{i=1}^{n} Q_i Q_i'\right]^{-1}\right\} \tag{2.11}$$

or equivalent by maximizing

$$\text{Det}\left\{\left[\sum_{i=1}^{n} Q_i Q_i'\right]\right\} \tag{2.12}$$

1. Select lower and upper values (l_i, u_i) for each quantitative demographic variable.

2. Transform all quantitative variables to -1, 0, +1 and temporarily set aside GSAs with one or more 0s.

3. Select n (number of GSAs), such that n is equal to power of 2.

4. Search for an orthogonal fraction.

5. If one exists, go to step 7.

6. If none exists, either reduce the number of parameters in the model and go to step 4, or search for a nonorthogonal design and go to step 7. (The search for nonorthogonal designs involves reintroducing the GSAs with one or more 0s that were set aside in step 2. It also involves, in the absence of computer programs that find D-optimal solutions, a trial-and-error search procedure, but some useful and simple strategies for finding good nonorthogonal design exist and are discussed in more detail later.)

7. Compute the D- or A-criterion using the -1, 0, +1 codes assigned in step 2 or using the original raw values for the demographic data, and decide whether to stop or to continue, by either returning to step 3, step 4, or continuing the trial-and-error search for a nonorthogonal design.

These steps are now illustrated with an example.

Suppose we are interested in obtaining estimates of the effects on usage of the following GSA characteristics:

Factor A: percentage of households with head of household greater than 64 years of age

Factor B: median income of GSA

Factor C: percentage of households in GSA with at least one child less than 18 years of age

Factor D: urban/rural,

Factor E: area (two, north and south)

Since there are five factors, each at two levels, optimal designs exist for n = 8, 16, and 32 observations (among others, which are multiples of 4). Unfortunately, these known designs may be infeasible because they may require sampling in a GSA that does not exist. We shall consider the problem of finding a feasible design with n = 8. Standard fractional factorial theory would result in the possible plan given in table 2-3.[8]

TABLE 2-3

LEVELS OF EACH FACTOR FOR AN n = 8 DESIGN

	Factors				
GSA Observation	A	B	C	D	E
1	-1	-1	-1	1	1
2	-1	-1	1	-1	1
3	-1	1	-1	-1	-1
4	-1	1	1	1	-1
5	1	-1	-1	1	-1
6	1	-1	1	-1	-1
7	1	1	-1	-1	1
8	1	1	1	1	1

Source: Authors' example

[8]For examples of fractional factorial designs, see O. Kempthorne, The Design and Analysis of Experiments (New York: John Wiley & Sons, Inc., 1957).

This is a known optimal design for n = 8. Suppose in the particular telephone exchanges being studied that the eight GSAs with the above set of values of t_{ij} do not all exist. Then one must examine other known optimal designs. For this case, there are a total of sixty 2^{5-2} designs.[9] Efficient methods exist for determining whether any of these designs are feasible.[10] Suppose it is determined that no set of eight GSAs exists to allow an n = 8 orthogonal design to be specified. Then no design for n = 16 or n = 32 exists.[11] Suppose we are unwilling to assume that some of the effects are negligible (step 6). We must then consider nonorthogonal designs. This may be done in several ways. First, one can find an orthogonal design with the fewest number of infeasible points, and replace the infeasible points with nearby feasible ones including those that had been set aside in step 3 because they had some factor coded with a zero. Alternatively, the infeasible points may simply be deleted from the nearly orthogonal design under consideration, thereby reducing n. In this last approach, moving some of the remaining points to other nearby feasible points may improve the D-criterion. Again these nearby points may include those points that had been set aside in step 3. Second, the idea of "staying close" to an orthogonal design may be abandoned. In this case one would resort to other common approaches such as a "one-at-a-time design" in which observation 1 has all factors set to −1, and each subsequent observation has only one of the factors set to +1.

While there are programmable algorithms that seek to maximize D as suggested earlier, research continues into improving the algorithms, and developing others that use the A-criterion. Research also

---

[9] C. A. Mount-Campbell and J. B. Neuhardt, "On The Number of 2^{n-p} Fractional Factorials of Resolution III," Communications in Statistics --Theory and Methods, Vol. A10(20), pp. 2109.

[10] See, for example, Joe J. Pignatello, "Cost-Optimal Fractional Factorial Designs: A Methodology for Designing Experiments," (Unpublished Ph.D. dissertation, The Ohio State University, 1982).

[11] J. W. Knight and J. B. Neuhardt, "Computer-Aided Design of Fractional Factorial Experiments Given A List of Feasible Observations," IIE Transactions (June 1983): 142-149.

continues into the best ways to apply these algorithms.[12] Our experience is that on fairly small problems (three to five independent variables) a simple interactive computer program that alleviates the computational burden of forming the design matrix, X, given design points, and that computes the determinant and trace of X^tX is all that is needed. The analyst can then pose various designs in a trial-and-error fashion as the seven steps listed earlier are followed, locating very good designs quickly. As discussed earlier, two general strategies followed in posing the trial-and-error solution are (1) to choose GSAs that are as widely dispersed in demographic space as possible or (2) to start with infeasible orthogonal designs and either replace the infeasible points with the nearest available feasible points or simply delete them and reduce n. The second strategy seemed to work best in the West Virginia pilot study.

The Question of N, Total Observations

In the previous section, three design questions were answered under the assumption that the total sample size, N, was known. We now assume that a given set of GSAs to be sampled is known, together with the fraction of the total sample that is to be used in each GSA. This allows us to address a fourth design question: what is the total number of observations, N? To answer this question we formulate the following three steps for determining N so as to satisfy precision requirements of the estimates of the β's:

(1) Given a study objective of being able to estimate as precisely as required the parameters, β, of the GAM model, then the way to specify the requirement is to state the required average width of the 100 (1-α) percent confidence interval, C_i, on the elements of β.

---

[12]See, for example, D. M. Steinberg and W. G. Hunter, "Experimental Design: Review and Comment," Technometrics XXVI No. 2 (1984). For an optimal computer algorithm see T. J. Mitchell, "Computer Construction of 'D-Optimal' First-Order Designs," Technometrics XVI No. 2 (1974): 211-220.

(2) The minimum number of observations required to achieve the requisite width for each β_i confidence interval is denoted N_i and may be computed by the following equation:

$$N_i=[(2z_{\alpha/2})^2 W_{ii}\sigma^2]/C_i^2, \tag{2.13}$$

where W_{ii} is the i^{th} diagonal element of the matrix W defined in the previous section (assumes indexing starts at zero), σ^2 is the assumed error variance in GAM, and $z_{\alpha/2}$ is the standard normal value with $(1-\alpha/2)$ 100 percent of the standard normal curve lying below it.

(3) Select N such that $N \geq N_i$ for each i of interest. The calculation of N_i and its subsequent use in step 3 need not be made for every variable in the GAM model. It may instead be restricted to include only those n_i corresponding to variables whose parameters are of interest to the researchers.

An alternative and somewhat simplified approach to the one above may be applied if all independent variables have been coded -1, 0, or +1 as described in the previous equation. In this case one would compute the average of the squares of the C_i values specified in step one above. Denoting this quantity by C^2, N can be determined by satisfying the following inequality

$$N \geq [(2z_{\alpha/2})^2 tr(W)]\sigma^2/C^2 m \tag{2.14}$$

where the new term C^2 and m are respectively the sum of the squared confidence interval widths and the number of parameters in the model.

As an example of the simplified calculation, suppose we are planning to use the one-at-a-time design for n = 6 in the previous section and suppose the dependent variable is the fourth root of local calls per month (technically, "calls" are not usage, but a surrogate for usage). Further, suppose that we estimate σ^2 to be 200, and we want 16 to be the average of the squared widths of the 95 percent confidence interval, averaged over all effects. For the one-at-a-time design, $tr(W) = 36$ and $m = 6$, so that $N = 1{,}200$ when computed with equation (2.14). This implies that the plan should sample at least 200 households in each of the 6 GSAs.

Some Considerations in Binary Response Usage Data

Throughout this chapter, the linear model was assumed with a transformation of usage made, if necessary, to assume constant error variance. One situation will occur that is not commonly considered in such design problems, that of the usage data being binary-valued (0 or 1, say). This occurs when the usage is, for example, number of toll calls per month. Typically, a large percentage of households will make no calls in one month (or perhaps two or three months). One may study those households for which at least one call was made, but what about the simple measurement, percentage of households not making calls? This is an example of a binary response.

As discussed by Taylor in modeling access,[13] one convenient way to model P_j, the fraction not making a call in GSA "j," is to use a model consisting of transforming the fractions observed

$$\ln[P_j/(1-P_j)] = 0_j, \text{ called the "log-odds" ratio,}$$

and the logit model

$$0_j = M + X_j b + e_j.$$

The right hand side of the equation is the same as the GAM model, except the data are totaled at the GSA level rather than the household level. The reason that this is a special case is that the error term, e, is binomially distributed with a variance that is not constant over

---

[13] L. D. Taylor, Telecommunication Demand: A Survey and Critique (Cambridge, Mass.: Bellinger Publishers Co., 1980).

the js. Optimal designs have been discussed for the one demographic variable case.[14] The solution is to attempt to find two GSAs with the proportion of households making calls of approximately .2 and .8. Because of the intractability of the design problem for more than one demographic variable, and the general lack of research into this problem, we recommend the use of design approaches as discussed previously in this chapter for which the error variance is assumed constant.

Summary

This chapter has contained two telephone usage models, HUM and GAM, which, with some simplifying assumptions, are useful in leading to the application of D-optimal or A-optimal experimental design theory in order to determine

1. How many GSAs should be sampled
2. Which GSAs should be sampled
3. What proportion of the total sample should be placed in each selected GSA

To help relieve the computational burden associated with D-optimal and A-optimal design theory a simplified approach, which transforms the demographic variables into two- or three-level factors, is suggested. This approach has the goal of achieving an orthogonal fractional factorial design in which the total sample is split equally among the GSAs in the design. Such designs are known to be D-optimal, but they may not exist when one is restricted to the GSAs available. In this case it is suggested that a good practical approach is to find a nearly feasible design and then to examine designs made feasible either by deleting the infeasible GSAs from the design or by replacing the infeasible GSAs with the most similar feasible GSAs.

---

[14]For examples, see S. D. Silvey, Optimal Design (London: Chapmen and Hall, 1980), and L. V. White, "An Extension of the General Equivalence Theorem to Nonlinear Models," Biometrika LX (1973): 345-8.

A section is included that then computes the required total sample size, given a design. Finally, the case often arising in telephone usage studies where there is a significant binary characteristic to the data (i.e., people either make calls or they do not make calls) was considered. In this case it was concluded that because of a general lack of research into optimal design of such studies that the methods in the earlier part of the chapter would suffice, even though some assumptions leading to the application of D-optimal theory are violated by the binary character of the usage data.

CHAPTER 3

STEPS IN THE PLANNING PROCESS FOR A TELEPHONE USAGE STUDY

The purpose of this chapter is to provide an outline of the steps leading to a plan for studying telephone usage. The steps provided show alternative courses of action. A discussion of these alternatives will highlight the advantages and disadvantages of each. Further, the steps are stated in general terms but they are illustrated with specific examples. We have given the alternative steps and used general enough terms to incorporate sufficient flexibility to plan most studies under most conditions. In so doing, we have avoided taking a "cookbook" approach that could plan perhaps only one type of study. Many of the steps in the planning process can be accomplished without any specialized knowledge of experimental design and statistics although a completion of all steps in the process may require some reading in the experimental design literature.

For those commissions that cannot commit sufficient internal staff to the planning of a major usage study, this outline of planning steps can serve as a guide in developing a request for proposals (RFP) needed to seek a contract with outside sources. All or part of the planning steps can be contracted for, but significant staff involvement is recommended in order to assure the accomplishment of the commission's study goals.

The remainder of this chapter consists of a flow chart of the steps in the planning process followed by a detailed discussion of each of the steps. This discussion will fully describe the steps and summarize the implications where appropriate. Examples are included when necessary.

Overview of the Planning Process

Figure 3-1 is a flow chart of the process leading to a sampling plan to study telephone usage. In the figure each rectangle represents a step in the process. A diamond-shaped box in the figure represents a decision point. At these decision points one will branch to different parts of the flow chart depending on the outcome of the decision. Each box is marked with a code consisting of a capital letter followed by a decimal point and a number. This code is used for ease of reference in the section that follows this one. Dashed lines show areas where a return to previous steps in the process is most likely to occur because some prior step has had a result that proves infeasible or impractical after further analysis. Boxes marked with an asterisk indicate those steps requiring specific technical knowledge in the area of experimental design. While chapter 2 gives some guidance in the design area, those with the appropriate background who wish to acquire this specific technical knowledge are referred to a survey of experimental design that has been published recently. The survey article, "Experimental Design: Review and Comment" by David M. Steinberg and William G. Hunter contains an extensive bibliography.[1] In addition we recommend the following:

Federov, V. V. Theory of Optimal Experiments. Translated and edited by W. J. Studden and E. M. Klimko. New York: Academic Press, 1972.

Silvey, S. D. Optimal Design. New York: Chapman and Hall, 1980.

Mitchell, T. J. "An Algorithm for the Construction of D-Optimal Experimental Design." Technometrics XVI No. 2 (1974): 203-210.

Mendenhall, Wm. Introduction to Linear Models and the Design and Analysis of Experiments. Belmont, Ca: Duxbury Press, 1968.

---

[1]David M. Steinberg and William G. Hunter, "Experimental Design: Review and Comment," Technometrics XXVI No. 2 (May 1984): 71.

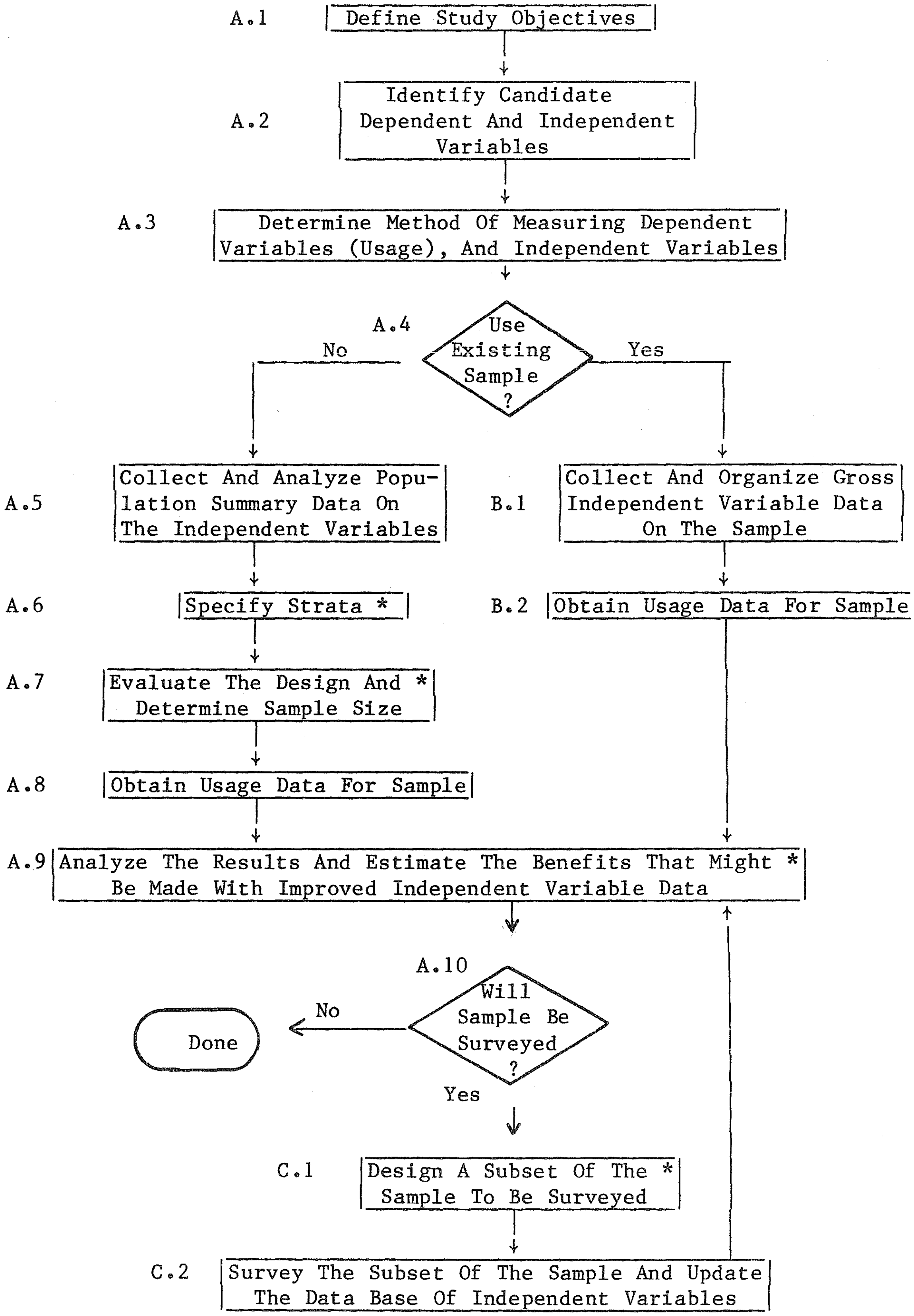

Fig. 3-1. Flow chart of steps in the planning process

Details of the Planning Steps

In this section each of the steps in the planning process is discussed in detail. A subsection is devoted to each step in the process and headed with the (letter.number) code marking the steps in figure 3-1. These are followed by a discussion of the process as a whole.

A.1 Define Study Objectives

The fundamental question here is, "what is it that we hope to learn by conducting the study?" The objective would then be to produce data that can yield what we hoped to learn. For example, suppose the question is, "what is the value of SLU (the Subscriber Line Usage Factor used in separations) for the state?" In this case an appropriate sample of households, and/or businesses would be measured for their interstate and intrastate usage so that numerical estimates of SLU can be computed from the data.

The appropriateness of the sample would be judged on the basis of how accurately SLU is known after the study. Thus, an objective for the study might be to estimate SLU with a 95 percent confidence interval no wider than plus or minus 1 percent of its value.

Once it is known what we want to learn and once we can be explicit about it, formulating the objective is relatively easy. For example, those developing cost studies using cost-causative principles may wish to know SLU (or some intrastate counterpart) computed at the busy hour. Also, those designing local measured rate options would be interested in revenue effects as well as distributional effects (how the effect distributes across the population) of particular rate designs. The former would require knowledge of usage levels and elasticities of demand while the latter would require data on how usage among households differs according to the demographic characteristics of the households. As a final example, a commission may want to know the distributional impact of certain pricing policies. This would suggest an objective of producing data that can be used to relate usage to demographic characteristics before and after the policy change. Policies regarding extended area service (EAS) are most often supported by actual usage data that shows the impact on costs and revenues of various EAS plans.

Study objectives should contain two basic elements. First, they should indicate what knowledge we hope to gain. Second, they should indicate what acceptable amount of uncertainty about that knowledge can remain after the study. We recognize that it is often very difficult to be as explicit about the second basic element as it is about the first. The example given earlier to estimate SLU with a 95 percent confidence interval no wider than plus or minus 1 percent is a case in point. While the phrase "95 percent confidence interval no wider than plus or minus 1 percent" is a very precise and explicit way to state the amount of uncertainty about SLU that would be acceptable, it is very hard to get those who will use the study results to make such statements. There is a further problem in that the study might cost one-tenth as much if the 95 percent confidence interval could have a width of plus or minus 2 percent. In this case the larger interval might have been the objective rather than the narrower interval. However, as difficult as it may be, we recommend that an attempt be made to state as explicitly as possible the allowable level of uncertainty as a part of a study objective. If it turns out that that part of the objective has a significant (and perhaps unnecessary) impact on the cost of the study, then certainly the "define study objective" step can be revisited and perhaps modified.

A.2 Identify Candidate Dependent and Independent Variables

Dependent variables are the response variables that will be measured for the individual sampling units selected as participants in the study. Of course, in our present context these response variables are usages. Examples are peg count (number of calls); originating minutes of use for local, intrastate toll, and/or interstate toll; originating minutes of use by time of day; and minutes of use over specific routes. It is often the case that studies can produce information about ancillary response variables (dependent variables) such as whether or not subscribers have elected optional measured service.

Independent variables represent those characteristics of individual sampling units thought to have some influence on the units' response variables. This influence is to be modeled with a postulated mathematical relationship between a dependent variable and the independent variables. The purpose of gathering data is to estimate parameters of the model. Examples of independent variables are region of the state, exchange, household income, age of head of household, number of teenagers in the household, and sex of head of household.

Finally, the sampling unit is the basic item on which data are to be gathered and recorded. Examples are residential households, single-line businesses, multiline businesses, whole exchanges, a particular interoffice trunk, and measured-rate residential customers.

This step, identifying candidate variables and sampling units, is one that may need to be revisited in the course of developing a study plan. The most common reason is that data for the variable intended for the study may prove to be very difficult to obtain so that some other variable whose value is more easily obtained may be selected to serve as a proxy. For example, peg count is often used as a proxy for usage.

