Sample Weights for the Puget Sound Transportation Panel Using Stratification Anchors in the Public Use Microdata Sample and Probabilistic Models for Self-Selection

Jun Ma

Ph.D. Candidate
Department of Civil and Environmental Engineering
and
The Pennsylvania Transportation Institute
131 Research Office Building
The Pennsylvania State University
University Park, PA 16802
Voice: (814) 863-1897
Fax: (814) *865-3039*

and

Konstadinos G. Goulias

Assistant Professor of Civil Engineering
Department of Civil and Environmental Engineering
and
Research Associate
The Pennsylvania Transportation Institute
209 Research Office Building
The Pennsylvania State University
University Park, PA 16802
Voice: (814) 863-7926
Fax: (814) 865-3039
E-mail: Goulias@ecl.psu.edu

Paper Submitted for Presentation at the 75th Transportation Research Board Meeting, January 7-11, 1996, and Publication in the Transportation Research Record

ABSTRACT

A general purpose transportation panel survey contains sampling stages such as initial response (e.g., recruitment of individuals to answer general questions and subsequent or contemporaneous recruitment for the travel diary portion) and panel participation (e.g., continuous enticement to participate in the repeated portion of the survey). In an attempt to contain costs, sample stratification is used as a tool to decrease the number of recruits needed. In addition, panel participation depends on the respondents' characteristics (e.g., the panel is more likely to loose people with a given set of characteristics). If the panel data are to be used as a tracking device of the population under study one needs to perform adjustments so that the resulting dataset represents the population. This adjustment is done using weights that are either based on observed frequencies of a given set of variables, when benchmark -datasets exist such as the CENSUS and there are anchor variables in the two datasets, and/or based on probabilistic models of survey participation.

The Puget Sound Transportation Panel (PSTP) offers a unique opportunity to assess the performance of these methods at each weighting stage when a benchmark dataset is also available. The weighting scheme here contains sequential weighting stages for exogenous and endogenous sampling stratification, missing data, pre-wave self-selection, and multi-wave panel attrition in the PSTP. To derive weights, instead of using marginal distributions from CENSUS 1990, multivariate frequencies from the *5%* Public Use Microdata Sample (PUMS) are used. Weights are first created to correct for exogenous stratification by county of residence. Then, weights are created to adjust the initial Random Digit Dialing sample (RDD) to match PUMS. Next, a pre-wave self-selection weight in the two distinct samples in PSTP (RDD and choice-based transit samples) is applied. After this, a weight that corrects for the bias in the choice-based stratification is also applied. At the end, a weight for multiwave panel attrition is created using an ordered probit model and used to adjust the multi-wave data. At all these stages comparisons are made among the weighted PSTP data, un-weighted PSTP data, and PUMS. In general, comparisons between weighted samples at several stages show the PUMS-based weighting scheme performs as expected very well. Weights based on models of self-selection do not perform equally well. This is due to sampling and non-sampling errors, possible incorrect model selection, and limited number of explanatory variables available.

I. INTRODUCTION

PSTP was designed as a population tracking device. One of its objectives is to provide unbiased measures of changes occurring during the time span considered (in this case 1989 to 1993). For example, to provide descriptive statistics on demographic changes and concomitant changes in mode choice during the period 1989 and 1993 in the four counties covered by PSTP. When samples are collected, sampling error (i.e., the deviation between sample characteristics and population characteristics) is always present. The extent and effect of this "error" has been examined using diagnostic tests applied to the joint frequencies of key variables in PSTP and a CENSUS based sample (i.e., PUMS *5%* sample). Since PUMS is a large sample of the Puget Sound population, joint frequencies for common variables between PUMS and PSTP can be created and compared. In addition, when no other information is available regarding sample selection issues and the PSTP frequencies are not coinciding with the PUMS frequencies one can use PUMS as the benchmark and the weight creation "base" information.

The method here reflects the findings of a preliminary diagnostic analysis. These are:

1. The initial RDD sample is not entirely representative of the Puget Sound Region.
2. The transit sample recruits need different treatment than the RDD sample.
3. Systematic self-selection/missing information appears in all stages of the survey and they are:
   a) Initial RDD
   d) Refusal to provide income information
   c) Refusal to participate in the travel diary phase
   d) Failure to return travel diaries
   e) Panel attrition
4. Exogenous (county of residence) and endogenous (mode choice-based) stratification.

Before proceeding further a note of clarification is needed. The PSTP data scrutiny has been unprecedented in transportation. This has resulted in diagnosing a variety of data problems and related solutions that may give an impression of poor data quality in PSTP. The same type of effort in searching for data problems is not taking place when one deals with the usual transportation surveys and has to work within tighter time constraints. It should be noted, however, that the data quality in PSTP is substantially superior to many survey data encountered in practice.

II. PSTP SAMPLE COMPOSITION

The PSTP sampling phases also define the sample composition. The initial contacts are 4867 households recruited by random digit dialing (RDD) and 308 households recruited from transit samples in the region. Of the 4867 RDDs, 2636 respondents agreed to participate in the travel diary portion of the survey (called Initial Respondents-in this paper). The remaining 2231 RDDs were systematically different from the initial respondents. From among the initial respondents 1545 households returned their travel diaries and they were also systematically different from the initial respondents. From among the 308 transit households, 167 returned their diaries by systematic self-selection. Since systematic self-selection took place even before the true panel survey (i.e., repeated contact over time) commenced, some form of data adjustment is needed. This is a phenomenon that has not received sufficient attention in all past analyses of PSTP data that focused on the exogenous and endogenous stratification and attrition between Waves 1 and 2. Figure 1 provides a time sequenced depiction of the sample composition. The figure does not show the stratification by county.

