

DESCRIPTION AND EVALUATION OF A DESIGN FOR
SAMPLING HOUSEHOLDS "AT RISK" OF WELFARE DEPENDENCY

by

John Tarnai
Social & Economic Sciences Research Center
Washington State University
Pullman, WA 99164-4014

Steven G. Heeringa
Sampling Section
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106-1248

Don A. Dillman
Social & Economic Sciences Research Center
Washington State University

Ernst Stromsdorfer
Department of Economics
Washington State University

Greg Weeks
Washington State Institute for Public Policy
The Evergreen State College
Olympia, WA 98505

Presented at the 46th Annual Conference of the American Association for Public Opinion Research, Phoenix, Arizona, May 16-19, 1991.

ABSTRACT

DESCRIPTION AND EVALUATION OF A DESIGN FOR SAMPLING HOUSEHOLDS "AT RISK" OF WELFARE DEPENDENCY

by

John Tarnai, Steven G. Heeringa, Don A. Dillman,
Ernst Stromsdorfer, & Greg Weeks

This paper describes and evaluates a unique technique for obtaining a general population sample, with probability proportional to being "at risk" of public assistance. In 1987, the Washington State Legislature commissioned a five year longitudinal study of public assistance recipients and persons "at risk" of becoming eligible for assistance. This paper describes a sample design for obtaining a multistage sample of persons "at risk" of public assistance. The sample design involved selecting a stratified random subsample of 100 persons already on public assistance as an indicator sample, and using their address information to identify 100 geographic area segments for selection of a general population sample. The indicator sample was selected with equal probability, and the corresponding sample of area segments were therefore selected with probability proportionate to the number of public assistance addresses located within the area segment boundaries. This paper evaluates the sample design by comparing the obtained incidence of public assistance recipients obtained in the "at risk" sampling approach, with the number of public assistance recipients estimated for each of the 100 area segments. This general sample design has applications to other situations, where indicator data is available, and where the concept of being "at risk" is useful.

Contact: John Tarnai
Social & Economic Sciences Research Center
Washington State University
Pullman, WA 99164-4014
(509) 335-1511

Sampling Households "At Risk" of Public Assistance

DESCRIPTION AND EVALUATION OF A DESIGN FOR
SAMPLING HOUSEHOLDS "AT RISK" OF WELFARE DEPENDENCY

by

John Tarnai, Steven G. Heeringa, Don A. Dillman,
Ernst Stromsdorfer, & Greg Weeks

Introduction

In 1987, the Washington State Legislature commissioned a five year longitudinal study of public assistance recipients and persons "at risk" of becoming eligible for assistance. The primary purpose of this study is to determine the causes of public dependency in Washington State and to identify the major influences on movement on and off of public assistance. The study was designed as a longitudinal panel survey of 2,100 Washington State residents, selected from one of two populations, residents already on public assistance, and residents "at risk" of public assistance. Designing a sampling plan for the sample of public assistance recipients was a straight forward process with few difficulties. However, designing a sampling plan for persons defined as "at risk" of public assistance was considerably more difficult.

The purpose of this paper is to describe the development of a design for obtaining a multistage sample of persons "at risk" of public assistance. The first task of this plan was to define what is meant by being "at risk" of public assistance. Since the goal of the study was to examine movement on and off of public assistance, "at risk" meant people who were at some moderate to high probability of going on public assistance. However, since the sample was to be obtained through area sampling methods, it would be impossible to measure this risk prior to the actual sample draw. Thus, the design used an indicator approach to sampling.

This approach involved selecting a stratified random subsample of 100 persons already on public assistance, and using their address information to identify 100 area segments for

Sampling Households "At Risk" of Public Assistance

selection of a general population sample. The indicator sample was selected with equal probability, and the corresponding sample of area segments were therefore selected with probability proportionate to the number of public assistance addresses located within the area segment boundaries. Thus, an area segment which contained three public assistance addresses had three times the selection probability of an area segment with only one public assistance address.