A.3 Determine Method of Measuring Dependent Variables (Usage) and Independent Variables

The measurement of usage can be very easy (inexpensive) in some electronic offices where a change in a customer code can initiate automatic measurement. The measurement of usage can also be very difficult (expensive) in step-by-step, cross-bar, and some electronic offices where special equipment for monitoring and recording, or additional traffic-sensitive equipment for remote monitoring, or a special generic computer program may be needed in order to initiate measurement.

Consultation with the participating telephone company officials can help to minimize the cost to the company of a usage study. Some purposes of usage studies such as estimating SLU, are sufficiently important and so frequently done as to justify both the investment in additional programs and equipment, and the costs of measurement--even in the high cost offices. However, in the larger telephone companies there should be enough electronic offices with the requisite programs already installed so that a study that concentrates its sample in these offices will minimize the cost of the study to the company, yet still offer sufficient opportunity to design a valid study. On this point the literature indicates there is good reason to believe that studies done only in electronic offices will contain a bias in the results. In part, the bias arises from the fact that over the years telephone companies have selected offices to install electronic equipment because of the particular calling characteristics of the subscribers located in the area served by the office. As time passes and older offices are replaced with electronic offices for reasons other than their calling characteristics this bias will be reduced. But it is still the case that a SLU-type study could be affected by the bias introduced through restricted office selection.

In most cases, values of the independent variables will not be available from the telephone company. The independent variables represent characteristics of the sampling unit (household) whose telephone usage is to be measured. The only way to obtain these characteristics is through a direct survey of the sample done as a part of the usage study or a comparable survey that has already been done for other purposes. If the independent variables are demographic variables, then the U.S. Bureau of the Census has summary statistics that can be substituted for the unknown actual values in the households.

By substituting census summary statistics for actual household information two potentially serious problems are avoided. They are:

1. A survey of a sample of households is expensive if done so as to maximize the response rate

2. Sampled households lose their anonymity if a survey is done

On the other hand, census summary data substituted for actual household data will give lower quality statistical results. Nevertheless, the question of whether or not to survey the sample is a critical determinant of the cost and quality of the study. As can be seen in figure 3-1, our recommendation is to rely upon existing survey data to conduct the study to a conclusion and only then decide whether a survey would sufficiently improve the results to justify its cost.

A.4 Use Existing Sample?

A second key question that can affect the cost of the study to the telephone company involves whether to use usage data routinely collected in SLU studies or to request that measurement take place on a new sample. The basic issues here are how well suited the existing SLU data are to the study purpose and the cost of measuring usage on some additional lines. Assuming SLU studies are conducted on a random sample of households and stratified by exchange we have analyzed the potential results of such a plan in one West Virginia exchange (Charleston). Our analysis shows that while the SLU sampling plan is well suited to estimating SLU, it is not well suited to the purpose of examining the relationship between demographics and usage. In particular, if one intends to estimate the parameters in a linear model that relate usage to the age of head of household, income of household and the presence of children in the household, then (as will be seen later) the variance of the estimators can be as much as ten times as large as would be achieved using the same number of observations with a feasible, D-optimal or A-optimal (henceforth, optimal) sampling plan that stratifies on demographic characteristics. Thus, a significant gain in the statistical quality of the results can be achieved by using a new stratified sample rather than the existing SLU sample of households.

As mentioned previously, the cost of measuring usage on a sample of households can be reduced to a minimum by restricting the sample to those offices that already contain the appropriate equipment and programs to measure local telephone usage. Still, the costs are not zero. For example, the participating households in the sample must be identified by the company, then usage measurement can be initiated by processing a service order for each household. In our work with the C & P Telephone Company in West Virginia, we requested cost estimates for the various tasks associated with a special study. Electronic offices not possessing the generic program needed for special studies would cost $5,000 to $10,000 to acquire and install the programs according to the company. Arrangements to make measurement possible in a nonelectronic office could cost in the neighborhood of $30,000. The company also reported that its average cost for processing service orders over their last test period was thirty-five dollars per order. One should recognize, however, that this figure is an average embedded cost obtained by dividing the total annual cost of the equipment and personnel used to process orders by the number of orders processed. Thus, if a commission were to request a study with sufficient lead time to allow the company to schedule the processing of the additional service orders so as to not require hiring new people or increasing overtime, then the real additional cost of these orders could be virtually eliminated.

While the cost-quality tradeoff discussed above is specific to West Virginia we doubt that it would be significantly different in most other states. Our general recommendations would be to plan on not using the existing SLU data exclusively. Instead, a systematic sampling plan should be developed if at all possible. The SLU data can then be used to supplement the data obtained under an optimal sampling plan. As an alternative, should the cost of working with the SLU data be relatively low for the commission, it could be obtained and analyzed first, with the intention that if statistically unsatisfactory results were obtained, the commission would implement an optimal sampling plan.

Since the present step, A.4, is a decision point, two branches emanate from it in figure 3-1, the "A" branch having to do with the stratified sampling plan, and the "B" branch having to do with the SLU data. We shall first consider the "A" branch through A.8, then consider B.1 and B.2. Step A.9 is common to both branches.

A.5 Collect and Analyze Population Summary Data on the Independent Variables

Basically, as discussed in chapter 2 an optimal sampling plan is achieved by dividing the population into several different and carefully selected population groups called strata and then randomly sampling each stratum. The independent variables identified in step A.2 should be measures of the characteristics describing the strata and should be used to classify individual sampling units into their respective strata. For example, if family income is the only independent variable then one might wish to divide the population into perhaps three strata. One stratum would contain all low income households, the second would contain all medium income households and the third would contain all high income households. Of course, many more than three strata could be used if the researcher felt it necessary to approximate more closely the continuum that income really is. As another example, suppose both income and the presence of children are the independent variables. In this case six strata may be appropriate using two sets of low, medium, and high income groups, one set with children and one set without.

The problem is that in order to stratify households one must know the characteristics of all the households in the population. To find out such information for a large group is an enormous task, so that one must search for one or two easily observable characteristics of households that correlate with the independent variables in the study and then stratify the households using these correlated variables. In demographic studies, such a correlated variable exists: the location of the household, since people tend to live in homogeneous neighborhoods. If independent variables or some variables highly correlated with them are not easily observable, and if the population is large, then there is almost no hope of developing an optimal sampling plan at a reasonable cost.

Thus, what is meant by gross independent variable data are the summary statistics on the independent variables. These summaries should be matched with the correlated variables. For example, the median income of households in a particular census block group can be obtained from U.S. Bureau of the Census tables. Thus a summary statistic (median income) is readily available for an independent variable (income) and its corresponding correlated variable (location as indicated by block group, tract, county, and for state).

Once all relevant U.S. Census data and maps have been acquired for sample design purposes the additional information needed is the exchange boundary maps for those exchanges where measurement of usage is expected to take place.

The organization of the collected materials consists of inventorying and cataloging the population living within the telephone exchanges where the study is to take place. In the case of a demographic study this entails transferring exchange boundaries from telephone maps to U.S. Census maps of the greatest available detail.

This will allow one to easily determine and list those GASs[2] contained in an exchange area.

The reason the whole population within the appropriate exchange area must be catalogued in this step is that such information is necessary in order to determine the appropriate stratification.

A.6 Specify Strata

In this step three basic questions are to be answered. They are

1. How many strata are to be sampled
2. Which strata are to be sampled
3. What fraction of the total number of sampling units sampled should be in each strata

In the context of a demographic study these questions need to be stated more specifically, as they are in chapter 2: (1) how many GSAs are to be sampled, (2) which GSAs are to sampled, and (3) what fraction of the total number of households sampled should be in each GSA.

It is in this step that many of the technical issues discussed in chapter 2 arise. Chapter 4 contains the best illustration of these issues since it contains a detailed description of this step as it was performed in the West Virginia pilot. The reader is referred to these two chapters for detailed discussion of the relevant issues.

A.7 Evaluate the Design and Determine Sample Size

An evaluation of numerous rival designs naturally occurs as part of step A.6, but such evaluations are based on simple models derived

---

[2]GSAs were defined in chapter 2 as geographic sampling areas--a geographic area in which household samples are located and for which summary federal census statistics may be used as a proxy for the actual statistics of the sampled household. Our recommendation is that they be initially defined as the smallest possible such areas.

from complex situations through a series of assumptions. In reality, the sampling plan may be inadequate to accomplish the purposes of the study with the kind of precision desired. Therefore, we recommend the use of a simulation model like the one discussed in appendix B. Such a model can be used to generate hypothetical household usage data according to the proposed sampling plan. The whole sampling plan can be repeated many hundreds of times if desired. There are three specific advantages to doing this: (1) Sets of realistic data are generated that can be used to develop and practice data analysis techniques. (2) Parameter estimates can be computed from each data set using the data analysis procedures; the collection of all such estimates obtained from all replications of the sampling plan allows one to evaluate the distribution of parameter estimates. This helps to validate the methods used in step A.6. (3) One can test the robustness of results as related to the various assumptions that have been made to this point. While analytical calculations can be made to determine the total sample size that should be used in the study, the simulation program will be helpful in refining the sample size while validating the calculations. An explanation of the total sample size calculation is found in chapter 2, the West Virginia example calculation in chapter 4, and simulation results for West Virginia in appendix B. The reader is referred to those chapters for further details on this step.

A.8 Obtain Usage Data for Sample

This step pertains to the issuance of a data request to the participating telephone company. The data request should ask the company to identify a random sample of a certain number of telephone lines within each of the strata specified in step A.6. It should then ask the company to monitor the amount of usage on those lines over a predefined period of time and ask the company to report the results in a specific format. Appendix D contains a model data request developed during the West Virginia study.

A.9 Analyze the Results and Estimate the Benefit That Might Be Made with Improved Independent Variable Data

In this step a complete analysis of the data can take place as if all data collection is complete. The step is a point in the planners process flow chart where the A.5 through A.8 branch and the B.1, B.2 branches merge. Standard statistical packages can be used to analyze the data collected from the optimal stratified sampling plan (steps A.5 through A.8). Analytical techniques employing maximum likelihood methods should be applied to analyze any data collected from preexisting random samples (steps B.1 and B.2). The reasons for these different analytical approaches are discussed later in this chapter. In either case a careful analysis will yield estimates of the error variance, the parameters of the model being analyzed, and the variances (or confidence intervals) of the model parameters.

At this point the study could be considered complete depending on whether the quality of the results is sufficient. If the quality is not sufficient one may wish to return to step A.4 and decide, for example, whether to supplement the data already collected with the SLU data that the company had collected during the same time frame as the study. Another alternative would be to increase the sample size of the designed sampling plan and remonitor usage for the entire sample. Yet another alternative that has the potential of greatly improving the results (although at great cost and not without other problems) consists of replacing the census data on the population with better household data obtained by actually conducting a survey. This last alternative is the major decision point in the flow chart labeled A.10 and is discussed later.

B.1 Collect and Organize Gross Independent Variable Data on the Sample

This step is similar to step A.5 in that easily obtained measures that correlate with the independent variables describing a household are used as a proxy for actual household data. In a demographic study, the location of the household is useful in identifying the U.S. Census block group summary statistics that would serve as the proxy. The step is different from A.5 in that summary statistics need be matched only to the households in the companies' SLU sample rather than a complete inventory of the population as is required in step A.5. The reason is that in the controlled study described in A.5 through A.9, the plan decided where sampling should be done on the basis of the demographics of the various block groups (or GSAs), but in the SLU sample the sample has already been determined, leaving only the step of identifying the demographics of the participating household. If privacy issues are not a problem, the addresses of the households in the SLU study could be requested from the company. U.S. Census maps and address locations could then be used to identify the appropriate county, city, tract, and block group codes for each sampled household. The U.S. Census summary statistics for each household's location code would then be recorded as the proxy for actual household data.

If privacy issues are a problem, then the company could be asked for address ranges for each SLU household (e.g., 1200-1500 West 15th Street) or the company could be asked to provide the U.S. Census location code for the block groups containing SLU households.

B.2 Obtain Usage Data for Sample

This step is similar to step A.8 except that the sampling plan has already been predetermined by the company so that all the data request need do is request the usage data and the format in which the data should be provided.

A.10 Will Sample Be Surveyed?

This step is the second of the two key decisions that affect both the quality and the cost of doing the study. The simulation results in appendix B indicate that for the West Virginia study a substantial improvement in the quality of statistical results is possible if the sample is surveyed in order to gain better information about the household characteristics. However, such a survey is not without its problems. First, the anonymity of households is lost, which raises privacy issues and possibly legal questions. Second, if only a mailed questionnaire is used, the response rate can be expected to be less than 20 percent and one cannot rule out the introduction of bias into the study. If both a mailed questionnaire and a follow-up telephone call is used, the response rate may be as high as 75 percent and less survey bias would be introduced, but the approach is much more costly and, of course, could not be used on households with unlisted numbers, which would, in itself, bias the survey somewhat.

Our recommendation is that unless the privacy issues are so severe as to strictly prohibit it, a survey of the sample should be undertaken.

C.1 Design a Subset of the Sample To Be Surveyed

Since the cost of the survey is generally a function of the number of questionnaires administered, it is possible that some cost savings can be achieved by surveying only a part of the sample. For example, if the A.5 to A.8 branch has been undertaken in a demographic-usage study, then the study design will consist of a list of GSAs in each of which a sample of size n was taken. One approach would be to survey one-half of n in each GSA. The households so surveyed would have their characteristics updated in the data base to reflect the survey results and the remaining households in a GSA could be updated by the average of those in the GSA that were surveyed. If the B.1-B.2 branch has been followed, then the "design" consists of a list of GSAs where SLU households are located, together with the code number in each GSA. One might then select about half or fewer of the SLU households from those available, using an approach similar to step A.6, and then survey those selected. The difference between this design approach and that given in A.6 is that in A.6 the population consisted of all households in an exchange while in this step the population consists of all SLU households in an exchange.

C.2 Survey the Subset of the Sample and Update the Data Base of Independent Variables

The survey technique most likely to give the highest response rate and be the most reliable is to mail to each of the households in the survey sample a questionnaire with a cover letter explaining that someone representing the commission would be calling to obtain the answers to the questionnaire. Households with unlisted numbers would be asked to mail the completed questionnaire back. Once the household information has been associated with usage data in a computer data base all connections between the information and telephone numbers, names, and addresses can be destroyed so as to restore the anonymity of the participants. Such controls will probably need to be carefully worked out to avoid problems with the privacy issue, but specific measures are beyond the scope of this study. In fact, in the West Virginia pilot study a decision was made a priori not to do a survey of the households.

Discussion

In this chapter a general set of steps for accomplishing a usage study was presented. In several cases certain courses of action were recommended, but two key decisions remain. One is whether to use the SLU sample or to obtain a new sample based on an experimental design better suited to the purposes of the study. The other decision pertains to whether or not a survey of the sample (or some portion thereof) should be conducted. The best decision in these two cases depends on the particular situation in a state and the choice of variables included in the study. Based partly on our experiences in the West Virginia study and on experience in other studies, some very rough estimates of costs were made. These are displayed in table 3-1.

TABLE 3-1

ESTIMATES OF DIRECT COST ELEMENTS

	SLU Sample	Optimal Stratified Sample
Do not survey	\$5•N	\$3•N to \$38•N Depending on schedule
Survey	\$55•N	\$53•N to \$88•N Depending on schedule

Note: N is the total sample size.
Source: Authors' estimates based on some C & P telephone company informal estimates

Some of the costs in table 3-1 would be borne by the company and some by the commission, depending on how the tasks are divided. The range of costs in the stratified Sample column is intended to reflect the various interpretations of the cost of processing a service order as discussed in chapter 1. No cost for analysis nor any fixed costs have been included, and it is assumed that the sample would be taken in offices already possessing measurement capability. While there are differences in the cost of a study depending on the decisions made in the two key instances, there are also differences in the statistical quality of a study design resulting from these two key decisions.

An understanding of this tradeoff between cost and quality can be obtained by performing simulation studies of the various sampling plans. Appendix B contains a description and the prototype computer code of a simulation model developed and used to examine this issue in the West Virginia case. The study by Park et al., provided the basis for defining the relationship between household usage measured as number of calls raised to the .27 power and the three demographic parameters:

1. Logarithm of household income in thousands of dollars

2. Age of head of household in scores of years

3. Logarithm of the household size

For purposes of our experimentation with the simulation model, we used several values derived from the Park study. In particular we used an intercept value of 2.1, and rates of change in the usage measure for the three variables listed above of -.004, .068, and 1.1 respectively. The variance of the household usage measure that is unaccounted for by the demographic variables was .314. Using these figures and assumed distributions of the population parameters present in Kanawha County, West Virginia, simulation runs were made for four cases.[3] In each case a sample of households was generated along with the telephone usage of the sample households. It was expected that the data generated by the simulation program had the same appearance that a set of actual data would have had if the sampling plan had been implemented. It was also expected that the simulated data had the same statistical properties as actual data would have had. The four cases were as follows:

Case 1: A stratified sample is planned but census data are used in place of actual household demographic data for analytic purposes.

Case 2: A stratified sample is planned and a survey is conducted to obtain actual demographic data on a household basis.

Case 3: A random sample is used and a survey is conducted to obtain actual demographic data on a household basis.

---

[3]Charleston is the major city in Kanawha County and the Charleston exchange provides telephone service to almost all of the county.

Case 4: A random sample is used but census data is used in place of actual household demographic data for analysis purposes.

Case 4 is a more complicated situation than the other three cases. The complication arises from the fact that

1. The design is random rather than fixed

2. Random errors in the demographic variables are present because GSA averages are used as proxies for individual household demographic values[4]

A short-hand reference to these two items is random design and errors in variables. When either of the two items does not apply, as in cases 1 to 3, then least squares is an appropriate analysis criterion leading to unbiased estimates of the variable coefficients. When errors in variables are present in a random design, the least-square criterion results in biased estimates of the variable coefficients making maximum likelihood estimation techniques more appropriate.[5] As a practical matter one may want to avoid case 4 situations since many statistical analysis packages do not contain general maximum likelihood techniques and if they do contain them a usual assumption is that the distributions of errors in the variables are joint normal distributions. This assumption is at best of questionable validity in the case of demographic data (e.g., household size is a particular problem with respect to this assumption).[6] We did not have computer programs for maximum likelihood techniques available in the course of this study, and we did no attempt to acquire such programs (which are

---

[4]Medians or some other summary statistic may be used place of averages.

[5]For a discussion of these issues see for example, M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, 3d ed., 3 vols. (New York: Hafner Publishing Co., 1973) 2: 391-435.

[6]A computer program named LISREL that can be used to obtain maximum likelihood estimates is referred to in R. E. Park et al., Charging For Local Telephone Calls (Palo Alto: The Rand Corporation, 1981), R-2535-NSF.

commercially available) because using them would have involved issues beyond the scope of this study. Instead we took an approximate approach that we felt was adequate for comparative purposes, in evaluating the quality of data that would be obtained in a case 4 situation. In this approach we treated the case 4 design as a fixed design that had been arrived at randomly. For example, suppose a completely random set of households is selected and N_i of these households are found to be in the i^{th} GSA, i=1,...,M, when M is the number of GSAs available. A fixed design would now mean that all subsequent replication of the sample would be constrained to locating exactly N_i randomly selected households in the i^{th} GSA, i=1,...,M. This constrained sample selection allows the use of least squares and, more importantly, allows the variance of the estimates of variable coefficient to be computed from $\sigma_e^2\ (X^TX)^{-1}$ where X is the design matrix and σ_e^2 is the error variance in the model. We recognize that this approach is an approximation in that the calculation ignores the bias and the variance in the least-squares estimates that would be introduced if the N_is could change from replication to replication (the actual situation in case 4).

It should be noted that even in the calculation $\sigma_e^2\ (X^TX)^{-1}$ the initial design specified in the design matrix X is arrived at randomly. Our approach here was to simulate the completely random selection of households in Kanawaha county, place each household in its GSA, then form the X matrix using the codes for GSAs and the relative frequencies with which each GSA contained one of the sample households. As mentioned earlier the first three cases did produce usage data and demographic data similar to that which would have been obtained in a field study under each of the three cases. These data were analyzed using General Linear Model analyses instruction sets in the Statistical Analysis System (SAS) computer package.[7]

In order to compare the results of the four cases we use the variance of the estimates of the model parameter as introduced earlier. The reason for this is that if a study goal is to achieve a certain

---

[7]SAS Institute Inc., SAS User's Guide: Basics, (Cory, N.C.: SAS Institute Inc., 1982).

precision in the estimates of demographic effects on usage, the variance of these estimates is needed to determine the sample sizes required to achieve the specified precision. The resulting variances of the estimators in the four cases are given in table 3-2.

TABLE 3-2

NORMALIZED VARIANCES OF THE ESTIMATION OF THE COEFFICIENTS OF THE VARIABLE LOG INCOME (X_1), AGE OF HEAD OF HOUSEHOLD (X_2), AND SIZE OF HOUSEHOLD(X_3)

	SLU Sample	Optimal Stratified Sample
	Case 4	Case 1
No survey	$\sigma^2_{\beta_1} = 8.6$	$\sigma^2_{\beta_1} = 1.1$
	$\sigma^2_{\beta_2} = 40.2$	$\sigma^2_{\beta_2} = 3.6$
	$\sigma^2_{\beta_3} = 30.3$	$\sigma^2_{\beta_3} = 6.8$
	Case 3	Case 2
Survey	$\sigma^2_{\beta_1} = .58$	$\sigma^2_{\beta_1} = .31$
	$\sigma^2_{\beta_2} = 1.09$	$\sigma^2_{\beta_2} = .85$
	$\sigma^2_{\beta_3} = 3.11$	$\sigma^2_{\beta_3} = 3.0$

Note: $\sigma^2_{\beta_1}$ is the estimate of the variance of the linear model estimator, which is the coefficient of the variable X_i, i=1,2,3.
Source: Authors' calculations

The advantage cases 1 and 2 have over cases 4 and 3 respectively as seen in table 3-2 is due to the fact that nearly D-optimal designs were selected for the stratified sample. This means that the amount of multicollinearity naturally present in the population in Kanawha County was minimized in the stratified sample leading to a better quality of result for the same sample size. The random sample, of course, being random makes no attempt to correct for the multicollinearity, which leads to higher variance of estimation. It can also be seen in table 3-2 that a substantial improvement occurs when a survey is conducted especially when a random sample is involved. The reason is that without a survey the household characteristics of a sample household are not known exactly and so must be measured inaccurately by using GSA average (or median) statistics as proxies for the sample households instead of their actual values. As stated earlier, this represents an errors-in-variables situation which again, contributes to a higher overall variance in estimators. What the simulation data enabled us to determine as the relative magnitude in the increases in variances due to these two sources. The size of the differences between the "survey" and " no survey" cases as compared to the differences between the SLU and optimal sample cases were something of a surprise: the improvement over the "SLU sample, no survey" case is extremely large if a survey is conducted, while the "optimal stratified" sample cases do not show such dramatic improvement from conducting a survey.

The fundamental question remains--which case produces the most cost-effective study? While we cannot answer this question in general because some of our results are specific to Kanawha County and the opportunities for optimal samples will vary from locality to locality (as will the details of census data and cost data), we can combine the information in tables 3-1 and 3-2 in order to make rough cost calculations and pose a solution for Kanawha County. Although it is rather complicated, the approach we used in obtaining this solution should be generally applicable.

Suppose now, as an example, it was concluded that a sample size of one hundred obtained under the most favorable conditions (i.e., case 2) was sufficient to produce narrow enough confidence intervals for each β_i to satisfy the objectives of the study.

One can then apply the information of table 3-2 to determine what sample sizes would be required in the other three cases in order to produce confidence intervals that have at most the same width as the case 2 sample of one hundred. These relative sample sizes have been estimated and are given in table 3-3.

TABLE 3-3
EQUIVALENT SAMPLE SIZES

	SLU Sample	Stratified Sample
Do not survey	Case 4: Income------2,775 Age---------4,730 Size--------1,010	Case 1: Income-----360 Age---------425 Size--------225
Survey	Case 3: Income------190 Age----------130 Size---------104	Case 2: Income-----100 Age---------100 Size--------100

Source: Author's calculation using figures in table 3-2 and equation 2.13

In table 3-3, income, age, and size refer to the independent demographic variables of importance in the West Virginia study.[8] The cell in the lower right corner is case 2 where a sample size of 100 is chosen for illustrative purposes. The other cells show the sample sizes that would be needed in order to estimate with equal confidence intervals the parameters of the corresponding variables of a general linear model. For example, if usage data was collected for a new optimally stratified sample of 100 households, which is the baseline case in the table, a certain confidence interval would result for each of the parameters in a linear model with income, age, and size as the independent variables. If a SLU sample of 190 was used and a

[8]Actually, the presence of children was one of the variables of interest, but household size is used as a proxy in this analysis.

survey was conducted, the confidence interval for the parameter corresponding to income would be equal to the comparable interval in the baseline case. With that SLU sample of 190 the age and household size parameters would have smaller confidence intervals than in the baseline case since they would require only 130 and 104 sample sizes respectively in order to have the same size confidence intervals as the baseline case had with its sample of size 100. The rest of the table is interpreted in a similar manner.

Suppose the costs from of table 3-1 are combined with the sample sizes of equivalent or better quality for all three parameters. Then the total direct sampling costs of the one hundred sample baseline or equivalent studies is given in table 3-4.

TABLE 3-4
COSTS OF EQUIVALENT STUDIES

	SLU Sample	Stratified Sample
Do not survey	n = 4,730 $ 23,650	n = 425 $1,275 to $16,150
Survey	n = 190 $ 10,450	n = 100 $5,300 to $8,800

Source: Authors' calculations

As can be seen in table 3-3, a significant cost saving is possible if steps A.1 through C.2 are followed. If the final step is A.10, a study of relatively low cost can be conducted. However, it should be pointed out that these cost figures are based on several subjective estimates and, where two figures are given, the smaller figures assume that the processing of service orders is requested with sufficient lead time so that the work can be done without requiring additional resources. It is also the case that these results are peculiar to West Virginia and may be quite different in other states. It should be further pointed out that the set-up, analysis, and other indirect costs are probably not equal for all the cells in table 3-3, and they are not included in table 3-4. For example, it is often the case that several hundred survey attempts would need to be made in order to get one hundred household respondents. The results displayed in these tables can, however, provide a valuable guide, supplemented by good judgement, as to the probable best decisions to make for the two key decision points in figure 3-1.

CHAPTER 4

THE WEST VIRGINIA PILOT STUDY

At the outset of this study six widely dispersed state commissions were asked about their willingness to participate in the usage study as a possible pilot state. Most expressed interest; West Virginia was finally selected. Our experience there in applying the method given in chapter 3 is reported in this chapter. In this case, budget limitations for the pilot study caused an a priori determination that no demographic survey of the sample could be made. Also, since the study developed some novel study methods, one of the main purposes of the pilot was to try them out. For this reason we followed steps A.1 through A.8, but the data request to the telephone company also asked for SLU data (steps B.1 and B.2). Assuming the company cooperates, data will ultimately be made available to compare the results of the two methods with our theoretical calculations and simulation results. In any event the data will not be available in time to be included in this report.

Details of the Study

Our description of the pilot study will follow the planning method depicted in figure 3-1 in a fashion similar to the previous chapter.

A.1 Define Study Objective

A meeting was held with members of the West Virginia staff who were interested in a usage study. As a result of that meeting it was determined that the study should be designed to estimate relationships between the originating telephone usage in a household and variables pertaining to the household.

A.2 Identify Candidate Dependent and Independent Variables

At the meeting described above, the candidate variables were also discussed and decided upon. The dependent variables (usage) in the study would be originating local exchange usage and interexchange (or toll) usage. The staff felt it would be useful to have these usage figures separately for 3 or 4-hour blocks of time throughout each day, and to the extent possible, information on distance and duration of each call. The independent variables can be grouped into two categories. First, demographic variables consisting of age of head of household, household income, and the presence of children in the household. The age and income variables were of particular interest to the commission staff, but children were included because other studies have shown it to be an important variable. The second grouping of variables is geographic variables consisting of neighborhood type (urban, suburban, rural) and region within the state (Northwest, Southwest, Central, etc.).

A.3 Determine Method of Measuring Dependent Variables (Usage), and Independent Variables

A meeting was held jointly with the commission staff members and representatives of the telephone company. A number of usage measurement techniques were discussed along with rough estimates of the cost of each. Since local measured service options are available in many West Virginia communities, it was thought that additional households or businesses could generally be measured at little or no additional cost, but there were some communities for which this was not the case. For example, in some exchanges the local measured rate option is offered even though the equipment there cannot provide the requisite measurement. Instead remote measurement is provided by the equipment in a nearby office for those few customers who have elected measured service. As a result, traffic-sensitive equipment is involved in providing the link between the two offices, and any measurement of additional lines in a study would require an increase in the traffic-sensitive facilities. As another example, some offices contained the requisite measurement capabilities for measured service but did not contain the generic program that would allow the use of customer codes other than the measured rate code. Therefore, additional lines measured in these offices would be indistinguishable from measured rate customers and erroneous billing would result.

The outcome of the meeting was an agreement to use only central offices with measurement capability on non-measured service lines so as to minimize the cost of initiating measurement of the sample of lines in the study. At that time the company was asked to provide a list of such exchanges and a copy of the exchange boundary maps for each of the listed exchanges. These materials were needed for steps A.5 through A.7.

A.4 Use Existing Sample?

In West Virginia a telephone company-wide random sample of about 1,300 households is routinely measured in order to estimate SLU for separations purposes. Our analysis of a purely random sampling approach within Kanawha County indicated that it was well suited for estimating SLU, but was not particularly well suited for estimating the relationship between demographics and usage. However, our analysis was without the benefit of any actual usage data and is certainly subject to check with actual data. This is one reason that our answer to the question at step A.4 is both "yes" and "no." By that we mean that the SLU data as well as usage data from a stratified sampling plan will be requested from the company. There are other reasons for this dual request as well. For example, it proved difficult to design a stratified sampling plan that was both feasible and useful in any exchanges other than the Charleston exchange. This particular problem would be less serious in more populous states, but more serious in less populous states. For this reason, we elected to employ the best available stratified sampling plan in Kanawha County in order to study the effects of the demographic variables on usage and then to employ the SLU sample to study regional effects and the effects of urban, surburban, and rural locations on usage. Although steps B.1 and B.2 have not been completed at this time as part of the pilot study, the model data request in appendix D asks for information about the SLU sample needed to accomplish B.1 and B.2 in such a way as to insure the anonymity of the participating households.

A.5 Collect and Analyze Population Summary Data On the Independent Variables

In the case of West Virginia this step consisted of obtaining and studying the demographic characteristics of the GSAs. Demographic characteristics of households in the seven counties from different regions of the state were tabulated in histogram form as shown in figures 4-1, 4-2 and 4-3. The characteristics shown are

x_1 = median income of the area,

x_2 = percentage of the households in the area with head over 64 years of age,

x_3 = percentage of the households in the area with at least one child, age 17 years old or less (in 1979).

Not illustrated in these figures is the presence of the joint variation of these characteristics. It is in fact the case that as x_2 increases both the median income and the percentage of households that have at least one child decrease. There seemed to be no clear relationship between median income and the percentage of households with children.

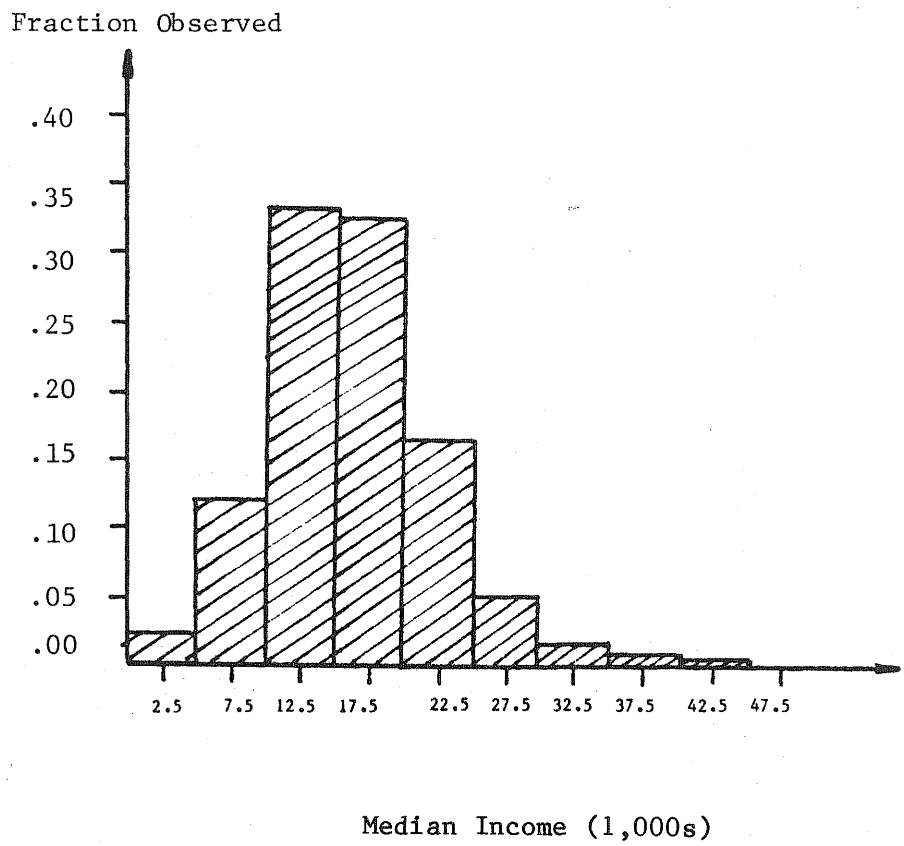

Fig. 4-1. Distribution of median income in thousands of dollars

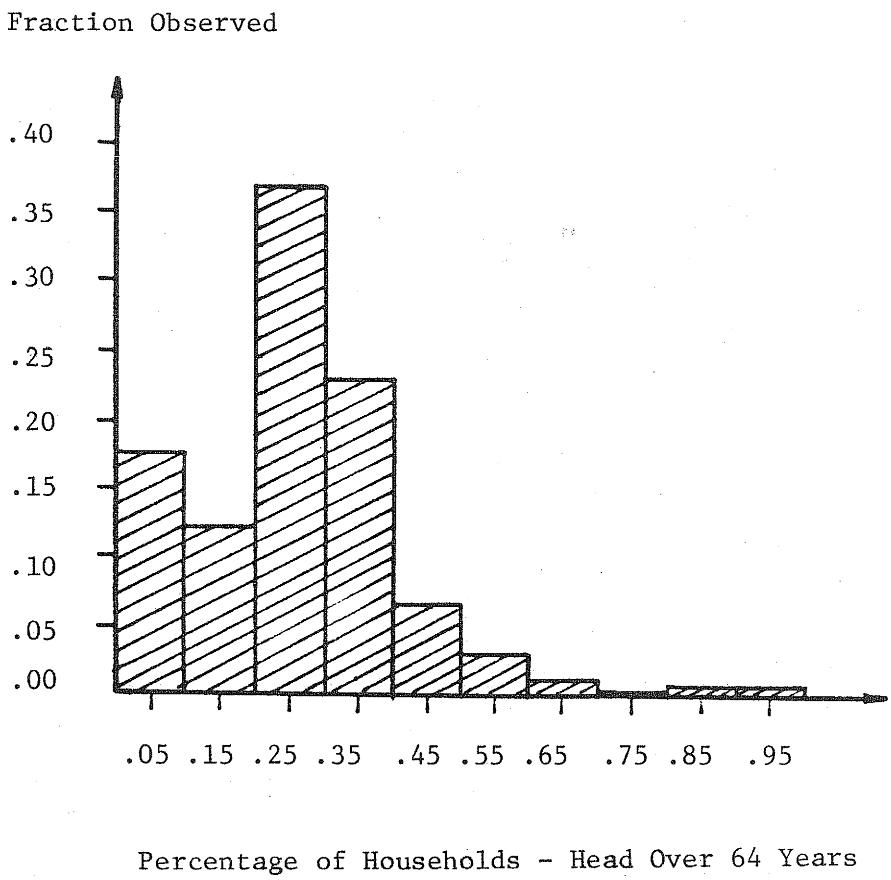

Fig. 4-2. Distribution of the percentage of households with head over 64 years of age

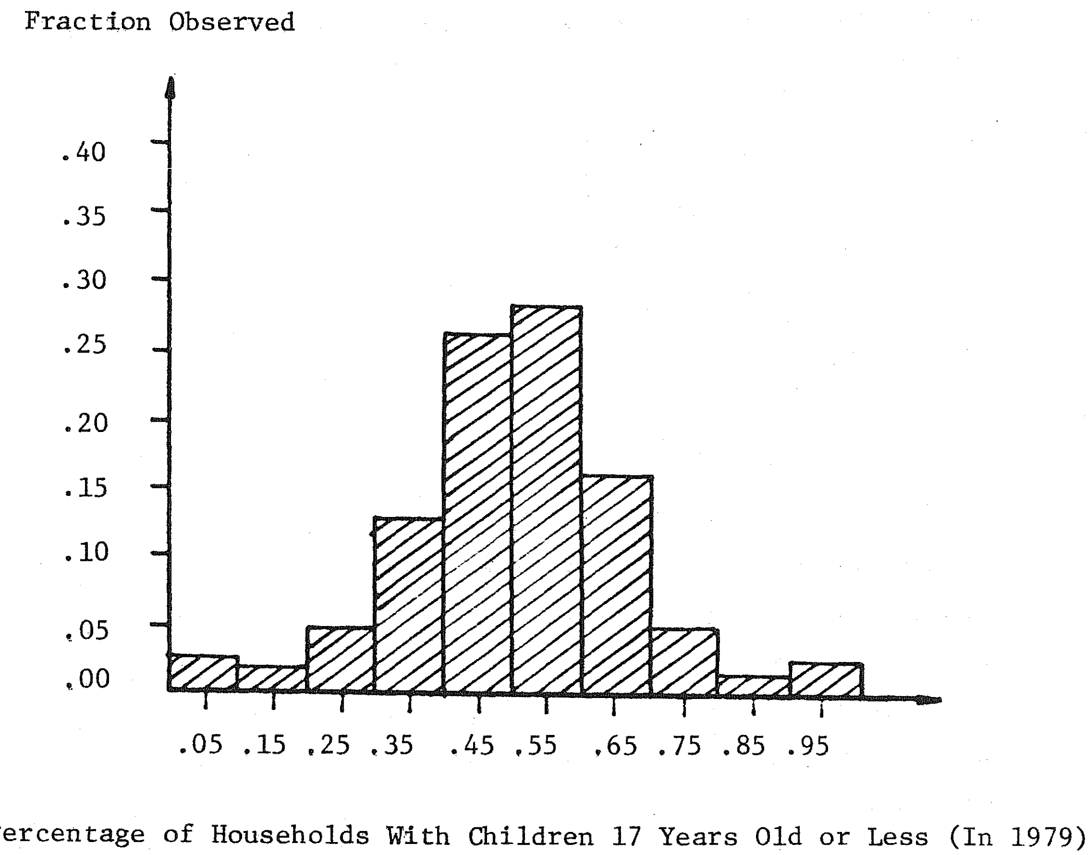

Fig. 4-3. Distribution of the percentage of households with children 17 years old or less (in 1979)

A.6 Specify Strata

In order to facilitate the design process for the sampling plan in Kanawha County, the values of the three demographic variables were coded as qualitative variables. After examining figures 4-1, 4-2, and 4-3, the following ranges were chosen in order to transform the numerical values of the variables into the qualitative codes -1, 0, and +1:

$$t_1 = \begin{cases} -1 \text{ if } x_1 \leq 12{,}000 \\ +1 \text{ if } x_1 \geq 24{,}000 \\ 0, \text{ otherwise} \end{cases}$$

$$t_2 = \begin{cases} -1 \text{ if } x_2 \leq .1 \\ +1 \text{ if } x_2 \geq .3 \\ 0, \text{ otherwise} \end{cases}$$

$$t_3 = \begin{cases} -1 \text{ if } x_3 \leq .35 \\ +1 \text{ if } x_3 \geq .65 \\ 0, \text{ otherwise} \end{cases} \quad (4.1)$$

A good deal of judgment must be exercised in choosing the ranges used to make the above transformation. The reason is that there are conflicting objectives involved. On the one hand it is desirable to achieve as large a separation as possible between GSAs coded -1 for a variable and those coded +1. For example, if GSAs with median income over $50,000 a year could be in a group coded +1 for income, and GSAs with median income less than $5,000 a year could be in a group coded -1, then any income effect on usage would be more readily detected than would be the case if all GSAs with median income more than $15,000 were coded +1 and those with median income less than $14,000 were coded -1. On the other hand, this wide gap between GSA coded +1 and -1 will cause the number of GSAs in each of these groups to be very small. In fact, in seven counties in West Virginia that we examined as part of the pilot, no GSAs had median income more than $50,000 and only two percent had median income $5,000 or less. Thus, what one must balance is having as much separation between -1 and +1 as possible while still insuring the existence of GSAs coded -1 and +1. The above coding scheme has the potential of dividing the population into twenty-seven strata--one for each combination of values of the qualitative variables.

Table 4-1 shows for Kanawha County the number of GSAs that possess each of the twenty-seven possible combinations of qualitative codes for their demographic characteristics. This tabulation illustrates the restraints on the experimental design. For example, notice the combinations of demographic characteristics with no GSAs (zero entry). In this case the zero entries cause the design problem to be somewhat complicated since no orthogonal fraction is possible using these qualitative variables to stratify the population. We therefore decided to achieve a "good" design based on the determinant criterion discussed in chapter 2.