A second major purpose of this paper is to evaluate the sample design by comparing the obtained incidence of public assistance recipients obtained in the "at risk" sampling approach, with the number of public assistance recipients estimated for each of the 100 area segments. Accomplishing this required a visual examination of the household listings for each area segment and comparing this to a sorted listing of public assistance addresses. Then, all public assistance addresses that fell within the boundaries of an area segment were counted. Except for rural areas, where post office addresses made this matching impossible, the correlation between estimates of public assistance rates obtained from the public assistance records and from the sample itself should be highly positive.

Background

The Family Independence Study (FIS) was designed as a longitudinal panel survey of 2,100 Washington State respondents selected from AFDC beneficiaries and persons "at risk" of welfare dependency. Interviews in the first year of the study were conducted face-to-face, with telephone follow-up as needed to obtain complete information from respondents.

The study questionnaire was designed to get information about family economics, including labor market behavior, public assistance history, household composition, educational experiences, and selected information about the psychosocial characteristics of respondents. A number of questions were adapted from questionnaires used in the Panel Study of Income Dynamics (PSID) and the Census Bureau's Survey of Income and Program Participation (SIPP).

Sampling Households "At Risk" of Public Assistance

Both of these surveys deal with populations and policy issues similar to the FIS. The questionnaire was designed as a three part instrument, with each part obtaining different information. Part I of the questionnaire consisted of an 11-page booklet that obtained identifying information about respondents and households. A major reason for separating this part of the questionnaire from the main body of the questionnaire was to ensure greater confidentiality and anonymity to survey responses. The main body of the questionnaire consisted of a 93-page booklet, containing questions on 27 different topics, including: migration history, family history, periods of working for pay, self-employment, looking for work, not looking for work, unemployment compensation, schooling and training, periods of food stamp receipt, AFDC receipt, medical assistance receipt, public assistance history, other assets and income, housing, life events, personal health, child health, school activities, child care, and food economics.

A key feature of this questionnaire was its use of a timeline to assist respondents with recalling their labor market and public assistance experiences during the past year. The timeline was divided into weekly periods since June 1987. This timeline was used by interviewers to determine the number of separate periods of labor market activity or public assistance that a respondent had.

The third part of the questionnaire was its own booklet of questions and was designed to be completed by respondents without the assistance of the interviewer. This part of the questionnaire consisted of six pages of psychosocial questions that obtained information on respondents' self-esteem, sense of personal control, depression, social support, financial support, and emotional support. Respondents were instructed to complete this questionnaire by themselves and to seal the completed questionnaire in an attached envelope. This procedure was used to encourage anonymity and honesty in responding to sensitive psychosocial information. Dillman's (1978) total design procedures were adopted to increase the likelihood of questionnaires being returned.

Sampling Households "At Risk" of Public Assistance

Respondent Selection: For the beneficiary sample, the respondent was the AFDC beneficiary. For the "at risk" sample, respondent selection procedures were more involved. The goal for this sample was to select the respondent who, in general, would be most "at risk" of welfare dependency. This was also done to increase comparability between the beneficiary and "at risk" samples. Selecting a respondent in an "at risk" household required that the interviewer first enumerate the persons aged 16 or older in the household. For each person living in the household, the interviewer obtained gender and age information. For females, information was also obtained on whether they had any children under the age of 16. The selection procedure then consisted of the following rules: (1) First, select a woman between 16 and 45 years of age and who has a child or children under 16 years of age. (2) If there is no such woman, then select a woman between 16 and 45 years of age. (3) If there is no woman of this age, then select a woman over 45 years of age. (4) If there are no women in the household, then select the youngest man over the age of 16. The rules also specified that if more than one respondent at any step in the process was eligible, the youngest eligible person was to be selected.

Sample Design

The purpose of the survey was to study people currently on welfare and people "at risk" of welfare dependency. Therefore, the sample design included two probability samples of Washington State households: (1) a beneficiary sample of households on welfare; and (2) a sample of households "at risk" of welfare dependency. Three objectives guided the choice of sample design. The first objective was to select respondents in such a way as to match the study design for an independent evaluation study of the Family Independence Program (FIP). The second objective was to minimize interview costs by not having to interview sample members in all regions of the state. The third objective was to use the beneficiary sample to select the "at risk" sample proportional to their probability of risk of welfare dependency.