TABLE 4-1

FREQUENCY OF GSAS IN EACH STRATUM DEFINED BY THE VARIABLES SHOWN, KANAWHA COUNTY

		t_1 =-1			t_1 = 0			t_1 =+1		
		t_3			t_3			t_3		
		-1	0	+1	-1	0	+1	-1	0	+1
	-1	1	2	4	1	31	7	1	6	6
t_2	0	5	13	4	17	87	7	2	14	0
	+1	6	6	0	2	12	0	1	0	0

Source: Authors' coding of 1980 census data using equation 4-1

A second restraint on the experimental design results from the total number of households within each stratum since that may limit the number of households that can be sampled within the stratum. These are given in table 4-2 where it is seen that the most restrictive strata are the ones defined by (t_1 = -1, t_2 = -1, t_3 = -1), (t_1 = 0, t_2 = -1, t_3 = -1), and (t_1 = 1, t_2 = -1, t = -1).

TABLE 4-2

FREQUENCY OF HOUSEHOLDS IN EACH STRATUM DEFINED BY THE VARIABLES SHOWN, KANAWHA COUNTY

		t_1 =-1			t_1 = 0			t_1 =+1	
		t_3			t_3			t_3	
	-1	0	+1	-1	0	+1	-1	0	+1
-1	34	185	243	10	9994	1012	55	2128	1341
t_2 0	1285	3632	1397	3877	29420	1048	223	5024	0
1	1660	929	0	332	1258	0	203	0	0

Source: Authors tabulation of 1980 census data

A Preliminary Design Based on Qualitative Variables

As discussed in chapter 2, it is convenient to consider first the fraction of observations to be taken in each stratum, treating the question of total sample size as a separate issue. In order to decide which GSA's to select we will employ both the D-criterion and the A-criterion.

For our purposes here it is convenient to number consecutively all twenty-seven strata. Table 4-3 shows such a numbering scheme. Because we have chosen to code all design variables with -1, 0, or +1, these problems are simplified to determining which of the twenty-two existing GSAs should be used for taking samples. The proportions of the sample taken in the GSAs are all equal and add to one. The total number is then a separate calculation.

TABLE 4-3

NUMBERS USED TO IDENTIFY COMBINATIONS OF QUALITATIVE VARIABLES THAT REPRESENT EACH STRATUM

		$t_1 = -1$			$t_1 = 0$			$t_1 = +1$	
		t_3			t_3			t_3	
	-1	0	+1	-1	0	+1	-1	0	+1
-1	1	2	3	10	11	12	19	20	21
t_2 0	4	5	6	13	14	15	22	23	24
+1	7	8	9	16	17	18	25	26	27

Source: Authors' choice

Table 4-4 lists values of $DET(W^{-1})$ and TR(W) for several candidate designs. Notice that design 10 includes all feasible strata (twenty-two in all) and is inferior to all others that would use many fewer strata. An orthogonal design is included for comparison, but it is infeasible because stratum 9 contains no households. Table 4-4 also includes values in parentheses that are traces of the design matrix, except that they exclude the constant term in the analysis model. Since the present project is involved with estimating the coefficients of the demographic terms, the constant term is of no interest. Therefore, the values in parentheses may be more appropriate for comparison purposes if the trace criterion is used. After considering both the determinant and trace of the nine feasible designs listed in table 4-4, we conclude that design eight is the best overall.

TABLE 4-4

LIST OF CANDIDATE DESIGNS CONSISTING OF SELECTED STRATA

Design	Combinations (Table 5-3)	DET(W^{-1})	TR(W)*
1**	1,9,21,25	1.000	4.00(3.00)
2	3,7,19,23	.250	8.00(6.50)
3***	1,8,21,25	.563	5.11(4.00)
4***	1,6,21,25	.563	5.11(4.00)
5	1,3,8,21,25	.563	5.11(4.03)
6	1,3,7,21,25	.410	6.88(5.63)
7	1,3,7,15,19,25	.494	5.25(3.98)
8****	1,3,8,19,21,25	.642	5.50(3.62)
9	1,3,6,7,15,17,19,25	.523	5.60(3.93)
10	All 22 observations	.205	6.49(5.30)

*Terms in parentheses are the values excluding the constant term in the model

**The orthogonal design, which is infeasible

***Other good designs by both the determinant and trace criteria

****The "best" feasible design found by the determinant criterion

Source: Authors' calculations using equation (2.10) and 1980 census data coded by equation 4-1

A.7 Determination of Sample Size

Given the six strata determined in the previous section (design eight), we use the simplified calculation of the total sample size given in chapter 2. In order to make the calculations we asssume a requirement for each of the model parameters (excluding the mean) of a 95 percent confidence interval with width no greater than twice the value of β. Based on the -1, 0, +1 coding of the variables log (income), age, and presence of children, and information in the literature we should expect the individual C_is to be on the orders of magnitude of .001, .01, and .4 respectively. Thus C is calculated as follows:

$$[(.001)^2 + (.01)^2 + (.4)^2]/3 = .0534$$

Based on the simulations reported in appendix B we estimate the σ^2 is approximately 0.6 and with $\alpha = .95$, $Z_{.975}$ = approximately 2, m = 3, and tr(W) = 4.00 we have $N \geq (2(_2))^2(4.00)(.6)/((.0534)(3)) = 240$. Since the design consists of equal proportions of the total sample being allocated to the six GSAs, then forty households should be randomly sampled in each of the six GSAs.

It should be noted that if the sample size had been determined from the formula that ultimately selects the maximum of N_1, N_2, and N_3, the required sample would have been greater than the entire population of Kanawha county. The problem is that, according to the literature, income appears to have a very small effect on telephone usage compared to the error variance in the model. Thus, this latter sample size reflects a study design that attempts to use a large number of samples in order to accurately estimate a very small effect. This would probably not be a cost-effective design even if the population were large. This shows one of the advantages of the A-criterion, which avoids this pitfall, which is the basis for the sample size calculation involving the trace of the design matrix, and which was the one finally used for the pilot study.

Our sample size calculation is based on an assumption of an infinite population. That, in fact, is not the case, since there are just over thirty-thousand households in Kanawha County. In general, when a required sample is more than about 10 percent of a finite population, then sample sizes calculated on the basis of an infinite population may be as much as 5 percent too large. However, general results about sample-size adjustments when populations are finite do not apply well in the case we have here of an optimized, stratified plan. In fact the finite size of each sampled stratum is a bigger problem than is the total population size vis-a-vis the total sample size. This issue is discussed at some length in the next section.

Summary and Additional Issues

In the West Virginia pilot study five variables and their relationship to local and toll, outgoing-residential telephone usage were found to be of interest to the commission staff. A study was planned that consisted of a bifurcated approach. First, readily available state-wide SLU data would be requested that could be used to examine differences in usage due to regional differences and urban-rural differences. Second, a stratified sampling plan was developed for Kanawha County (Charleston) that would produce household usage data suited to the study of the three remaining demographic variables' effects on usage. West Virginia is not a heavily populated state and as a result has fewer people in tracted areas, fewer tracted areas, and less opportunity for developing good stratified sampling plans than larger states. As was indicated to us by the commission staff, it appears also that West Virginia is unique in having neighborhoods that are more heterogeneous with respect to the demographic variables we were examining than would be expected elsewhere. If this is true, the error-in-variables problem may be more severe in West Virginia than elsewhere. In any event the errors in the variables that result from using census demographic data as proxies for actual household data have an effect on the quality of the data that would be generated by implementing the sampling plan. However, our final sampling plan was based on an approach that was committed to a priori and, therefore, calls for the use of a stratified sample. It does not call for a survey of the sample that would have minimized the errors-in-variables problem.

Reflecting upon how the pilot approach worked out, it is our belief that a stratified design is still justified but that the use of actual household data in place of census data is almost certainly cost justified if the legal aspects of any privacy issues can be resolved. We also believe it is worthwhile to conduct the study, first using census data as was done in West Virginia, and then supplementing the data with a demographic survey if analysis indicates the initial data are inadequate.

There are several other aspects of the pilot study not previously discussed in detail that it is appropriate to include here. They are ancillary results, measured service assumptions, multi-period usage data collection, and finite population problems.

Ancillary Results on Measured Service

Often when a study is conducted, information about variables that were not meant to be part of the study is naturally and easily collected. The case in point in the West Virginia study is local measured service (LMS), which is an optional rate offered in many West Virginia communities. The C & P Telephone Company and the West Virginia Public Service Commission have been pursuing many different LMS options aggressively. There is a natural interest in knowing if there are any emerging patterns among the households that elect LMS. When the forty randomly sampled households in each of the six GSAs are selected, their election or not of LMS will be recorded, and similar information will be available for each of the SLU households. Since both these samples will be matched with demographic data as part of a usage study, it will be possible as a byproduct of the usage study to identify and examine emerging patterns of measured service selection, should they exist.

Measured Service Assumption

While the West Virginia LMS situation will create a beneficial byproduct of the usage study it also creates a cause for concern in that LMS can affect usage, and raises questions about whether or not we should stratify on the basis of customer's using flat rate or LMS. Since optional LMS is a recent development in West Virginia we assumed that people elected LMS because they were already low users on the flat rate. Thus, we decided not to stratify on the basis of rate structure under the theory that the calling habits of LMS customers are not significantly affected by the rates. Had LMS been available for many years or had it been mandatory we could not have made this assumption.

Multi-Period Usage Data Collection

With respect to multi-period usage data collection the reader should be reminded that among the modeling assumptions that allowed us to develop a stratified sampling plan and specify the sample size was the assumption that the study was to be a single-period, cross-sectional study. If the sample is retained and multiperiods of usage data are collected then some reduction in the original sample size would be possible. The problem of determining how much of a reduction is possible so as to obtain the same precision in estimates of βi is intractable. This is due to autocorrelation of usage data within each of the sample households and due to time-related systematic effects on the usage. Inasmuch as several periods of usage data can be collected at nominal additional cost--compared to the cost of setting up measurement on the sample in the first place our conservative strategy was to design the best possible single-period sampling plan and then request multiperiods of usage data on the sample. In this way, study costs are kept under control but study results can be somewhat better than originally anticipated.

The Finite Population Problem

The final issue deals with the question of finite populations. A case in point is that our stratified sampling plan calls for forty households in each of six GSAs, but one of those GSAs contained only thirty-four households, according to the 1980 census figures. All of our calculations of sample sizes implicitly assume that samples are drawn from an infinite population even when the population is a particular stratum. If the sampling is done with replacement, the equivalent of an infinite population is achieved and the assumptions behind our calculations are correct.

Many authors have treated the problem of sampling from finite populations in order to estimate population parameters. In these treatments a prominent suggestion is to adjust the sample size with a finite population factor.[1] However, the situation we have here is not so simple even when the additional problem of errors-in-variables is set aside. Rather than the estimation of population parameters being our goal we want instead to estimate the parameters of the relationship between some set of known population characteristics and the telephone usage by the population. This involves several "finite populations" (strata), each of a different size. To our knowledge there are no finite population factors that have been shown to optimize a stratified sampling plan by jointly adjusting sample sizes to the several differently sized strata.

One approach that seems intuitively appealing is to apply a simple finite population factor to each stratum separately after having optimized the design under an infinite population assumption. The problem with this approach is that we have no way of knowing whether or not this approach results in a correct sample size. Instead we have elected to sample with replacement when the size of a stratum dictates it. This we know is a conservative approach in that it will result in a slightly higher sample size than would be absolutely necessary. In order to simplify the data request to the telephone company, however, we do not mention both sampling with replacement and sampling without replacement. Instead, in the one stratum where its size is clearly too small (thirty-four households) we simply ask for a 100 percent sample, which can be treated in any appropriate manner when analysis is done.

---

[1]For example, see W. Allen Wallis and Harry V. Roberts, Statistics, A New Approach (New York: The Free Press, 1956), pp. 368-75.

It should be noted that randomly selecting forty households with replacement from a population of thirty-four households will reduce the cost of the study since no more than thirty-four households need to be set up for measurement, yet the quality of the statistical results are not substantially affected by this procedure.

APPENDIX A

SUMMARIES OF STUDIES CONTAINING ASPECTS OF TELEPHONE USAGE

Seven articles containing information about the statistical properties of telephone usage and peg count data are summarized below.

(1) Pavarini, Carl. "The Effect of Flat-to-Measured Rate Conversions on Local Telephone Usage." Pricing in Regulated Industries: Theory and Applications. Edited by J. Wenders. Denver: Mountain States Telephone and Telegraph Co., 1979.

Two basic data sources provide analyses of usage: first, 7,253 customers measured over (it would appear) a 3 year period from 1972 to 1975. These residence customers were on a flat rate, from 73 switching offices over a 9 state area and were urban; second, 377 customers in 3 switching offices bordering metropolitan Denver, Colorado studied in 3 periods, April-July, 1970, August-November 1970, and April-July 1971; during the last of these periods (1971) the customers were on the optional "metropac" rate. The first data source provided basic statistics regarding (long term) individual-to-individual distribution of local calls per day and minutes per call. The second data source examined daily total use (averaged over 4 months) before and after the rate change. The statistics in both data sources represent household usage (per unit time, daily or monthly) averaged over either 4 months, or 3 to 4 years. The basic source of variation is, then, differences in households in both data sources, seasonal in data set 2, and to a lesser extent, temporal variation of a given household (which, it is implied, is of lesser consequence since time averaging occurred over 4 months or 3 to 4 years).

In what follows, the joint distribution of (M, L) is studied, where

M = random variable, calls per day (aver ged over a long time period)

m_i = value of M, customer "i"

L = random variable, minutes per message (time average)

l_i = value of L, customer "i"

P = sample space or population, over customers (time average)

$E(L)$, $E(M)$ = expected values (over customers)

$s(L)$, $S(M)$ = standard deviation of L,M

L_k, M_k = conditional E(L), E(M) given switching office k

Relationships obtained were

1. (M,L) roughly independent (if 3 to 5 percent of customers are eliminated having very low and very high values of m_i and l_i, respectively). The statements below are conditional on independence.
2. $M^{.27}$ approximately normally distributed with $s(M)/E(M)=.8$, s(M) independent of E(M).
3. L approximately lognormal, independent of E(L); $s(L)/E(L)=.51$, independent of E(L).
4. M_k normally distributed (across offices)

   L_k lognormally distributed

   $E(M_k) = 3.85$, $s(M_k) = .88$ calls per day

   $E(L_k) = 4.51$, $s\ (L_k) = .42$ minutes per call

   correlation $(M_k.L_k) = .26$

   $E(L_k) = 3.99 + 1.36E\ (M_k) + e$, $s(e) = .481$

Results: Data set two (377 customers, 3 switching offices)

The statistics studied were

U_1 = average monthly usage (total "metropac" minutes, averaged over 4 months, 8-11/70

u_{1i}= value of U, customer i

U_2 = average monthly use, averaged for 4-7/70

u_{2i}= value of U, customer i

U_3 = average monthly use, 8-11/71 (after rate change)

u_{3i}= value of U_3, customer i

First, the conditional regression of $U_2:u_{1i}$ was studied (natural logarithms were taken first), resulting in an equation with a statement of fairly constant variation of residuals. There was no inclusion of the magnitude of the residual error.

Second, the model: "the change effect is stochastic across customers," was selected (the effect on a customer is probabilistic, not deterministic). This model was

$$(U_3:u_{1i}) = e^x u_{1i}^w$$

x normal, parameters E(x), s(x)

The values of the parameters for the switching offices were

Office	Number sampled	E(x)	s(X)	w
1	48	3.42	.445	.174
2	78	2.82	.250	.311
3	251	2.23	.263	.390

(2) Infosino, W. J. "Relationships Between the Demand for Local Telephone Calls and Household Characteristics." Bell System Technical Journal LIX (July-August 1980).

This paper had two analysis phases, a household analysis (several hundred samples of households) and an aggregate analysis of the data averaged over households, within wire centers (ten samples). The first analysis involved household-related calling rates to household demographics (reported in a mail survey, 40 percent response). The second, aggregate analysis, involved wire center average calling rates as related to wire center aggregate demographics estimated from 1970 census data. Such data involved weighting those characteristics of census known to be partially within the area served by the wire center.

Analysis 1--Household Analysis

Two geographic areas were selected to provide a comparison. These were Metropolitan Los Angeles (10 wire centers, 705 total residential households, 1972-3 study) and Metropolitan Cincinnati (8 wire centers, 293 residential households, 1975-6 study). In both areas, flat rate customers were studied, although the Cincinnati customers had an extended area option, exercised by some.

In addition to relating household calling rates to reported demographics of each household, several "aggregate" values of variables not reported on in the mail survey were included that were available from census data, on an area-wide basis. The linear model specified was

$$(CR)^{.5} = B_o + B_1 X_{1i} \ldots + E_i,$$

where $(CR)^{.5}$ = square root of household calling rate, calls per day, and

X_{1i} = presence or absence of "characteristic" i, a dummy variable taking on values 0 or 1.

Actually, for that data collected, the value of $(CR)^{.5}$ could be adequately described by a + bCR, so the above model was replaced by one involving CR rather than the square root of CR. Excluding the Beverly Hills and Madison wire centers as "nonrepresentative" resulted in a coefficient of determination of $R^2 = .349$ in the California study. Variables included were race, sex of head of household, number of people in household wire center, income of household, and wire center (qualitative). The Cincinnati study resulted in a value of R^2 of .35, almost identical to that of the California study.

Then, the "coefficients" related to each wire center in the model were regressed against census data characteristics for these areas. Excellent fits resulted, with telephone density (telephones per square mile) and (only in the California model), racial characteristics. Finally, a "wire center" model was developed for each area. The California model was the simplest:

$$AVG(CR) = -1.34 + 1.1P + 2.3R + .00204D,$$

where AVG(CR) = wire center average call rate,

P = average number of people per household,

R = fraction black and Spanish,

D = residence main telephone density.

The value of R^2 for this data (eight wire centers, or data points) was equal to .98. The Cincinnati model was more complicated with a value of R^2 = .52. Two reasons for a lower value suggested by the authors are that the range in D was much lower in Ohio, and sample sizes were smaller (twenty-eight households on the average, per wire center).

This study mentions the techniques that can be valuable in studying (in geographic areas of relatively homogeneous demographics) average calling rates or usage, as related to area average demographics. It would seem that, in addition to the usual sex, age, and income characteristics, telephone density per unit area, should be used in any model. Also, 8 to 10 wire centers seemed adequate for their purposes, but 20 to 25 households per center seemed inadequate. One interesting note is that, in sampling 178 customers in the Los Angeles Madison wire center, the resultant customers were nonrepresentative (presumably in that electronic switching customers only were included, and the majority of the Madison center customers were served by the nonelectronic type). Also, the range of D in 117 to 2,917 telephones per mile seemed far less adequate than the values of 1,684 to 6,794, found in the California study. Finally, households were sampled from 1 to 2 months each, with no correction for holidays or season of the year.

Infosino indicated that it would seem reasonable to select eight to ten wire centers, and at least fifty to seventy-five households per wire center. The wire centers should be selected to provide as great a range as possible in values of telephone density, average people per household as well as other characteristics (unless, of course, the study is primarily directed toward specific wire centers). This is to establish relationships between usage and demographics. On the other hand, if an estimate of population usage is required, stratified sampling of wire centers should be employed.

(3) Park, Rolla E.; Mitchell, Bridger M.; Wetzel, Bruce M.; and Alleman, James. Charging for Local Telephone Calls: How Household Characteristics Affect the Distribution of Calls in the GTE Illinois Experiment. Palo Alto, Ca.: The Rand Corporation, R-2535-NSF, 1981.

The GTE Experimental data were analyzed, relating household monthly calls (in 3 successive months) to demographics, and to institution of measured rate (3 successive months a year later). A total of 641 households were monitored, representing a random sample of 3 exchanges, supplemented by an overrepresentation of the lower and upper 10 percent users (in calls per month, established by a 10-month monitoring of all customers). Telephone interviews completed the demographic data on each household, with an approximate 12 percent loss for various reasons.

An elaborate model was constructed to account for (1) persistency of given households over time, (2) demographics, (3) effect of change in rate structure, and (4) effect of demographics in rate structure effect. The dependent variable was the number calls per month raised to the 0.27 power.

Results were

1. Demographics accounted for 30 percent of the variation; of significance were size of household, age of head, and number of teenagers (note possible correlation among these)

2. Correlation, month to month, same household, was .23

3. Coefficient of variation, within exchange, was .8 (authors note this agrees with Pavarini's study)

4. Eleven percent of the "before-after" rate change effect was due to demographics

5. Lorenz curves are used to describe variation across households in usage; these display cumulative fraction of calls vs cumulative fraction of households making calls

(4) Perl, Lewis J. Economic and Demographic Determinants of Residential Demand for Basis Telephone Service. Boston: National Economic Research Associates Inc., 1978.

The study focused on economic and demographic factors that influence the demand for basic telephone service by residential households. A total of 36,703 households, which were obtained from the public use sample of the 1970 census of population, were monitored. Using the

U.S. Federal Census data, enough information could be obtained about the effect of telephone prices on demand for telephone service; thus, Revenue Accounting Offices (RAOS), which are a collection of continuous Bell exchanges grouped together for accounting purposes, were used. Then, the monthly service charge was estimated by combining the data from AT&T's Market Research Characteristics and System (MRTS) with the RAOS' data.

For the purpose of estimating the relationship between demand for basic telephone service households and various socioeconomic factors, three alternative models were used. They were the linear, the logit, and the probit models. Each of these three models have broadly similar implications regarding the impact of specific variables on the demand for telephone service. However, the effects of each factor of the three models on the demand for telephone service may have quite a different magnitude of results. In addition, each of the three models has numerous advantages and disadvantages. Therefore, it is hard to evaluate the linear, logit, and probit functions. In the paper, the author went to great lengths to describe the statistical properties and interpreted the results as the statistical properties were applied to the analysis of the demand for telephone service of the three models.

Using the author's notation, the linear, the logit, and the probit models are shown as follows:

Model 1, Linear Model

$$P(T/X) = b'X$$

where P(T/X) = the probability of telephone service for a household with characteristics described by the vector X

X = A vector of characteristics that describe each household's status with respect to the factors that influence demand for telephone service

b = A vector of coefficients, each of which describes the effect on the probability of telephone service of a unit change in one of the components of the vector X

Model 2, Logit Model

$$P(T/X)/(1-P(T/X)) = EXP(b'X)$$

Model 3, Probit Model

$$P(T/X) = 1/2\pi \int_{-\infty}^{b'x} e^{-\mu^2/2} d\mu$$

The parameters, b, of the linear model were obtained by using the ordinary least-square method. However, the logit and probit models used the maximum likelihood approach to obtain the b. The results of the parameters of the three models are listed in table A-1.

The empirical results of the study are summarized as follows:

1. There is a strong positive correlation for a household having basic telephone service with respect to the income, age, and education

2. There is also an inverse relationship for a household having basic telephone service with respect to the price of telephone service and the number of persons per household

3. The regression coefficients of the seven indicator variables showed that the demand of telephone service is higher for employed household heads and urban areas than for the unemployed household heads and rural areas; also, the negro-headed household, male and female individuals, male-headed household with no spouse present, and Southern region have lower demand for telephone service than other racial groups, household types, and regions

4. In general, the linear, the logit, and the probit models have the similar effects for each of the variables; The logit and probit models appear to perform somewhat better than the linear model

(5) Wellenaus, B. "The Effect of Income and Social Class on Residential Telephone Demand," Telecommunications Journal XXXVI (1969): 227-230.

In the article, for the purpose of examining the effect of income and social class (defined by the occupational level) on residential telephone demand, data that had been collected periodically in the metropolitan area of Santiago by the University of Chile in 1967 are analyzed. A stratified random sample out of a total 2,546,900 inhabitants as monitored. Each strata had about 3,200 households.

TABLE A-1

REGRESSION COEFFICIENTS RELATING THE PERCENTAGE OF HOUSEHOLDS WITH BASIC TELEPHONE SERVICE TO SELECT ECONOMIC AND DEMOGRAPHIC VARIABLES

Unit of Variable	Variables	*Linear Model Regression Coefficient	Logit Model Regression Coefficient	Probit Model Regression Coefficient
	Intercept income adjusted	65.2865	-.9101	-.4463
Dollars	Monthly service charge	-4.2836	-.1325	-.0701
Dollars	Installation charge	-.1946	-.0222	-.0136
Indicator	Measured rate service	-1.9105	-.2425	-.1271
Years	Age of head	.3917	.0350	.0200
$100	Income	.1353	.0154	.0080
$100-years	Age*income	-.00143	-.0006	-.0005
Number	Persons per household	-.5419	-.0486	-.0223
Years	Education	1.1191	.1164	.0648
Indicator	Urban	3.7448	.3551	.1958
Indicator	Negro	-9.8747	-.5953	-.3577
Indicator	Source	-3.7241	-.3480	-.1906
Indicator	Employment	-6.8609	.2337	.1679
Indicator	Other male head	-17.1626	-.6235	-.3489
Indicator	Male individual	3.5044	-1.290	-.7468
Indicator	Female individual	-1.8054	-.1908	-.1382

*The relationships between the demand for telephone service and some factors are nonlinear in the linear model. In order to explain the nonlinearities, several additional variables were included. These additional variables are listed below.

Variable	Regression Coefficients
Age squared	-.002625
Income squared	-.000167
Income times monthly charge	.0088
Age times monthly charge	.0488

Source: Tables 2, and 3 of Perl (1978)

The telephone demand conducted as response variable was defined as "the number of households with a telephone set divided by the total number of households."

The author uses a Pearson chi-squared, test-of-independence-test to analyze the influence of two social-economic variables of the households on the telephone demand. According to the test, the main results are

1. The income factor significantly influences telephone demand
2. The social class has a significant effect on local residential demand
3. There is high correlation between income and social class in Santiago
4. The low income groups and social classes both have a significant effect on residential telephone demand; however, for the high income and social classes groups, the second variable is not significant if the other variable has already been included in the model

(6) Garfinkel, L., and Linhart, P. B. "The Revenue Analysis of Local Measured Telephone Service." Public Utilities Forthnightly, October 9, 1980, pp. 15-21.

In this article a method is developed that is derived from the study by Carl Pavarini (see above) to analyze the relationships between the aggregate revenue and local measured telephone service.

In the method, the authors discuss how the four elements--number, duration, distance, and time of day--determine the customers' monthly bills under a local measured service tariff. Although there is no measuring equipment on individual lines under flat rates, the paper shows that the distribution of telephone usage under flat rates shows that 50 percent of users make only 20 percent of calls. This is a skewed distribution that implies that the individual calling rates are wildly dispersed. See figure A-1.

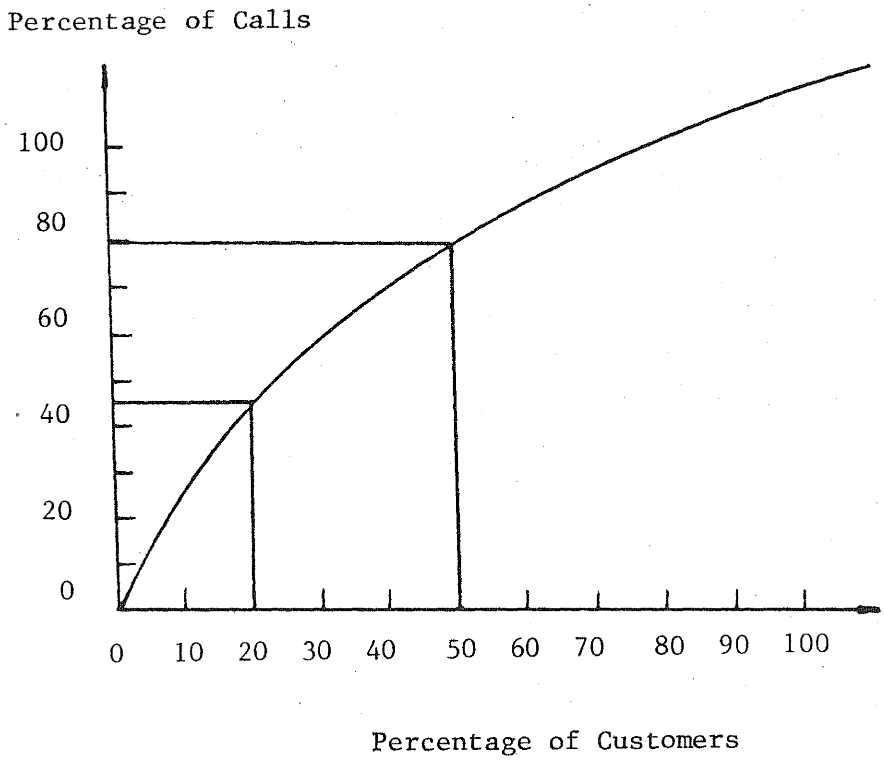

Fig. A-1 Flat rate residential percentage of calls vs. percentage of customers

(7) Park R. E., and Wetzel, B. M. Charging for Local Telephone Calls: Pricing Elasticity Estimates from the GTE Illinois Experiment. Palo Alto, Ca.: The Rand Corporation, R-2535-NSF, 1981.

The purpose of the article was to estimate price elasticities for local telephone calls and minutes of conversation. The experimental data of the article was obtained from the General Telephone and Electronics local measured service experiment in three small cities in central Illinois--Jacksonville, Clinton, and Tuscola. The GTE charged the telephone usage price according to the number of calls and minutes of calling. The experimental measured-service tariffs that are different from other places are nonoptional and include no allowance of free calls.

In order to examine the relationship of telephone use with respect to the experimental prices and other factors, the authors developed a model including shift, dummy, and trend variables. The shift variable is used to reflect the changes that were caused by the introduction of usage prices. The dummy and trend variables are used to control for cross-sectional differences and smooth changes in use over time. Two different sets of factors which determine the single-party and multiparty use were discussed in the article, and the authors join the single-party and multiparty parts of the equation to form a single estimating equation.

The results of the article are

1. Telephone use appears in a seasonal pattern--it is higher in the winter than in the summer; the lowest telephone use appears in September

2. Average telephone use by single-party customers was higher than the average use by multiparty customers before the measured-service tariff was offered

3. After introduction of the measured service rate in September 1977, single-party use decreased abruptly and multiparty use began to increase

4. Usage prices of local measured-service tariffs have a small but significant effect on local telephone use with price elasticities ranging up to about 0.1 in absolute value

APPENDIX B

THE SIMULATION MODEL AND RESULTS

The purpose of this appendix is to present the simulation model for telephone usage and examine the cost-optimal tradeoff between different design parameters and their associated study costs.

The remainder of this appendix is organized into five sections. In the first section, the purpose of the simulation model is briefly discussed. In the second section, the simulation model and its related parameters are presented. In developing the simulation model we analyzed the statistical properties of three demographic characteristics (income, age, children). They are also shown in this section. In the third section, the function of the simulation main program and subprograms are presented. Different models are associated with the different sampling plans. The simulation results for certain sampling plans that were investigated in the West Virginia pilot study are presented in the fourth section.

Individuals interested in receiving a listing of the FORTRAN source code for the simulation programs together with the input data used in the West Virginia pilot study should request them by writing:

Publications Office
The National Regulatory Research Institute
2130 Neil Avenue
Columbus, Ohio 43210

The Purpose of the Simulation Model

In this study, the Telephone Usage Simulation System (TUSS) is used to develop general statistical sampling models that would be used to examine the effects of various levels of aggregation of sampling unit (e.g., households, central office, exchange, and the effects of temporal aggregation).

Originally, the word "simulate" meant to "imitate or feign." This meaning suggests one important characteristic of simulation: to simulate is to imitate the operations of various kinds of real-world facilities or processes. The facility or process of interest is usually called a system, and in order to study it scientifically we often have to make a set of assumptions about how it works. These assumptions, which usually take the form of mathematical or logical relationships, constitute a model that is used to gain some understanding of how the corresponding system behaves.

If the relationships that comprise the model are simple enough, it may be possible to use mathematical methods (such as algebra or calculus) to obtain exact information on questions of interest; that is called an analytic solution. However, most real-world systems are too complex to allow realistic models to be evaluated analytically, and these models must be studied by means of simulation. In a simulation we use a computer to evaluate a model numerically over a time period of interest, and data are gathered to estimate the desired true characteristics of the model.

Since physical models are often relatively expensive to build and unwieldly to move, mathematical models are often preferred. In a mathematical model, mathematical symbols or equations are used (instead of physical objects) to represent the relationships in the system. Therefore, in this appendix we restrict our attention to a particular type of mathematical model of a system that we call a simulation model.

As was discussed in the main body of this report, much of the work in designing a study of telephone usage patterns involves the proper stratification of the population by demographic characteristics, geographic characteristics and customer class, and the determination of a scientific sampling plan. To do so would require development of a method to satisfy these two requirements and to test the methods by working with the state commission to produce a data base that is adequate for estimating the parameters of a functional relationship between local telephone usage and the characteristics of subscribers. Thus, we first decide which relationships to include in our simulation model, then we use a computer that simulates or mimics the system under study in order to determine and keep track of the implications.

However, selection of methods of monitoring usage will entail a consideration of the tradeoff between stratified sample and a random sample, and between obtaining household demographic data from census records and conducting a survey. Resolution of these questions also requires a reasonable model of random household usage over time, including parameters associated with the demographic characteristics of the household as well as variation over time. Such a model will be contructed to insure it is as consistent as possible with existing studies. Since the model will deal with household usage, it can be manipulated to generate anticipated effects of aggregation of households in data reporting. Presumably, given the information requirements of the study, and limitations on allowable data collection costs, the model would then be used to examine those feasible alternative sampling plans that satisfy the information requirement.

We conclude this section by giving several reasons why we use the simulation method (SMTU) in our study. First, an advantage of this approach is that most complex, real-world systems with stochastic elements cannot be accurately described by a mathematical model that can be evaluated analytically. Thus, a simulation is often the only type of investigation possible. Second, simulation allows one to estimate the performance of a sampling plan under some projected set of conditions. Third, alternative proposed samples designs can be compared via simulation to see which best meets the specified requirement. Finally, in a simulation we can maintain much better control over experimental conditions than would generally be possible when experimenting with the system itself.

The sections of this appendix that follow contain a detailed description of TUSS and a discussion of the sampling plan that was used to optimize the tradeoff between sampling cost and data quality.

The Simulation Model: An Overview

The simulation model was developed to analyze the impact of various sampling plans on the capability of estimating demographic effects on usage. By performing the simulation studies, the tradeoff between cost and quality of the various sampling plans can be examined. We begin this section with an overview of the structure of the simulation model, and then discuss the statistical properties of the independent variables. Finally, the design parameters of the simulation model are presented.

Usage Model Structure

As we have mentioned before, this is a study of the means by which one can efficiently and effectively study patterns of telephone usage. The main goal is to develop a method that reduces the cost of studying the effects of household characteristics on telephone usage.

Empirical studies of this type are often described as either cross-sectional studies or longtitudinal studies. The cross-sectional studies could involve gathering data across many individual sampling units for a short period of time, the longitudinal studies would gather data across perhaps fewer individuals, but would continue to collect periodic samples over a long period of time. While subscriber characteristics that determine usage can be discovered and related to usage with a cross-sectional study, the longitudinal studies are better at detecting trends, seasonal variations, and responses to price changes or regulatory policy. Longitudinal data are, on the other hand, more difficult to analyze because of the autocorrelation in the data. The actual elapsed time required to collect sufficient data for an analysis is greater in longitudinal studies than in cross-sectional studies for obvious reasons. Furthermore, a well-designed cross-sectional study can be easily fine-tuned and repeated periodically in order to produce data for a longitudinal study. Thus, for these several reasons we have concentrated our efforts on producing a cross-sectional study design in the pilot state (West Virginia).

A one-time period cross-sectional model is given as follows:

$$(\text{calls})^{.27} = M + \beta_1(\text{Income}) + \beta_2(\text{Age}) + \beta_3(\text{HHSIZE}) + h + e$$

where

$(\text{calls})^{.27}$ = Transformed usage of household during one time period,

M = intercept term,

h = household-specific error that is normally distributed with $E(h) = 0$, $V(h) = \sigma^2$,

e = within-household error that is normally distributed with $E(e)=0$, $v(e)=\sigma_e^2$,

β_j = regression coefficients, $j=1,2,3$.

The independent variables used in this linear equation are listed in table B-1.

TABLE B-1

THE SET OF DEMOGRAPHIC VARIABLES FOR EXPLAINING HOUSEHOLD CALLING RATES

Variable	Definition
Income	= Logarithm of household income ($000)
Age	= Age of head of household in scores of years
HHSIZE	= Logarithm of number of persons in the household

Source: Authors' adaptation of the Park et al. study

Design Parameters (NonSurvey Case)

Since the independent variables cannot be observed without conducting a survey, census summary data are substituted for actual household data. In the case of West Virginia, the demographic characteristics data that were easily obtained from the U.S. Federal Census statistics are

X_1 = median income of the area,

X_2 = percentage of the households in the area with the age of household head at least 65 years,

X_3 = percentage of the households in the area with at least one child, age 0-17 years.

The regression coefficients for the variables given in table B.1 are shown in table B-2.

TABLE B-2

THE MAXIMUM LIKELIHOOD ESTIMATES OF THE RATE OF CHANGE IN USAGE DUE TO THE THREE FACTORS

Intercept	β_1 (Income)	β_2 (Age)	β_3 (HHSIZE)
2.1	-.004	.068	1.1

Source: Park et al. (1981)

The effect of household size transformed to an effect of the number of children that is measured by a combination of household size and teenagers effects from the Park et al. study. The parameter we used here is 1.1. The estimate effects are shown in table B-3.

TABLE B-3

ESTIMATE EFFECTS CALCULATION

Number of children	Calculation	Effect
1	1.1 x ln (1+1.8)	1.13
2	1.1 x ln (2+1.8)	1.47
3	1.1 x ln (3+1.8)	1.72

Source: Authors' calculation

There are two error terms included in the simulation model. We assume that both of the error terms are normally and independently distributed. The variances for these error terms are σ_ε^2 (within-household error) = .07 and σ^2 (household-specific error) = .307.

Since the simulation model must select random households according to the restrictions of a particular sampling plan we analyzed the statistical properties of the three demographic characteristic variables in the Kanawha County area of West Virginia. For example, given that a household is randomly selected from a particular GSA, we needed sufficient knowledge about the distribution of the three demographic characteristics in the GSA in order to specify the three values for the household in question.

For illustrative purposes we detail the results of this analysis of the four GSAs used in the simulation of a stratified sampling plan. The assumptions borne out during the analysis are that the income and the age of the head of household are both approximately log normally distributed (natural logarithms) and correlated, while the number of children in a household is related to age, but appears uncorrelated with income other than through age. We estimated a correlation coefficient of .13 between income and age. Table B-4 gives the available census information concerning income and age for the four GSAs in the stratified sampling plan. Table B-5 gives the distribution parameter corresponding to the census data in table B-4.

Tables B-6, B-7, B-8, and B-9 show the cumulative probability of the number of children for given age groups for the four selected.

TABLE B-4

THE DEMOGRAPHIC CHARACTERISTICS STATISTICS OF THE FOUR SELECTED GSAS

GSA	Median Income (1,000's)	Percentage of Households with Head Older than 64	Percentage of Households with Children
1	4.4	30	46
2	6.9	0	26
3	24.4	40	31
4	29.4	0	77

Source: U.S. Census data

TABLE B-5

THE STATISTICS FOR THE FOUR SELECTED GSAS*

GSA	Mean of Log Normal for Income	STD of Log Normal for Income	Mean of Log Normal for Age of Household	STD of Log Normal for Age of Household
1	1.48	.76	4.03	.25
2	1.93	.76	3.66	.25
3	3.19	.76	4.12	.15
4	3.38	.76	3.66	.25

*Note: All logarithms are natural logarithms
Source: Authors' calculation from U.S. Census data

TABLE B-6

THE CUMULATIVE PROBABILITY OF NUMBER OF CHILDREN ASSUMED FOR GIVEN HOUSEHOLDER'S AGE RANGES FOR GSA 1

	The Cumulative Probability for				
Householder	None	1	2	3	4
10 to 24 years	.413	.793	.959	.992	1
25 to 34 years	.209	.481	.815	.953	1
35 to 44 years	.099	.292	.654	.881	1
45 to 54 years	.383	.702	.891	.963	1
55 to 64 years	.686	.910	.975	.993	1
65 to 74 years	.781	.942	.977	.990	1
75 years and over	.796	.957	.980	.992	1

Source: Authors' calculation based on U.S. Census data

TABLE B-7

THE CUMULATIVE PROBABILITY OF NUMBER OF CHILDREN ASSUMED FOR GIVEN HOUSEHOLDER'S AGE RANGES FOR GSA 2

	The Cumulative Probability for				
Householder	None	1	2	3	4
10 to 24 years	1	1	1	1	1
25 to 34 years	.606	.741	.908	.977	1
35 to 44 years	.286	.439	.726	.906	1
45 to 54 years	1	1	1	1	1
55 to 64 years	1	1	1	1	1
65 to 74 years	1	1	1	1	1
75 years and over	1	1	1	1	1

Source: Authors' calculation based on U.S. Census data

TABLE B-8

THE CUMULATIVE PROBABILITY OF NUMBER OF CHILDREN ASSUMED FOR GIVEN HOUSEHOLDER'S AGE RANGES FOR GSA 3

	The Cumulative Probability for				
Householder	None	1	2	3	4
10 to 24 years	.428	.798	.960	.992	1
25 to 34 years	.217	.486	.817	.954	1
35 to 44 years	.102	.295	.655	.882	1
45 to 54 years	.397	.708	.893	.964	1
55 to 64 years	.712	.917	.977	.993	1
65 to 74 years	.811	.950	.980	.992	1
75 years and over	.826	.963	.983	.993	1

Source: Authors' calculation based on U.S. Census data

TABLE B-9

THE CUMULATIVE PROBABILITY OF NUMBER OF CHILDREN ASSUMED FOR GIVEN HOUSEHOLDER'S AGE RANGES FOR GSA 4

	The Cumulative Probability for				
Householder	None	1	2	3	4
10 to 24 years	.372	.778	.956	.991	1
25 to 34 years	.188	.467	.810	.952	1
35 to 44 years	.089	.284	.650	.880	1
45 to 54 years	.345	.683	.884	.961	1
55 to 64 years	.618	.890	.969	.991	1
65 to 74 years	.704	.922	.969	.987	1
75 years and over	.717	.940	.873	.989	1

Source: Authors' calculation based on U.S. Census data

GSAs. These probabilities are approximate in that they were rounded to one to four children, giving no possibility of more than four children.