Sampling Households "At Risk" of Public Assistance

Beneficiary Sample: The beneficiary sample design was a two-stage stratified random sample of welfare recipients. The sample frame was the March 1988 warrant roll computer tape, which was accessed through the Washington State Department of Social and Health Services (DSHS). This tape contained information on 81,767 welfare recipients and 43 Central Service Offices (CSOs) from which welfare recipients received services.

The primary stage of the sample design involved assigning the 43 CSOs to 18 strata. The criteria used to define the 18 strata included (a) geographic region and urban/rural status; and (b) caseload size and proximity to major geographic regions. Fourteen of the strata included multiple CSO's, but four included only one large CSO serving a major metropolitan area, and these four were thus self-representing. Strata were formed so as to keep the total caseload size for any stratum between 3,500 and 5,000 beneficiaries. In the fourteen strata with multiple CSOs, a single primary stage unit (PSU) was selected to represent the stratum. Selection of PSU's was done with probability proportionate to size (PPS) as measured by the 1987 average monthly beneficiary caseload for the CSO's.

One additional criteria in forming the 18 strata was to match the FIP evaluation sample design, for a separate administrative study of FIP that was being undertaken. To accomplish this, the "Keyfitz method" (Keyfitz, 1951) for PPS selection was applied to retain five strata as treatment sites, and five strata as control sites. In the five treatment strata, CSOs would adopt the FIP program within the following six months. In the five control strata, CSOs would not adopt FIP for at least five years. In the remaining eight strata, FIP could be adopted at any time within the next five years.

Thus, one feature of the sample design was the creation of three sample domains. The first domain consisted of five CSOs representing the FIP experimental areas. The second

Sampling Households "At Risk" of Public Assistance

domain consisted of five CSOs representing the FIP control areas. The third domain consisted of eight CSOs representing the areas where FIP could be implemented at any time.

The second stage sampling of beneficiaries within CSOs consisted of a systematic random sample drawn from the March 1988 tape listing the 81,767 welfare recipients in the state. For each CSO selected in the primary stage, a systematic random sample of beneficiaries was selected, according to the following selection equation:

$$f(\text{Domain}) = [\text{MOS}(\text{cso})/\text{MOS}(\text{stratum})] * [\text{k}(\text{stratum})/\text{MOS}(\text{cso})]$$

The two-stage sampling procedure gave each beneficiary an equal probability of being selected for the study.

At Risk Sample: The concept of being "at risk" of welfare dependency is a probabilistic one. All individuals can be considered to be at some risk of welfare dependency, although some individuals, such as single female parents, will be at greater risk than others, such as single males. The question was could a sample design be developed that incorporated this idea of differential probability of risk of welfare dependency. The goal of such a sample design would be to allocate the sample in proportion to the probability of risk of going on welfare. Since the purpose of the study was to examine people who were "at risk" of welfare dependency, the sample should be more likely to include people with greater risk. A problem in developing this sample design was how to identify who was at greater risk of welfare dependency. Prior to making contact with households selected for the study, there would be no way of knowing whether individuals were at low or high risk of welfare dependency. One fact that is known, however, is that income is related to where people live, and income is also related to risk of welfare dependency. People with higher incomes tend to be at lower risk of welfare dependency, and also live in areas occupied by other people with high incomes. People with lower incomes tend to be at greater risk of welfare dependency, and tend to live in areas of low income housing. We decided to take advantage of this fact, and to develop a sample design

Sampling Households "At Risk" of Public Assistance

that used density of welfare beneficiaries as an indicator of the probability of being "at risk" of welfare dependency.

The at risk sample design was a multi-stage area probability sample of "at risk" households in Washington State. This sample was a statewide probability sample of Washington's general household population. The probability of selecting a household was proportional to its likelihood of being "at risk" of welfare dependency.

The primary stage of sample selection was the same stratified sampling of CSOs which was used for the sample of beneficiaries, including the CSO domain categories of experimental, control, and additional. The second stage of sample selection was the selection of 100 area segment units representing the State. The area segment units were defined objectively on the basis of the 1980 Census data file (STF-1B) including census blocks and enumeration districts (EDs) in Washington State. Area segment units were defined to include a minimum of 96 occupied housing units, and were selected with probability proportionate to the number of welfare beneficiaries residing in the area. The PPS procedure involved sorting the datafile listing of welfare beneficiaries by domain, CSO, and by zipcode, and allocating the sample to each domain in proportion to its share of the statewide total welfare caseload (Sudman, 1976, pp 134-135). Then a systematic random sample of 100 beneficiaries was selected to serve as indicator cases. Thus, if three welfare beneficiaries reside in one area segment, the probability of that area segment being selected is three times that of area segments with only one welfare beneficiary.

The process of linking the indicator case to a unique area segment was a laborious one, that involved locating the address on a map, identifying the Census tract, block, or enumeration district (ED) containing the address, and then identifying the same tract, block, or ED on the Census STF-1B listing, and then joining tracts until an area segment of at least 96 households had been created. Another area segment was also created, consisting of the tracts immediately

Sampling Households "At Risk" of Public Assistance

following the one just created, in order that a check for "zero area segments" could be conducted. So that area segments containing no welfare beneficiaries would have some chance of being included in the sample, those zero area segments were objectively linked with the immediately adjacent area segment.

To ensure the correct statewide representation of the at risk population a check was conducted to search for area segments that contained no welfare beneficiary cases. Persons living within such zero segment areas would have a zero probability of being selected into the at risk sample. To ensure that everyone had some probability of being selected into the at risk sample, it was necessary to identify zero segment areas and objectively link them to selected area segments.

For practical reasons the process of searching for zero segment areas was constrained to the 100 area segments that followed sequentially the 100 selected area segments in the 1980 census block listing. Each of these area segments was identified and mapped so that physical addresses could be compared to street addresses on the beneficiary warrant roll tape. To determine whether an area segment was a zero-segment required searching the computer tape to see if an address on the tape could be found within the boundaries of the area segment map. If just one beneficiary address could be found then the area segment was a non-zero segment and could be ignored. If no matching address could be found, then the search continued until we were certain the search was fruitless, and the segment was considered as a zero segment. Only three area segments of the 100 segments searched, were found to be zero-segment areas, and had to be linked to other selected area segments. This ensured statewide representation, and ensured that all area segments had a known, non-zero probability of being selected.

All 100 area segments created in this process were then mapped, and sent to field staff to conduct a complete household enumeration of the area segment (Kish, 1965, pp 334-350).

Sampling Households "At Risk" of Public Assistance

Once all households within an area segment had been enumerated, it was possible to implement the final stage of sampling. The third stage of sampling consisted of a systematic sample of households within the 100 area segments. A total of 15 households were selected from each segment for inclusion in the sample. The selection equation for the "at risk" sample is of the following form:

$$f(\text{Household}) = 100[\text{MOS}(\text{domain})/\text{MOS}(\text{state})][\text{MOS}(\text{AS})/\text{MOS}(\text{domain})][15/\text{NoHH}(\text{AS})]$$

Results

Response Rate: Of the 2,100 households sampled for the beneficiary sample, 112 were excluded from the sample because respondents were deceased or disabled, they had moved out of state, they had moved out of the study area, or they were duplicate cases. This reduced the sample to 1,988 potential respondents. A total of 1,316 of these households completed interviews. A total of 263 households refused to participate in the study, for a refusal rate of 13%. We were unable to complete interviews with 408 households because of language problems, moves for which we had no new address, cases for which we had post office box addresses only, households that did not answer our letters or telephone calls, and cases where we had a telephone contact but were unable to complete an interview prior to the cutoff date for the end of data collection.

Of the 1,543 housing units selected for the At Risk sample, 189 were excluded from the sample because respondents were deceased or disabled, houses were currently vacant with no one living there, or respondents had already been selected in the Beneficiary sample (duplicates). This reduced the sample to 1,354 potential respondents. A total of 798 of these households completed interviews. A total of 341 households refused to participate in the study, for a refusal rate of 25%. We were unable to complete interviews with 217 households because of language problems, households that did not answer our letters or telephone calls, and cases

Sampling Households "At Risk" of Public Assistance

where we had a telephone contact but were unable to complete an interview prior to the cutoff date for the end of data collection.