Under the random sampling plan the GSAs would be selected randomly and could be located in any one of the 235 GSAs. Therefore, statistics similar to those shown in table B-5 through B-9 were needed for all 235 GSAs in Kanawha county. The statistics of the demographic characteristics for all the GSAs are not shown here.

The simulation process uses as input the model and parameters as described earlier in this section and it uses assumed distribution and parameters such as those given in the preceding several tables. It then generates a set of households and their telephone usage according to particular sampling plan. The output of the simulation is a set of hypothetical household data. These data consist of the following information about each household in the sample:

- Telephone usage for one period
- The GSA in which the household is located
- The income of the household
- The age of the head of household
- The number of persons (including children) living in the household

These data are intended to be representative of the data one would get if a sample of real households was selected according to the sampling plan, their usage was measured, and a survey was conducted to determine their demographic characteristics.

The next section discusses the computer simulation program in greater detail.

Computer Program

The simulation program was written in Fortran-77, and all computation was done on a Digital Equipment PDP-44 system.

Briefly the entire program consists of the following three major steps.

1. Generate N random vectors of household demographic characteristics according to a sampling plan that associates each vector with a particular GSA. In the West Virginia case two of the components of the vector were bivariate normal and the third component was a conditional multinomial with age being the condition variable. The set of all vectors is denoted X with each vector denoted X_i.

2. Calculate the usage $U_{i(j)}$ using the model $U_i = M + X_i\beta + e$ where the index i denotes household i.

3. Store U_i and X_i on a disk file that can be read by other programs to analyze the data. Also associate a GSA number with each household in the file.

The simulation program requires the use of techniques involving the generation of random vectors from a bivariate normal population with a specified variance-covariance matrix and random vectors from standard normal population with a specified mean and variance.

The S1M3 subroutine generates η-variate normal deviates, $V^1 = (V_1,\ldots,V_\eta)$ with mean μ and covariance Q based upon the following formula:

$V = CZ + \mu$, where

Z = Independent standard normal variables (that is, zero mean and unit variance)

C = a unique lower triangular matrix such that $Q = CC'$

The CMAT subroutine generates C matrix, such that the elements of C are determined recusively as follows. This method has been discussed by Moonan and Wold.[1]

$$C_{i1} = \sigma_{i1} / \sqrt{\sigma_{11}} \qquad 1 \leq i \leq n$$

$$C_{ii} = \sqrt{\sigma_{ii} - \sum_{k=1}^{i-1} C_{ik}^2} \qquad 1 \leq i \leq n$$

---

[1]William J. Moonan, "Linear Transformation To a Set of Stocastically Dependent Normal Variables," Journal of American Statistical Assocation, 52 (1957): 247-252; Herman Wold, Random Normal Deviates, no. XXV: Tracts for Computers (place of publication: Cambridge University Press, 1955).

$$C_{ij} = \{\sigma_{ij} - \sum_{k=1}^{j-1} C_{ik}C_{jk}\} / c_{jj} \quad 1 < j < i \leq n$$

$$C_{ii} = 0 \quad i < j \leq n$$

The method for generation of a set of independent standard normal variables are based on the Box and Muller[2] method. The functions of the main program are

1. read in input data
2. generate X_k, h_{ij}, e_{ijk} and calculate U_{ijk} for each household
3. report the result

The functions of the subprograms are

S1M3: generation of random vectors from a multivariate normal population

CMAT: calculation of lower triangular matrix c, such that $CC^T = \sum$

NORMAL: generation of random vectors from a stand normal population DRAND: generation of single random numbers CHILD: generation of the number of children for each household GA: generation of the cumulative probability for the number of children at selected age of householder

The Simulation Runs and Their Results

Two types of sampling plans were simulated. The first was a completely random sampling plan where the probability of a household being selected from any particular GSA was proportional to the number of households in the GSA. This plan was intended to represent a SLU sample of households.

The second plan was a stratified sampling plan that selected households from the four preselected GSAs mentioned in previous

---

[2]G. E. P. Box and Mervin E. Muller, "A Note on the Generation of Random Normal Deviates," Annual Mathematic Statistics 29 (1958): 610-611.

sections. In this case the sample was split evenly among the four GSAs but the households themselves were randomly selected from their designated GSAs. The four GSAs had been selected to achieve a near-optimal experimental plan according to the determinant and trace criteria presented in chapter 2.

Once the simulation program was run, two separate data sets were available for analysis that could take place under different scenarios. The subsections that follow present those analyses.

Simulation Results

Statistical techniques were used to analyze the output data from the simulation runs. Typical goals are to produce point estimates and confidence intervals for model parameters. In chapter 3 we mentioned that there were two key questions in our study. One is whether to use the SLU sample or to obtain a new sample based on an experimental design. The other decision is whether or not to conduct a survey of the sample. While there are differences in the cost of a study depending on the decisions made in the two key instances, there are also differences in the statistical quality of a study design resulting from these two key decisions.

The simulation runs provide data for the two cases of SLU and the stratified sample, but in each of these situations one can either survey the sampled households to obtain the demographic information or substitute the known census data for the GSA in which the households reside. The implications of this second question can be answered through analysis of the simulation data and do not require additional simulation runs.

In chapter 3 of the report we defined the four cases resulting from the two questions "SLU or not?" and "survey or not?". They are as follows:

Case 1: A stratified sample is planned but census data is used in place of actual household demographic data.

Case 2: A stratified sample is planned and a survey is conducted to obtain actual demographic data on a household basis.

Case 3: A random sample is used and a survey is conducted to obtain actual demographic data on a household basis.

Case 4: A random sample is used but census data is used in place of actual household demographic data.

Based on the four different cases, four different analytic models were used. These are shown in the table B-10.

TABLE B-10

LINEAR MODELS OF SELECTED DESIGNS

Analysis Cases	Simulation Cases: SLU Sample	Simulation Cases: Stratified Sample
No Survey	Case 4 Model: $U = M+(X_2-D)\beta+(e_2+D\beta)$ or $U = M+ R_2\beta + f_2$ $cov(b) = \sigma^2_{f_2}(R_2'R_2)^{-1}$	Case 1 Model: $U = M + (X_1-D)\ \beta + (e_1+D\beta)$ or $U = M + R_1\beta + f_1$ $cov\ (b) = \sigma^2_{f_1}(R_1'R_1)^{-1}$
Survey	Case 3 Model: $U = M + X_2\beta + e_2$ $cov(b)=\sigma^2_{e_2}(X_2'X_2)^{-1}$	Case 4 Model: $U = M+ X_1\beta + e_1$ $cov(b)=\sigma^2_{e_1}(X_1'X_1)^{-1}$

Source: Authors' models

The notation in table B-10 is as follows:

U = (nxl) vector of household usage, during one period of time

M = (nxl) vector of intercept terms

X_1 = (nxs) matrix of demographic characteristics associated with stratified sample

X_2 = (nxs) matrix of demographic characteristics associated with SLU sample

e_j = (nxl) vector of random error term.(j=1,2)

D = (nxs) vector of demographic characteristic differences of household and the recorded census values

R_j = (nxs) vector of census records of demographic characteristics. (j=1,2)

f_j = (nxl) vector of random error term include D.(j=1,2)

β = (sxl) vector of regression coefficients

b = (sxl) vector of estimate of β

n = number of households in the sample

s = number of demographic characteristics for each household

Quantative Demographic Variable Model

One analysis assumed each demographic variable could be measured quantitatively so that the model variables are defined as given in table B-11.

TABLE B-11

QUANTITATIVE DEMOGRAPHIC VARIABLES

Survey	No Survey
Log income for household	Log median income for GSA
Age of head of household in scores	Average age of head of households in the GSA in scores
Size of household	Average size of households in GSA

Source: Authors' choice

Indicator Demographic Variable Model

One reason for the use of an indicator variable model is that one can not necessarily obtain the average age of householder and average household size from the U.S. Census summary statistics directly. Generally, the data we obtained from the census on these two demographic variables were recorded in percentage of the GSA population. Thus to explore the situation where such averages may not be available we analyzed the two data sets using a model based on qualitative (or indicator) variables.

Two of the three variables were transformed to indicator variables. In both cases the quantitative demographic variable was coded into a two-level indicator variable. Let t_{ij} be the two-level indicator variable and define the transformation for each variable as follows:

$$t_2 = \begin{cases} 1, & \text{if age of head of household} > 64 \text{ years} \\ 0 & \text{otherwise} \end{cases}$$

$$t_3 = \begin{cases} 1, & \text{if number of children} > 0 \\ 0 & \text{otherwise} \end{cases}$$

The income variable remains continuous. In the case where a survey of the sample household is conducted, each household is coded with income, and two zeros or ones depending whether the head is over 64 years old or whether there are children residing in the household. In the case where a survey is not conducted, each household is coded with the median income of its GSA, and two fractional values giving the fraction of households in the GSA that have heads over 64 years old and the fraction of households having children. The advantage of this coding scheme is that a linear model relating income, age, and children to telephone usage in each of these cases will possess precisely the same parameters, that will have the same values. In this case, the two models will be equivalent but, of course, the quality of the estimates of the parameters will differ in the two cases. As before, the choice of a random sample or a stratified sample also affects the quality of the estimates. The results of all four combinations of cases are shown in table B-12.

Based on the models in table B-10 and the model variables in table B-11 the two simulated data sets were analyzed. The results important to this study are given in table B-13.

TABLE B-12

NORMALIZED VARIANCE OF THE ESTIMATION OF THE COEFFICIENTS OF THE VARIABLE LOG INCOME, AGE OF HEAD, AND SIZE OF HOUSEHOLD IN TWO LEVEL INDICATOR VARIABLE MODEL

	Random Sample	Stratified Sample
	Case 4	Case 1
No survey	$\sigma^2_{\beta_1}$ = 8.28	$\sigma^2_{\beta_1}$ = 1.08
	$\sigma^2_{\beta_2}$ = 136.11	$\sigma^2_{\beta_2}$ = 20.84
	$\sigma^2_{\beta_3}$ = 75.52	$\sigma^2_{\beta_3}$ = 20.76
	Case 3	Case 4
Survey	$\sigma^2_{\beta_1}$ = 649	$\sigma^2_{\beta_1}$ = .34
	$\sigma^2_{\beta_2}$ = 4.321	$\sigma^2_{\beta_2}$ = 3.52
	$\sigma^2_{\beta_3}$ = 2.037	$\sigma^2_{\beta_3}$ = 1.87

Note: $\sigma^2_{\beta_2}$ is the estimate of the variance of the estimate of the linear model coefficient βi

Source: Authors' analysis of simulated data based on model structures given in table B-10 and the demographic variables given in table B-11, but with the independent variables transformed to indicator variables

TABLE B-13

NORMALIZED VARIANCES OF THE ESTIMATION OF THE COEFFICIENTS OF THE VARIABLE LOG INCOME, AGE OF HEAD, AND SIZE OF MODEL HOUSEHOLD IN QUANTITATIVE DEMOGRAPHIC VARIABLE MODEL

	Random Sample	Stratified Sample
	Case 4	Case 1
No survey	$\sigma^2_{\beta_1} = 8.56$	$\sigma^2_{\beta_1} = 1.11$
	$\sigma^2_{\beta_2} = 40.19$	$\sigma^2_{\beta_2} = 3.61$
	$\sigma^2_{\beta_3} = 30.3$	$\sigma^2_{\beta_3} = 6.76$
	Case 3	Case 4
Survey	$\sigma^2_{\beta_1} = .58$	$\sigma^2_{\beta_1} = .31$
	$\sigma^2_{\beta_2} = 1.09$	$\sigma^2_{\beta_2} = .85$
	$\sigma^2_{\beta_3} = 3.11$	$\sigma^2_{\beta_3} = 3.0$

Note: $\sigma^2_{\beta_i}$ is the estimate of the variance of the estimate of the linear model coefficient βi

Source: Authors' analysis of simulated data based on the models given in table B-10, and the demographic variables given in table B-11

APPENDIX C

ADDRESS LISTINGS OF STATE COORDINATING ORGANIZATIONS

This appendix contains an address list for state coordinating organizations that are part of the U.S. Bureau of Census' State Data Center Program. The list was reproduced from a list dated January 1984 that was obtained from the U.S. Bureau of the Census through the Community Development Division of the Governor's Office of Economic and Community Development in West Virginia. Persons working with census data to plan a usage study are urged to contact their respective state's coordinating organization for assistance in obtaining, understanding, and analyzing census data. Also contained in this appendix is a telephone contact list for census data users. The list is also a reproduction of a Bureau of the Census publication provided by the West Virginia office mentioned above.

<u>ALABAMA</u>

Alabama State Data Center
Center for Business and
Economic Research
University of Alabama
P.O. Box AK
University, AL 35486
Dr. Carl Ferguson, Director
*Mr. Edward Rutledge
(205) 348-6191

Office of State Planning
and Federal Programs
State Data Center
P.O. Box 2939
3465 Norman Bridge Rd.
Montgomery, AL 36105-0939

Alabama Public Library Service
6030 Monticello Drive
Montgomery, AL 36130
Mr. Anthony Miele
(205) 277-7330

<u>ALASKA</u>

Alaska Department of Labor
P.O. Box 1149
Juneau, AK 99802
*Ms. Barbara Baker
(907) 465-4513

*Denotes key contact person

Office of the Governor
Office of Budget and
Management
Division of Strategic Planning
Pouch AD
Juneau, AK 99811
Mr. Thomas Chester
(907) 465-2203

Department of Education
Division of Libraries and
Museums
Alaska State Library
Pouch G
Juneau, AK 99811
Mr. Lou Coatney
(907) 465-2942

Department of Community and
Regional Affairs
Division of Local Government
Assistance
Pouch BH
Juneau, AK 99811
Mr. Doug Griffin
(907) 465-4734

Institute for Social, Economic,
and Government Research
University of Alaska
707 "A" Street, Suite 206
Anchorage, AK 99501
Mr. Jack Kruse
(907) 278-4621

ARIZONA

The Arizona Department of
Economic Security
1300 West Washington, 1st Floor
P.O. Box 6123-045Z
Phoenix, AZ 85005
*Ms. Linda Strock
(602) 255-5984

Research Specialist
College of Business Admin.
Arizona State University
Tempe, AZ 85287
Mr. Tom Rex
(602) 965-3961

College of Business Admin.
Northern Arizona University
Box 15066
Flagstaff, AZ 86011
Dr. Ron Gunderson
(602) 523-2358

Federal Documents Section
Department of Library, Archives,
and Public Records
Capitol, Third Floor
17000 West Washington
Phoenix, AZ 85007
Atifa Rawan
(602) 255-4121

Dean of the Graduate College
Administration Building, Rm. 501
University of Arizona
Tucson, AZ 85721
Dr. Lee B. Jones
(602) 626-4031

ARKANSAS

IREC-College of Business Admin.
University of Arkansas
33rd and University Avenue
Little Rock, AR 72204
Dr. Barton Westerlund, Director
Sarah Breshears
*Dr. Forrest Pollard
(501) 371-1971

Arkansas State Library
1 Capitol Mall
Little Rock, AR 72201
Ms. Frances Nix
(501) 371-2159

*Denotes key contact person

CALIFORNIA

State Census Data Center
Department of Finance
1025 P Street
Sacramento, CA 95814
Ms. Linda Gage
*Mr. Bill Schooling, Director
(916) 322-4651

Saramento Area COG
800 H Street
Suite 300
Sacramento, CA 95814
Mr. Bob Faseler
(916) 441-5930

Assn. of Bay Area Governments
Hotel Claremont
Berkeley, CA 95705
Ms. Patricia Perry
(415) 841-9730

Regional Research Institute
of Southern California
600 S. Commonwealth Street
Los Angeles, CA 90005
Mr. Tim Douglas
(213) 385-1000

Source Point
Security Plaza Pacific
1200 3rd Avenue
San Diego, CA 92101
Ms. Karen Lamphere
(714) 236-5353

State Data Center Program
University of Calif.-Berkeley
2538 Channing Way
Berkeley, CA 94720
Ms. Ilona Einowski
(415) 642-6571

COLORADO

Division of Local Government
Colorado Dept. of Local Affairs
1313 Sherman Street, Rm. 520
Denver, CO 80203
*Mr. Reid Reynolds
Ms. Rebecca Picaso
(303) 866-2351

Business Research Division
Graduate School of Bus. Admin.
University of Colorado-Boulder
Boulder, CO 80309
Mr. Gerald Allen
(303) 492-8229

County Information Service
Department of Economics
Colorado State University
Fort Collins, CO 80523
Ms. Sue Anderson
(303) 491-5706

Documents Department
The Libraries
Colorado State University
Fort Collins, CO 80523
Ms. Karen Fachan
(303) 491-5911

CONNECTICUT

Comprehensive Planning Division
Office of Policy and Management
State of Connecticut
80 Washington Street
Hartford, CT 06106
*Mr. Theron A. Schnure
(203) 566-3905

DELAWARE

Delaware Development Office
99 Kings Highway
P.O. Box 1401
Dover, DE 19903
Mr. Nathan Hayward, Acting Dir.
*Mr. Doug Clendaniel
(302) 736-4271

*Denotes key contact person

Computing Center
University of Delaware
192 S. Chapel Street
Smith Hall
Newark, DE 19711
Mr. Bob Shaffer
(302) 738-8441

DISTRICT OF COLUMBIA

Data Services Division
Mayor's Office of Planning and Development
Room 458, Lansburgh Bldg.
420 7th Street, N.W.
Washington, DC 20004
*Mr. Albert Mindlin
(202) 727-6533

Metropolitan Washington Council of Governments
1875 I Street, N.W., Suite 200
Washington, DC 20006
Mr. John McClain
Ms. Susan Kalish
(202) 223-6800

FLORIDA

Division of Local Resource Management
Florida Department of Community Affairs
2571 Executive Ctr. Circle, E
Tallahassee, FL 32301
Mr. Matthew Brady
(904) 488-2356

GEORGIA

Georgia Office of Planning and Budget
270 Washington St., S.W.
Room 608
Atlanta, GA 30334
Mr. Clark Stevens, Director
*Mr. Tom Wagner
(404) 656-2191

Documents Librarian
Georgia State University
University Plaza
Atlanta, GA 30303
Mr. Jay McNamara
(404) 658-2185

Robert W. Woodruff Library for Advanced Studies
Emory University
Atlanta, GA 30322
Ms. Elizabeth McBride
(404) 329-6872

Main Library
University of Georgia
Athens, GA 30602
Ms. Susan C. Fields
(404) 542-8949

Georgia Dept. of Community Affairs
Office of Research & Information
40 Marietta Street, N.W.
8th Floor
Atlanta, GA 30303
Mr. Dave Wiltsee
(404) 656-3873

Documents Librarian
State Data Center Program
Albany State College
504 College Drive
Albany, GA 31705
Ms. Golda Jackson
(912) 439-4065

Documents Librarian
State Data Center Program
Georgia Southern College
Statesboro, GA 30458
Ms. Lynn Walshak
(912) 356-2183

*Denotes key contact person

State Data Center Program
Mercer University Law Library
Mercer University
Macon, GA 31207
Mr. Reynold Kosek
(912) 745-6811

University Computer Center
University of Georgia
Athens, GA 30602
Ms. Hortense L. Bates
(404) 542-3106

Price Gilbert Memorial Library
Georgia Institute of Technology
Atlanta, GA 30332
Mr. Richard Leacy
(404) 894-4519

HAWAII

State Department of Planning
and Economic Development
P.O. Box 2359
Honolulu, HI 96804
*Mr. Robert Schmitt
Ms. Maureen St. Michel
(808) 548-3082

Electronic Data Processing Div.
State Department of Budget
and Finance
Kalanimoku Building
1151 Punchbowl Street
Honolulu, HI 96813
Mr. Tom Yameshiro
(808) 548-4160

Hawaii Cooperative Health Systems
University of Hawaii
Moore Hall, #427
1890 East-West Road
Honolulu, HI 96822
Mr. Bain Henderson
(808) 948-6977

IDAHO

Division of Economic and
Community Affairs
700 W. State Street
State Capitol Bldg., Rm. 108
Boise, ID 83720
Mr. Dan Emborg, Administrator
(208) 334-2309
*Mr. Alan Porter
(208) 334-3416

University Research Center
Boise State University
1910 University Drive
Boise, ID 83725
Dr. Richard Hart, Director
(208) 385-3576
Mr. Basil Dahlstrom
(208) 385-1573

The Idaho State Library
325 West State Street
Boise, ID 83702
Ms. Helen Miller, State Librarian
Mr. Gary Bettis
(208) 334-2150