Analysis of non-respondents: For the beneficiary sample it was possible to conduct an analysis of non-respondents by comparing the demographic characteristics of respondents and non-respondents. This was done to ensure that nonrespondents were not systematically different from respondents and to ensure that the final sample of respondents adequately represents the population of AFDC cases. To accomplish this we used information from the March 1988 warrant roll tape and compared the characteristics of the entire population of AFDC recipients with the characteristics of AFDC recipients who had completed interviews for the study, and with the characteristics of AFDC recipients with whom we were not able to conduct interviews.

Table 2 presents the results of this analysis for three characteristics, gender, race, and number of persons in the household. Comparing the percentages for respondents and the total AFDC population, it is apparent that there are no substantial differences between these characteristics in the sample of interviewed respondents and the total population of AFDC cases. The few differences that do exist can be attributed to sampling error. This lack of substantial differences suggests that the sampling plan adequately portrays the AFDC population for the state of Washington.

Comparing the percentages for respondents and non-respondents, it is also apparent that there are few differences among these characteristics. In general, non-respondents are not significantly different from respondents. This suggests that interviewers were successful at reaching respondents, and that non-respondents are very much like our respondents.

The greatest difference between study respondents and both the AFDC population and the non-respondents is in the percentage of Black AFDC recipients. Among AFDC recipients, approximately nine percent are Black, whereas among respondents, only six percent are Black.

Sampling Households "At Risk" of Public Assistance

This difference is probably attributable to greater difficulty in completing interviews in CSO areas with greater Black populations.

Area Segment Beneficiaries

For purposes of weighting the at risk sample, we needed to estimate the number of welfare beneficiary cases living in each of the 100 area segments selected for the at risk sample. To accomplish this required a manual examination of the household listings for each area segment in comparison to a sorted listing of welfare beneficiary addresses obtained from the welfare warrant roll tape. This matching of street addresses was a very tedious process, prone to errors, and required considerable time to complete.

However, this procedure was used to estimate the number of welfare beneficiary cases residing in each area segment. We know that these estimates are underestimates of the actual number of welfare cases. Many addresses on the computer tape, particularly in rural areas, consisted of post office box addresses. Since the household listings contained street addresses, it was impossible to match post office box addresses with area segments. This turned out to be a major problem for most of the enumeration district area segments in the rural areas of the state. For area segments in these rural areas, it was virtually impossible to determine the number of welfare beneficiaries residing within the area segment boundaries. Sometimes, it was also not possible to specify the extent to which zip code boundaries overlapped into other area segments. Despite these problems, these estimates are the only ones available, and were therefore used for developing weights for the at risk sample.

We examined the composition of the "at risk" sample to determine how representative this sample was of households "at risk" of welfare dependency.

Over 100 households, or 13%, of the at risk sample were on AFDC at the time of the interview. Another 12% of households in this sample were considered "poor", as their total household income was at or below 125% of the federal poverty standard for their household size. These

Sampling Households "At Risk" of Public Assistance

two groups, comprising one-fourth of the total "at risk" sample are considered most at risk of welfare dependency. The other three groups of households in the at risk sample are at lesser risk of welfare dependency, including 30% with household incomes over 125% of the federal poverty standard, 10% male only households, and 35% households with females over child-bearing age and no children living at home.

Since the "at risk" sample was designed to select households with greater risk of welfare dependency, the sample as a whole should include a greater percentage of welfare beneficiaries than the average proportion of welfare beneficiaries in the state. We therefore compared the proportion of the sample that was on AFDC at the time of the interview, with the proportion of households in the state that were on AFDC in March of 1988. Four and a half percent of the state's households were receiving AFDC benefits in March of 1988, compared with over 13% of households in the "at risk" sample. Thus, the sample design was successful in selecting for households with a greater risk of welfare dependency.

We also examined the distribution of welfare beneficiaries in the 100 area segments, and correlated this against the number of "at risk" households that included a welfare beneficiary. If the "at risk" sample design selects households in proportion to density of welfare beneficiaries in area segment neighborhoods, then this correlation should be positive and significant. We determined the distribution of area segments by the number of welfare beneficiaries residing in area segments. While the distribution appears somewhat normal, it is positively skewed. The average number of welfare beneficiaries in area segments in the "at risk" sample is 13 households. In addition, because of the problems that were encountered in matching PO Box addresses to street addresses in area segments, the estimates at the lower end of the distribution, for one to ten AFDC households, are probably unreliably high. Thus, the true average number of welfare beneficiaries per area segment should most likely be higher.