ILLINOIS

Division of Planning and
Financial Analysis
Illinois Bureau of the Budget
William Stratton Bldg.
Room 605
Springfield, IL 62706
*Ms. Kathy Roberts
(217) 782-3500

Community Research Services
Department of Sociology, Anthropology, and Social Work
Illinois State University
Normal, IL 61761
Dr. Vernon C. Philmann
(309) 438-2387

*Denotes key contact person

Center for Governmental Studies
Northern Illinois University
DeKalb, IL 60115
Ms. Ruth Anne Tobias
(815) 753-0322

Center for Urban and Environmental Research and Services
Southern Illinois University at Edwardsville
Box 32
Edwardsville, IL 62026
Mr. Charles Kofron
(618) 692-3032

Chicago Area Geographic Information Study
Room 2101, Building BSB
P.O. Box 4348
University of Illinois at Chicago Circle
Chicago, IL 60680
Mr. Eric Heckman
(312) 996-5274

INDIANA

Indiana State Library
Indiana State Data Center
140 North Senate Avenue
Indianapolis, IN 46204
Mr. Ray Ewick, Director
*Ms. Carol O. Rogers
(317) 232-3733

School of Business
Division of Research
Indiana University
635 N. Jordan
Bloomington, IN 47405
Mr. Morton Marcus
(812) 335-5507

Division of Economic Analysis
Indiana Department of Commerce
440 North Meridian
Indianapolis, IN 46204
Mr. David A. Reed, Director
(317) 232-8959

IOWA

Office of the State Demographer
Iowa Office for Planning and Programming
523 East 12th Street
Des Moines, IA 50319
Dr. Edward Stanek, Director
*Ms. Mary Tavegia
(515) 281-3738

State Library Commission
Historical Building
Des Moines, IA 50319
Ms. Linda Maurer
(515) 281-4103

Center for Social and Behavioral Research
University of Northern Iowa
Cedar Falls, IA 50614
Dr. Robert Kramer
(319) 273-2105

Census Services
Iowa State University
318 East Hall
Ames, IA 50011
Dr. Willia Goudy
(515) 294-8370

Laboratory for Political Research
University of Iowa
321 Schaeffer Hall
Iowa City, IA 52242
Mr. Jim Grifhorst
(319) 353-3103

Census Data Center
Department of Public Instruction
Grimes State Office Building
Des Moines, IA 50319
Mr. Steve Boal
(514) 281-4730

*Denotes key contact person

Census Data Center
Iowa Department of Human
Services
Hoover State Office Building
Des Moines, IA 50319
Mr. Kent Westmas
(515) 281-4694

Ballou Library
Buena Vista College
Strom Lake, IA 50588
Dr. Barbara Palling
(712) 749-2127

KANSAS

State Library
State Capitol Building
Room 343
Topeka, KS 66612
*Mr. Marc Galbraith
(913) 296-3296

Division of the Budget
State Capitol Building
Room 152-E
Topeka, KS 66612
Mr. Daina Farrell
(913) 296-2436

Institute for Economic and
Business Research
325 Nichols Hall
The University of Kansas
Lawrence, KS 66044
Mr. Robert Glass
(913) 864-3123

Center for Urban Studies
Box 61
Wichita State University
Wichita, KS 67208
Mr. Mark Glaser
(316) 689-3737

Population Research Laboratory
Department of Sociology
Kansas State University
Manhattan, KS 66506
Mr. Donald Adamchak
(913) 532-5984

KENTUCKY

Urban Studies Center
Department SDC
University of Louisville
Gardencourt Campus
Alta Vista Road
Louisville, KY 40292
*Mr. Vernon Smith
(502) 588-6626

Office for Policy & Management
State of Kentucky
Capitol Annex
Frankfort, KY 40601
Mr. William Hintze
(502) 564-7300

State Dept. of Library &
Archives
State Library Division
300 Coffeetree Road
P.O. Box 537
Frankfort, KY 40602
Mr. James Nelson, Director
(502) 875-7000

LOUISIANA

Louisiana State Planning
Office
P.O. Box 44426
Baton Rouge, LA 70804
Mr. Wallace L. Walker, Director
*Mr. Thornton Cofield
(504) 342-7410

*Denotes key contact person

Division of Business and
Economic Research
University of New Orleans
Lake Front
New Orleans, LA 70122
Ms. Jackie Hymel
(504) 286-6248

Division of Business Research
Louisiana Tech University
P.O. Box 5796
Ruston, LA 71270
Dr. Edward O'Boyle
(318) 257-3701

Reference Department
Louisiana State Library
P.O. Box 131
Baton Rouge, LA 70821
Mrs. Blanche Cretini
(504) 342-4918

Experimental Statistics Dept.
173 Agriculture Admin. Bldg.
Louisiana State University
Baton Rouge, LA 70803
Dr. Nancy Keith
(504) 388-8303

MAINE

Division of Economic Analysis
and Research
Maine Department of Labor
20 Union Street
Augusta, ME 04330
Mr. Raynold Fongemie
(207) 289-2271

MARYLAND

Maryland Dept. of State Planning
301 West Preston Street
Baltimore, MD 21201
Ms. Constance Lieder, Secretary
*Mr. Arthur Benjamin
(301) 383-5664

Computer Science Center
University of Maryland
College Park, MD 20742
Mr. Eli Schuman
Mr. John McNary
(301) 454-4323

State Library Resource Center
Enoch Pratt Free Library
400 Cathedral Street
Baltimore, MD 21201
Ms. Anne Shaw Burgan
(301) 396-5328

MASSACHUSETTS

Center for Massachusetts Data
Executive Office of Communities
and Development
100 Cambridge Street, Room 804
Boston, MA 02202
*Mr. Charles McSweeney,
Coordinator
(617) 727-3253

University Office of Center for
Massachusetts Data
University of Massachusetts
117 Draper Hall
Amherst, MA 01003
Dr. George R. McDowell, Director
(413) 545-0176

MICHIGAN

Michigan Information Center
Department of Mangement
and Budget
Office of the Budget/LLPD
P.O. Box 30026
Lansing, MI 48909
Dr. Laurence Rosen
(517) 373-7910

*Denotes key contact person

MIMIC/CUS
Wayne State University
5229 Cass Avenue
Detroit, MI 48202
Mr. William Simmons
(313) 577-2180

The Library of Michigan
Government Documents Division
P.O. Box 30007
Lansing, MI 48909
Ms. F. Anne Diamond
(517) 373-0640

MINNESOTA

State Demographic Unit
Minnesota State Planning Agency
101 Capitol Square Building
550 Cedar Street
St. Paul, MN 55101
Mr. Thomas Gillaspy
*Ms. Eileen Barr
(512) 296-4886

Minnesota Analysis and Planning System
University of Minnesota - St. Paul
475 Coffey Hall
1420 Eckles Avenue
St. Paul, MN 55101
Ms. Patricia Kovel-Jarboe
(612) 376-7003

Office of Public Libraries and Interlibrary Cooperation
Minnesota Department of Education
301 Hanover Building
480 Cedar Street
St. Paul, MN 55101
Mr. Bill Asp
(612) 296-2821

*Denotes key contact person

MISSISSIPPI

Center for Population Studies
The University of Mississippi
Bondurant Building, Room 3W
University, MS 38677
Dr. Max Williams, Director
*Ms. Michelle Ratliff
(601) 232-7288

Governor's Office of Federal - State Programs
Department of Planning and Policy
Walter Sillers Building
Jackson, MS 39202
Mr. George Parsons, Director
Ms. Jeanie E. Smith

MISSOURI

Missouri State Library
308 High Street
P.O. Box 387
Jefferson City, MO 65102
Mr. Charles O'Halloran
*Mr. Jon Harrison
(314) 751-4552

Office of Administration
124 Capitol Building
Jefferson City, MO 65101
Mr. Ryan Burson
(314) 741-2345

B and PA Research Center
University of Missouri
10 Professional Building
Columbia, MO 65211
Dr. Ed Robb
(314) 882-4805

MONTANA

Census and Economic Information Center
Montana Dept. of Commerce
1429 9th Street
Capitol Station
Helena, MT 59620-0401
*Ms. Patricia Roberts
(406) 444-2896

Montana State Library
Capitol Station
Helena, MT 59620
Mr. Harold Chambers
(406) 449-3115

Bureau of Business and
Economic Research
University of Montana
Missoula, MT 59812
Ms. Maxine Johnson
(406) 243-5113

Center for Data Systems
and Analysis
Office of the Vice President
for Research
Montana State University
Bozeman, MT 59717
Ms. Lee Faulkner
(406) 994-4481

NEBRASKA

Bureau of Business Research
200 CBA
The University of Nebraska-
Lincoln
Lincoln, NE 68588
Dr. Donald Pursell, Director
*Mr. Jerry Diechert
(402) 472-2334

Policy Research Office
P.O. Box 94601
State Capitol, Rm. 1321
Lincoln, NE 68509
Mr. Andrew Cunningham
(402) 471-2414

Nebraska Library Commission
1420 P Street
Lincoln, NE 68508
Mr. John L. Kopischke, Director
Ms. Patricia Sloan, Fed. Doc.
(402) 471-2045

The Central Data Processing Div.
Nebraska Department of Adminis-
trative Services
1306 State Capitol
Lincoln, NE 68509
Mr. Robert S. Wright,
Administrator
Mr. Skip Miller
(402) 471-2065

NEVADA

Nevada State Library
Capitol Complex
401 North Carson
Carson City, NV 89710
Ms. Joan Kerschner
*Ms. Valerie Anderson
(702) 885-5160

Department of Data Processing
Capitol Complex
Blasdell Building, Rm. 304
Carson City, NV 89710
Mr. Bob Rigsby
(702) 885-4091

NEW HAMPSHIRE

Office of State Planning
State of New Hampshire
2 1/2 Beacon Street
Concord, NH 03301
*Mr. Jim McLaughlin
(603) 271-2155

New Hampshire State Library
Park Street
Concord, NH 03301
Mrs. Shirley Gray Adamovich
(603) 271-2392

Institute of Natural and
Environmental Resources
University of New Hampshire
James Hall, 2nd Floor
Durham, NH 03824
Mr. Owen Durgin
(603) 862-1020

*Denotes key contact person

NEW JERSEY

New Jersy Dept. of Labor
Division of Planning and Research
CN 388 - John Fitch Plaza
Trenton, NJ 08625-0388
*Ms. Connie O. Hughes
(609) 984-2593

New Jersey State Library
185 West State Street
Trenton, NJ 08625
Ms. Beverly Railsback
(609) 292-4282

Princeton-Rutgers Census Data Project
Princeton University Computer Center
87 Prospect Avenue
Princeton, NJ 08544
Ms. Judith S. Rowe
(609) 452-6052

Princeton-Rutgers Census Data Project
Center for Computer & Info. Services
Rutgers University
CCIS-Hill Center, Busch Campus
P.O. Box 879
Piscataway, NJ 08854
Ms. Gertrude Lewis
(201) 932-2483

NEW MEXICO

Economic Development and Tourism Department
Bataan Memorial Building
Santa Fe, MN 87503
*Mr. John Velasco
(505) 827-6200

New Mexico State Library
P.O. Box 1629
Santa Fe, NM 88003
Ms. Sandra Faull
(505) 827-2033

Bureau of Business and Economic Research
University of New Mexico
Albuquerque, NM 87131
Dr. Lee Brown, Director
(505) 277-2626

Center for Business Research and Services
Box 3CR
New Mexico State University
Las Cruces, NM 88003
Dr. Ken Nowotny
(505) 646-2035

NEW YORK

Division of Economic Research and Statistics
New York Department of Commerce
Twin Towers, Room 1005
99 Washington Avenue
Albany, NY 12245
Mr. Peter Ansell, Assistant Deputy Commissioner
*Mr. Mike Batutis
(518) 474-5944
(518) 474-6005

Law and Social Sciences Unit
New York State Library
Cultural Education Center
Empire State Plaza
Albany, NY 12230
Ms. Elaine Clark
(518) 474-5128

*Denotes key contact person

NORTH CAROLINA

North Carolina Office of State Budget and Management
116 West Jones Street
Raleigh, NC 27611
Ms. Francine Ewing, Director of State Data Center
(919) 733-7061

State Library
North Carolina Dept. of Cultural Resources
109 East Jones Street
Raleigh, NC 27611
Ms. Earlean Strickland
(919) 733-3343

Institute for Research in Social Science
University of North Carolina
Manning Hall 026A
Chapel Hill, NC 27515
Ms. Judy Moses
(919) 966-3346

NORTH DAKOTA

Dept. of Agricultural Economics
North Dakota State University
Agricultural Experiment Station
Morrill Hall, Room 207
P.O. Box 5636
Fargo, ND 58105
Dr. Jerome Johnson
*Dr. Richard Rathge
(701) 237-8621

North Dakota State Planning Div.
State Capitol, 17th Floor
Bismarck, ND 58505
Mr. Ronald Bostick, Director
(701) 224-2818
Ms. Kathy Lindquist
(701) 224-2094

Department of Geography
University of North Dakota
Grand Forks, ND 58202
Mr. Floyd Hickok
(701) 777-4593

North Dakota State Library
Randall Building
Highwary 83N
Bismarck, ND 58505
Ms. Ruth Mahan
(701) 224-2490

OHIO

Ohio Data Users Center
Ohio Department of Economic and Community Development
P.O. Box 1001
Columbus, OH 43216
*Mr. Jack Brown
(614) 466-7772

OKLAHOMA

Oklahoma State Data Center
Department of Economic and Community Affairs
Lincoln Plaza Bldg., Suite 285
4545 North Lincoln Blvd.
Oklahoma City, OK 73105
Ms. Cindy Rambo, Director
*Mr. Harley Lingerfeld
(405) 528-8200

Oklahoma Department of Libraries
200 N.E. 18th Street
Oklahoma City, OK 73105
Ms. Virginia Collier
(405) 521-2502

OREGON

Intergovernmental Relations Div.
Executive Building
155 Cottage Street, N.E.
Salem, OR 97310
Mr. Jack Carter
*Mr. Jon Roberts
(503) 373-1996

*Denotes key contact person

Bureau of Governmental Research
and Service
School of Community Service and
Public Affairs
University of Oregon
Hendricks Hall, Room 340
P.O. Box 3177
Eugene, OR 97403
Ms. Karen Seidel
(503) 686-5232

Center for Population Research
and Census
Portland State University
P.O. Box 751
Portland, OR 97207
Mr. Ed Shafer
(503) 378-4502

PENNSYLVANIA

Institute of State and
Regional Affairs
Pennsylvania State University
Capitol Campus
Middletown, PA 17057
Mr. Bob Surridge
(717) 948-6336

Department of Education
State Library of Pennsylvania
Forum Building
Harrisburg, PA 17120
Mr. John Gerswindt
(717) 2327

Governor's Office of Budget
and Administration
Bureau of Management Services
903 Health and Welfare Bldg.
Harrisburg, PA 17120
Mr. Ray Kaspar
(717) 787-1764

PUERTO RICO

Puerto Rico Planning Board
Minillas Government Center
North Bldg., Avenida De Diego
P.O. Box 41119
San Juan, PR 00940
*Mr. Suriel Sanchez
(809) 726-5020

General Library
University of Puerto Rico
Road #2
Mayaguez, PR 00708
Dra. Luisa Viga-Cepeda,
Director
(809) 832-4040

Department of Education
Carnegie Library
P.O. Box 759
Hato Rey, PR 00619
Ms. Carmen Martinez
(809) 724-1046

RHODE ISLAND

Rhode Island Statewide
Planning Program
265 Melrose Street, Rm. 203
Providence, RI 02907
Mr. Daniel Varin, Chief
*Mr. Chester Symanski
(401) 277-2656

Rhode Island Department of
State Library Services
95 Davis Street
Providence, RI 02908
Mr. Frank Iacono
(401) 277-2726

Social Science Data Center
Department of Sociology
Brown University
Maxcy Hall, Angel Street
P.O. Box 1916
Providence, RI 02912
Dr. James Sakoda
(401) 863-2550

*Denotes key contact person

Rhode Island Health Services
Research, Inc.
56 Pine Street
Providence, RI 02903
Mr. Lawrence Manire
(401) 331-6105

Rhode Island Department of
Community Affairs
Division of Housing and
Government Services
150 Washington Street
Providence, RI 02903
Mr. Joseph G. Simeone
(401) 277-2892

SOUTH CAROLINA

Division of Research and
Statistical Services
Budget and Control Board
State of South Carolina
Rembert C. Dennis Bldg., B/341
1000 Assembly Street
Columbia, SC 29201
Mr. Bobby Bowers, Chief,
Demographic Statistics
*Mr. Mike Mcfarlane
(803) 758-3138

South Carolina State Library
P.O. Box 11469
Columbia, SC 29211
Mary Toll, Documents Librarian
(803) 758-3138

SOUTH DAKOTA

Business Research Bureau
School of Business
Patterson Hall
University of South Dakota
Vermillion, SD 57069
*Ms. Karen Bihlmeyer
(605) 677-5289

The State Planning Bureau
South Dakota Department of
Executive Management
State Capitol Building
Pierre, SD 57501
Mr. Tony Merry, Commissioner
(605) 773-3661

Documents Department
The South Dakota State Library
Department of Education and
Cultural Affairs
800 N. Illinois Avenue
Pierre, SD 57501
Ms. Rose Oniewski
(605) 773-3131

Research and Statistics Unit
South Dakota Dept. of Labor
607 North 4th Street
Aberdeen, SD 57401
Ms. Mary Susan Vickers
(605) 622-2314

Vital Records Program
South Dakota Dept. of Health
Foss Building
Pierre, SD 57501
Mr. William D. Johnson
(605) 773-3355

Rural Sociology Department
South Dakota State University
Scobey Hall, 226
Brookings, SD 57006
Dr. Marvin P. Riley
Dr. Jim Satterlee
(605) 688-4132

TENNESSEE

Tennessee State Planning Office
James K. Polk State Office Bldg.
505 Deadrick Street, Suite 1800
Nashville, TN 37219
Mr. Lewis Lavine, Executive
Director
*Mr. Charles Brown
(615) 741-1676

*Denotes contact person

Center for Business and
Economic Research
University of Tennessee
Room 100, Glocker Hall
Knoxville, TN 37916
Ms. Betty Vickers
(615) 974-5441

TEXAS

Data Management Program
Governor's Office of Planning
and Intergovernmental Relations
P.O. Box 13561
Sam Houston Bldg., Room 411
Austin, TX 78711
*Ms. Bonnie Young
(512) 475-8386

Department of Rural Sociology
Texas A and M University System
Special Services Building
College Station, TX 77843
Dr. Steve Murdock
(409) 845-5115

Texas Natural Resources
Information System
P.O. Box 13087
Austin, TX 78711
Mr. John Wilson
(512) 475-3321

Texas State Library and
Archive Commission
P.O. Box 12927
Capitol Station
Austin, TX 78711
Mr. Allen Quinn
(512) 475-2996

UTAH

Office of Planning and Budget
State Capitol, Room 116
Salt Lake City, UT 84114
Mr. Kent Briggs, Director
Mr. Brad Barker
*Mr. Jim Robson
(801) 533-6082

Bureau of Economic and
Business Research
Business Building
University of Utah
Salt Lake City, UT 84112
Ronda Brinkerhoff
(801) 581-6333

Population Research Laboratory
Utah State University
Logan, UT 84322
Mr. William Stinner
(801) 750-1242

Department of Employment Security
174 Social Hall Avenue
P.O. Box 11249
Salt Lake City, UT 84147
Mr. Ken Jensen
(801) 533-2436

VERMONT

Vermont State Planning Office
Pavilion Office Building
109 State Street
Montpelier, VT 05602
Mr. Bernard Johnson
*Mr. David Healy
(802) 828-3326

Center for Rural Studies
University of Vermont
25 Colchester Avenue
Burlington, VT 05401
Mr. Fred Schmidt, Director
Mr. Sam McReynolds
(802) 656-3021

Vermont Department of Libraries
111 State Street
Montpelier, VT 05602
Ms. Patricia Klinck, State
Librarian
(802) 828-3265

*Denotes key contact person

Vermont Agency of Development
and Community Affairs
Pavilion Office Building
109 State Street
Montpelier, VT 05602
Mr. Barry Driscoll
(802) 828-3211

VIRGINIA

Department of Planning & Budget
445 Ninth Street Office Bldg.
P.O. Box 1422
Richmond, VA 23211
Mr. Stuart W. Connock, Director
*Ms. Julie Henderson
(804) 786-7843

Taylor Murphy Institute
University of Virginia
Dynamics Building, 4th Floor
2015 Ivy Road
Charlottesville, VA 22903
Dr. Charles Meiburg, Director
Dr. Julie Martin
Dr. Michael Spar
(804) 971-2661

Virginia State Library
12th and Capitol Street
Richmond, VA 23219
Ms. Linda Morrissett
(804) 786-2175

VIRGIN ISLANDS

Department of Commerce of the
Virgin Islands
P.O. Box 6400
Charlotte Amlie
St. Thomas, VI 00801
Mr. Richard Moore
(809) 774-8784 x214

WASHINGTON

Forecasting & Estimation Div.
Office of Financial Management
400 East Union
Mail Stop ER-13
Olympia, WA 98504
*Mr. Lawrence Weisser
(206) 754-2808

Washington State Library
State Library Building
Olympia, WA 98504
Mr. Roderick G. Swartz
Mr. Rushton Brandis
(206) 753-5424

Urban Data Center
University of Washington
Seattle, WA 98195
Mr. Edgar Horwood, Director
Mr. Bob Shawcroft
(206) 543-7625

Social Research Center/
Department of Rural Sociology
Room 133, Wilson Hall
Washington State University
Pullman, WA 99164
Dr. Annabel Cook
(509) 335-1511

Department of Sociology/
Demographic Research Laboratory
Western Washington University
Bellingham, WA 98225
Mr. Lucky Tedrow, Director
(206) 676-3617

Technical Information Services/
University Library
Eastern Washington University
Cheney, WA 99004
Mr. Jay Rea
(509) 235-2475

*Denotes key contact person

Office of Institutional Studies
Central Washington University
Ellensburg, WA 98926
Dr. John Purcell, Director
Mr. John R. Dugan
(509) 963-1856

WEST VIRGINIA

Community Development Division
Governor's Office of Economic and Community Development
Capitol Complex
Building 6, Room 553
Charleston, WV 25305
Mr. Miles Dean, Director, Governor's Office of Economic & Community Development
Mr. Fred Cutlip, Director, Community Development Division
*Ms. Katherine Shiflet
(304) 348-4010

Reference Library
West Virginia State Library Commission
Science and Cultural Center
Capitol Complex
Charleston, WV 25305
Ms. Karen Goff
(304) 348-2045

Office of Health Services Research
Department of Community Health
West Virginia University
900 Chestnut Ridge Road
Morgantown, WV 26505
Ms. Virginia Petersen
(304) 293-2601

WISCONSIN

Demographic Services Center
Department of Administration
101 South Webster Street
7th Floor
P.O. Box 7864
Madison, WI 53707
Mr. Don Holl
(608) 266-1067
*Mr. Robert Naylor
(608) 266-1927

Applied Population Laboratory
Department of Rural Sociology
University of Wisconsin
1450 Linden Drive
Madison, WI 53706
Ms. Doris Slesinger
Mr. Stephen Tordella
(608) 262-1515

*Denotes key contact person

APPENDIX D

PILOT STUDY DATA REQUEST

This appendix contains a data request for telephone usage data that is similar to the one intended to be used in the West Virginia pilot study. The data request may be used as a model for preparing similar requests in other states. If so used one should recognize that it is really three data requests in one. Depending on the objectives of a particular study any one of the pilot study data request sections may apply.