Sampling Households "At Risk" of Public Assistance

We calculated the percent of area segments with zero to four or more households in the "at risk" sample that were receiving AFDC benefits at the time of the interview. Almost half of the area segments in the sample, did not have any households with welfare beneficiaries. And, somewhat over half of area segments had one or more households receiving welfare benefits at the time of the interview. The average number of "at risk" households receiving welfare benefits at the time of the interview was one per area segment.

We computed the correlation between the number of welfare beneficiaries residing within each area segment, as determined by the matching process, and the number of welfare beneficiaries actually in the "at risk" sample in each area segment. This correlation, as predicted, was significant and in the positive direction; $r = 0.51$, which suggests that the "at risk" sample design was allocating households in proportion to their risk of welfare dependency.

Conclusion

This paper has described a sample design that disproportionately allocates a sample of households that are "at risk" of welfare dependency, which was of primary interest to this study for policy reasons. The paper describes and evaluates a unique technique for obtaining a general population sample, with probability proportional to being "at risk" of public assistance. This general sample design has applications to other situations, where indicator data is available, and where a concept of being "at risk" of something makes sense.

The concept of a differing "measure of size" or "measure of importance" for the observational units in a survey sample is extremely important but is difficult to understand. Kish (1965) and Cochran (chp 9A.3, 1977) provide a useful discussion of this statistical and analytical concept which underlies the sample design for the "at risk" general population sample.

The concept of being "at risk" of welfare dependency has been a useful one for the Family Income Study because it has provided a sample of households not on welfare from the

Sampling Households "At Risk" of Public Assistance

general population of Washington state that can be compared to a sample of households on welfare. This comparison sample of "at risk" households has more of the characteristics of the welfare sample, than a simple random sample of the state's population would have. This should make it easier to study movements on and off of welfare in these samples over the five years of this longitudinal study.

This study has also shown that the concept of being "at risk" of welfare dependency can be quantified in probability terms useful for the development of a sample design that is representative of a state's population. Policy makers are frequently interested in estimating the number of individuals likely to become welfare beneficiaries. This technique could be extended to help make these kinds of estimates and provide bounds on the degree of accuracy of the estimates.

References

Cochran, W. G. (1977) Sampling Techniques. (3rd Ed.) New York: Wiley.

Dillman, D. A. (1978) Mail and Telephone Surveys: The Total Design Method. New York: Wiley.

Heeringa, S. G. (1988) The Washington State Family Independence Study (FIS) Sample Design Description. June.

Keyfitz, N. (1951) Sampling with probability proportionate to size: adjustment for changes in the probabilities. Journal of the American Statistical Association. pp. 105-109.

Sampling Households "At Risk" of Public Assistance

Kish, L. (1965) Survey Sampling. New York: Wiley.

Kish, L. (1965) Sampling groups and organizations of unequal size. American Sociological Review. Vol. 30, pp. 564-572.

Sudman, S. (1976) Applied Sampling. San Diego: Academic Press.

Weeks, G. C., Gecas, V., Lidman, R. M., Seff, M., Stromsdorfer, E. W., & Tarnai, J. (1990) Washington State's Family Income Study: Results from the First Year. Unpublished report of the Washington State Institute for Public Policy, The Evergreen State College, Olympia, Washington. (January).

Sampling Households "At Risk" of Public Assistance

Table 1

Comparison of AFDC Respondents, AFDC Non-Respondents, and
the AFDC Population on Gender, Race, and Household Composition

	AFDC Non-Respondents	AFDC Respondents	AFDC Population
Gender			
Male	21.8	19.5	19.4
Female	78.2	80.5	80.6
Race			
Non-minority	74.5	78.9	74.2
Black	9.6	6.2	9.9
Asian	1.6	1.4	1.5
American Indian	5.7	4.4	4.9
Hispanic	4.9	5.4	5.8
Other	3.7	3.8	3.8
Persons in Household			
One	6.8	7.4	8.0
Two	34.3	30.7	31.5
Three	25.2	29.6	26.7
Four-Five	25.2	24.4	25.0
Six or More	8.4	8.0	8.8
Total Number of Cases	740	1,316	81,507