DATA REQUEST

The West Virginia Public Service Commission requests the assistance of the C & P Telephone Company in providing data for a study of telephone usage in West Virginia. The request has three parts that are organized into the sections that follow. To insure the anonymity of West Virginia residents whose amount of telephone usage will be observed during the study we ask that no specific information about individual subscribers be provided. Subscriber codes used to transmit information to the Commission should not be the same codes used for regular business purposes. For example, you should avoid the use of billing codes, telephone numbers, and addresses.

Subscriber Line Usage (SLU) Sample

In order to estimate the value of SLU, C & P routinely measures usage of a random sample of households in West Virginia. Please provide SLU usage data for the next SLU study period that can be coordinated with the sample described in the next section. The information provided for each household in the SLU sample should consist of the following:

1. A subscriber code developed for this study
2. The exchange name and the name of the street on which the residence is located and either the names of the two streets intersecting the resident's street on either side of the residence or the hundreds block designation of resident's address (for example, on Main Street between Elm and Willow, or the 1300 block of Main Street)
3. Whether or not the subscriber has elected one of the local measured service options, and if so which one
4. For each 4-hour period of each day of the study, tabulate both the number of calls of each type and the number of minutes of use of each type

The call types of interest are:

1. Local exchange
2. Interstate toll
3. Intrastate toll

Systematic Sample in the Charleston, WV Exchange

In addition to the SLU sample it is requested that an additional 240 households be sampled. In this case the sampled households are to be concentrated in 5 specific geographic areas in the Charleston exchange. The sample should be determined by randomly selecting 48 households from each of the geographic areas shown in figures D-1 through D-5[1]. By the terms "randomly selecting" we mean that the selection procedures used should insure that each household in a given sampling area should have an equally likely chance of being selected. An exception to this random selection rule is when the total number of households in a specified area is less than the sample size of 48 as is the case with the area in figure D-1 which had a population of 34 households in the 1980 census. In these cases the sample size in the sampling area is reduced to the actual population and every household in the area is sampled.

The information provided for each of the 240 (or less) households is as follows:

1. A subscriber code developed for this study
2. The number of the geographic sampling area (GSA) where the residence is located
3. Whether or not the subscriber has elected one of the local measured service options, and if so which one
4. For each 4-hour period of each day of the study, tabulate both the number of calls of each type and the number of minutes of use of each type

---

[1]The actual U.S. Census maps were used in the data request.

The call types of interest are:

1. Local exchange

2. Interstate toll

3. Intrastate toll

The timing of this part of the study should be coordinated so that the usage data are collected during the same time as the SLU data.

Business Sample

Within the Charleston exchange select randomly ten multi-line business subscribers and report the following:

1. A subscriber code developed for this study

2. For each 4-hour period of each day of the study, tabulate both the number of calls of each type and the number of minutes of use of each type. This measurement should be made on all the lines from one of the major facilities of the subscriber

In addition select randomly ten single-line business subscribers and report the following:

1. A subscriber code developed for this study

2. For each 4-hour period of each day of the study, tabulate both the number of calls of each type and the number of minutes of use of each type

The call types of interest are:

1. Local exchange

2. Interstate toll

3. Intrastate toll

Study Schedule

Usage measurement for all three parts of the study should commence on ____________________ and terminate on ______________________.

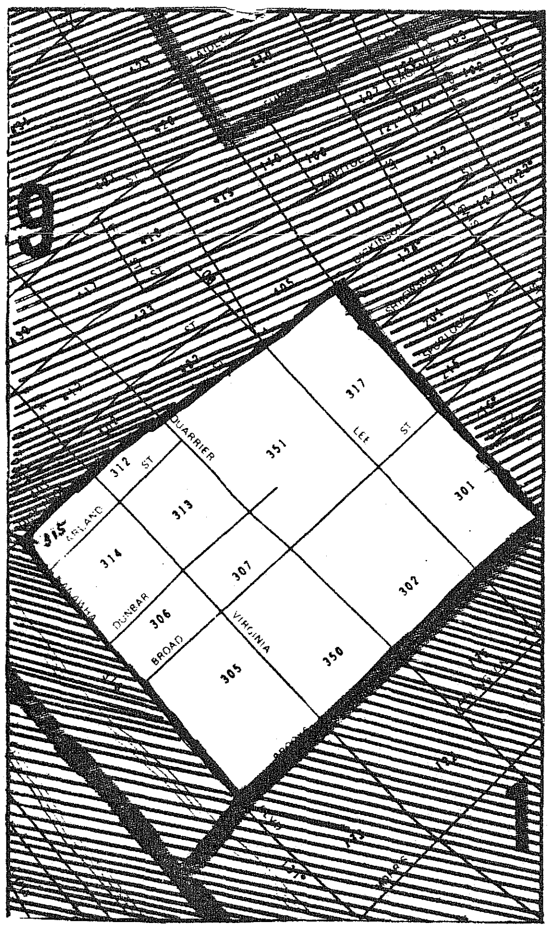

Fig. D-1 The GSA 1 of Block Group 3, Census Track 9, Place 280, Minor Civil Division 26, Kanawha County (39)

Source: Charleston, W.V., Metropolitan Map Series 1480, Map Sheet 4SW

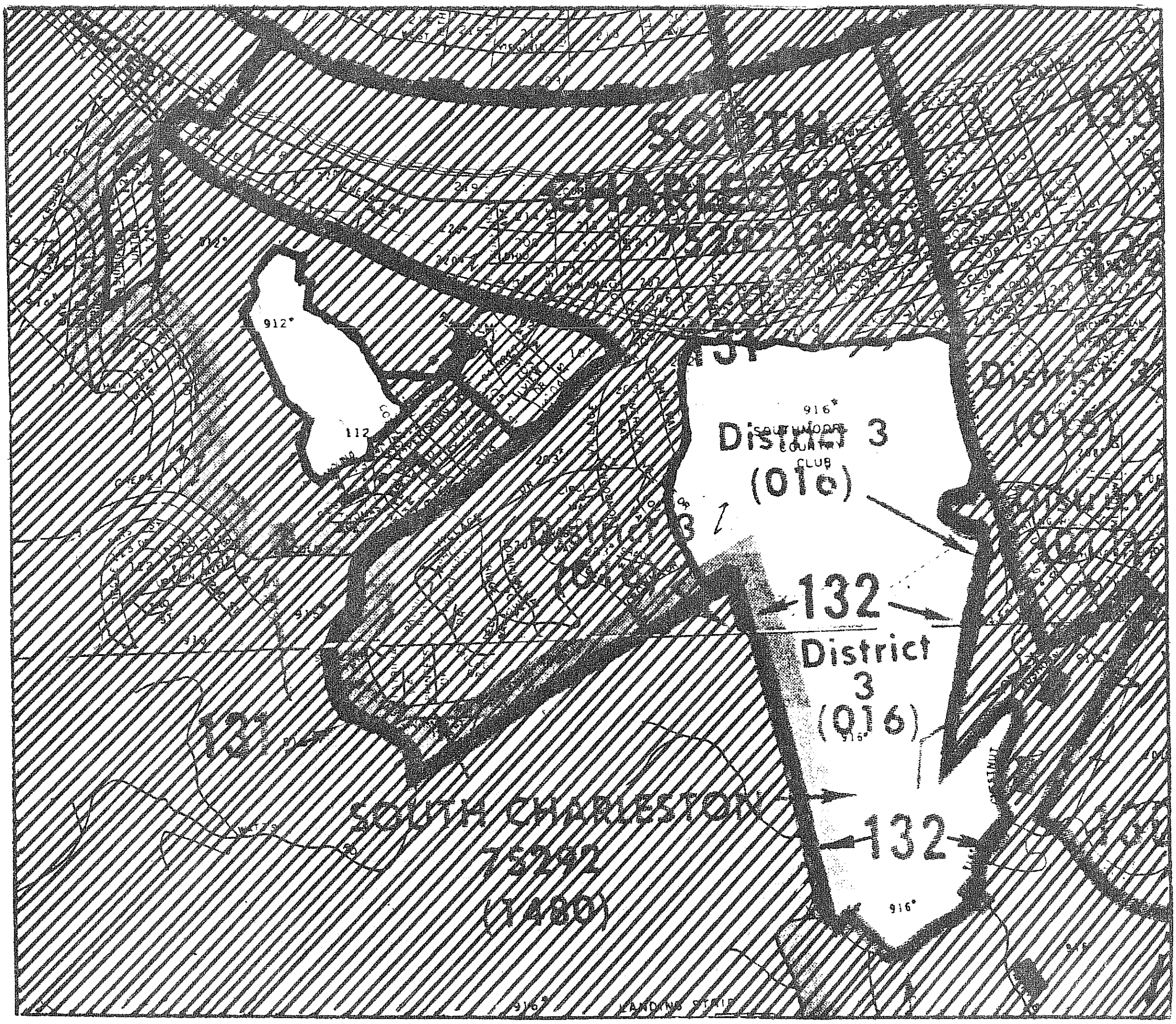

Fig. D-2 The GSA 3 of Block Group 9, Census Track 131, Place 1480, Minor Civil Division 16, Kanawha County (39)

Source: Charleston, W.V., Metropolitan Map Series 1480, Map Sheet 3 and 6

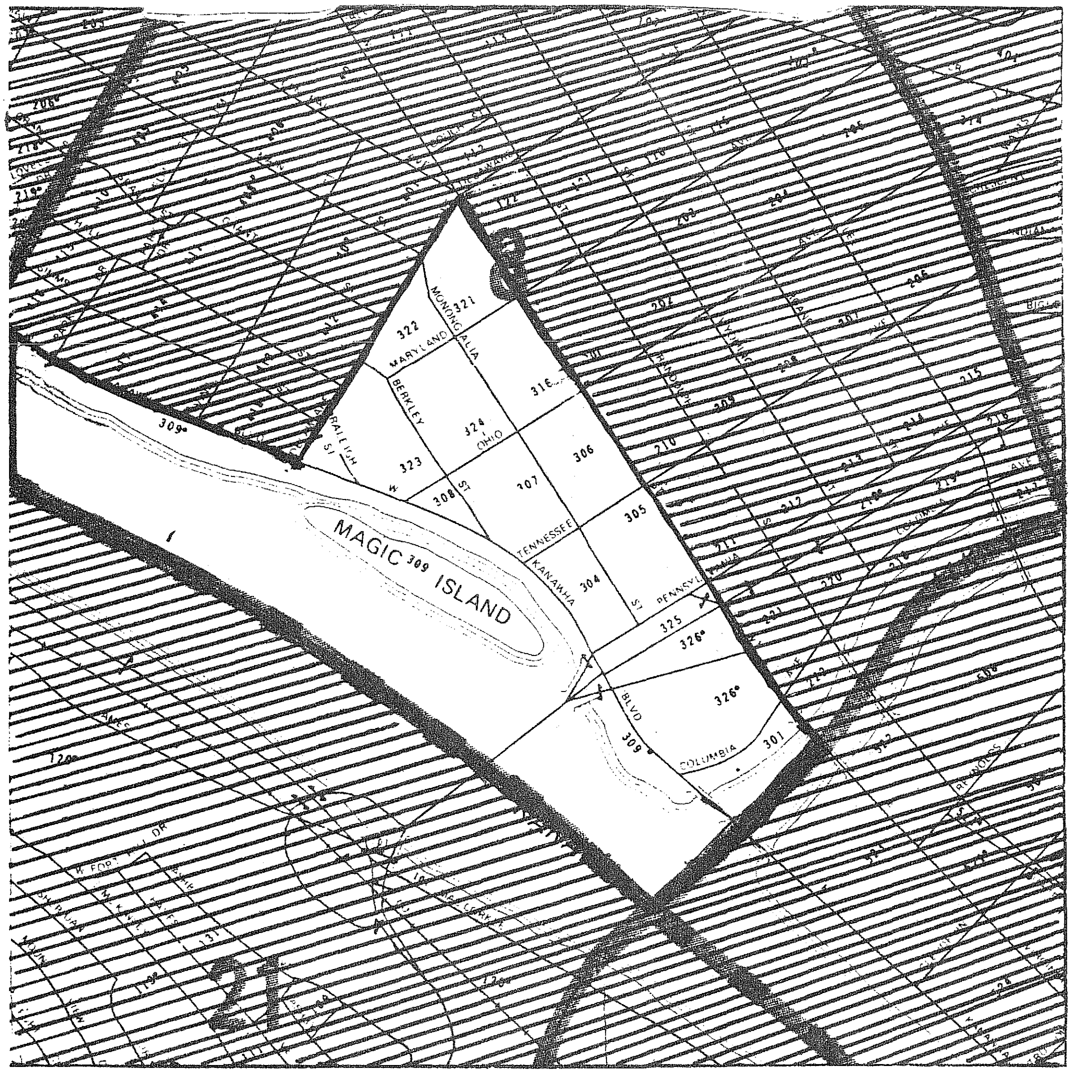

Fig. D-3 The GSA 8 of Block Group 3, Census Track 8, Place 820, Minor Civil Division 26, Kanawha County (39)

Source: Charleston, W.V., Metropolitan Map Series 1480, Map Sheet 4SW

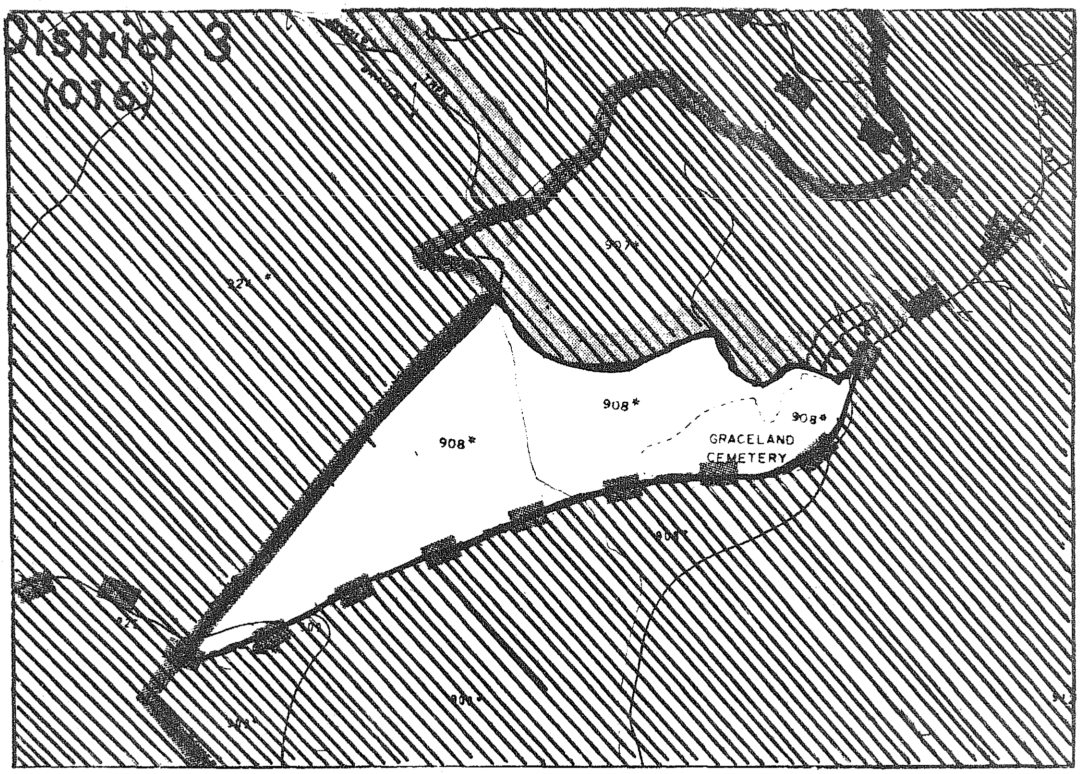

Fig. D-4 The GSA 21 of Block Group 9, Census Track 127, Place 9999, Minor Civil Division 16, Kanawha County (39)

Source: Charleston, W.V., Metropolitan Map Series 1480, Map Sheet G

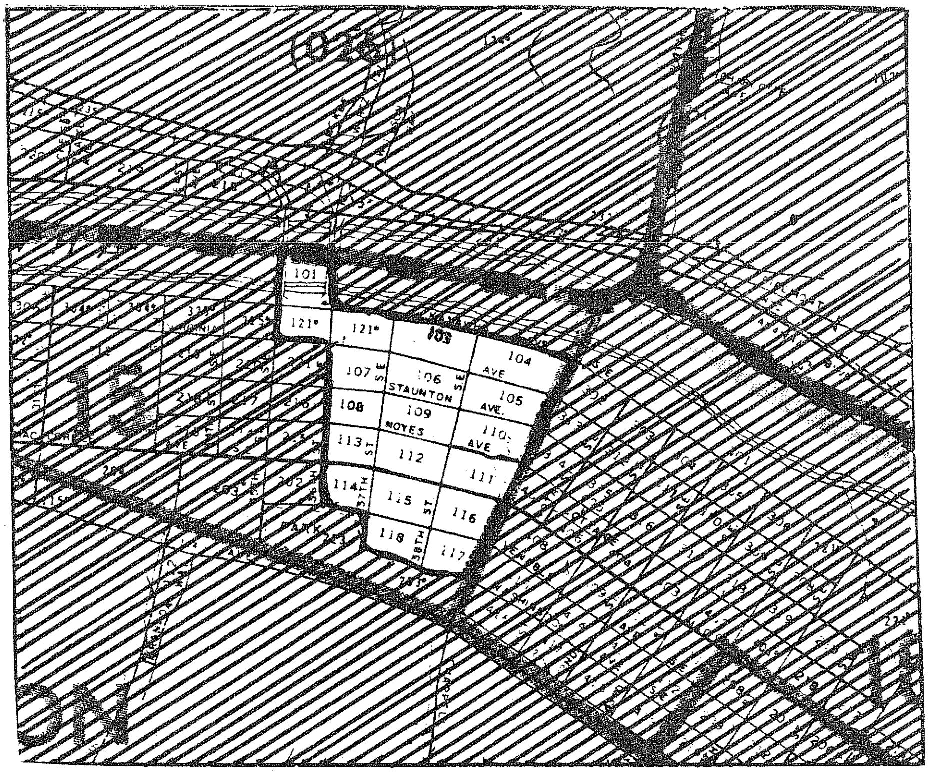

Fig. D-5 The GSA 25 of Block Group 1, Census Track 15, Place 280, Minor Civil Division 11, Kanawha County (39)

Source: Charleston, W.V., Metropolitan Map Series 1480, Map Sheet 5