Since PSTP is a stratified sample, in which the population is classified into several categories based on some predetermined or dependent variable, and within each population category, observations are drawn randomly, weights are needed to account for stratification (1). If the definition of categories of a given stratification variable is predetermined, the sample is called an exogenous stratified sample (county of residence here). If the definition of categories for the stratification variable (here part of the sample is recruited from transit using households and we would like to explain mode choice using the panel data) is dependent on other predetermined variables, the sample is an endogenous stratified sample. In transportation studies, gender, age, etc., are usually considered as predetermined and thus samples that are stratified based on any of these variables are considered exogenous stratified samples. Sometimes, however, it is of interest to study not only individuals' travel behavior from a major population segment such as the single-occupant-vehicle group (SOV) but also rare populations, which form a very small fraction of a population such as transit users. The classification of the rare population is based on the outcome of individuals' predetermined characteristics. To have rare population segments that are large enough and draw valid statistical inferences, we could increase the sample size so that the required number of rare population segments is satisfied. Increasing sample size, however, could be very costly. As an alternative, we can randomly recruit individuals directly from the rare population segment instead of the whole population. Suppose, for instance, we want to study the differences among different travel mode users. As mode choice is often the outcome of individuals' socioeconomic and traffic network characteristics, randomly recruiting individuals among transit users (e.g., by on-board bus survey) would take much less effort to obtain the same number of transit users than from the whole population. Consequently, the resulting sample is an endogenous stratified sample. This is convenient and decreases survey costs but it may result in complex weighting schemes that may not replicate the population structure of characteristics the sample is attempting to represent.

When a population is stratified several times, each according to a different variable (the number of strata based upon each variable can vary), and a fraction of the sample is drawn from each of several stratifications, a pooled sample consisting of all the fractions is called a multi-stratified sample. The PSTP is a multi-stratified sample. The stratification is based on household residential geographic locations and typical travel modes. Initially, telephone RDD survey technique was employed to recruit households from each of the four counties in the Puget Sound region. This was done by specifying the prefix of telephone numbers that differentiate the counties (this resulted in the 4867 RDDs). As transit users in the region are very few, compared to the SOV users, endogenous stratification sampling was used to “enrich” the PSTP sample by adding 308 transit user households. These were recruited primarily by sending letters to passengers on selected bus lines and contacting participants from previous transit surveys. These two sample components RDDs and Transit make up the 5175 households (Initial Contacts).

When the 5175 households were contacted, they were asked questions regarding their sociodemographic information such as household income, household size, the number of children, the number of vehicles in the household, and so on. They were also asked to participate in the panel survey that contains two-day travel diaries. Of the total 5175, 2944 initial contacts (called initial respondents) agreed to be panel members while the rest (called initial non-respondents) refused and thus lost contact since. The 2944 initial respondents received then two-day travel diaries and were required to fill out the diary at an assigned two-day period. From among the 2944 initial respondents, 1712 returned complete travel diaries, which made up the wave 1 panel (called wave 1 participants). Those that did not return complete travel diaries or return incomplete information is called the <u>wave 1 non-participants</u>.

A year later, the wave 1 participants were contacted and requested to fill out wave 2 travel diaries. From the 1712 wave 1 participants, 260 were lost due to panel attrition or moving out of the region (these are 15.2% of the sample). It should be kept in mind that out-of-the-region residential relocation and panel attrition cannot be differentiated in PSTP and their individual effects on descriptive statistics, behavioral models, and weight creation will be confounded. The “lost” wave 1 participants are called <u>wave 1 dropouts</u> and the wave 1 participants that returned complete wave 2 data are called <u>wave 1 stayers</u>. At this point an attempt was made to maintain PSTP representative of the population and additional households were recruited (called the <u>wave 2 new recruits</u>) immediately after wave 1 respondents returned their travel diaries. They were selected so that they possessed similar characteristics as the wave 1 dropouts. Among them, 203 newly recruited households returned complete travel diaries (<u>wave 2 refreshments</u>). No information is available on comparisons between the new recruits and the wave 2 refreshment. The wave 1 stayers and wave 2 refreshments, 1655 households, are the wave 2 participants. The wave 2 participants received travel diaries for the wave 3 survey in Fall 1992. Some of them dropped out and additional refreshments were recruited in the same way as that of wave 2. This process continues until today.

III. PSTP SAMPLE WEIGHTS ALGORITHM

At each stage of sampling in PSTP something happens that subtracts from its ability to represent the population. One way to counteract this process is to devise weights for each stage. The main sources of “bias” are due to stratification, pre-wave 1 self-selection, and attrition.

Stratification

Stratification based on county of residence and mode used introduced systematic bias in the sample that consists of the initial contacts. The former is exogenous stratification while the latter is endogenous stratification. Households in a county were selected disproportionally to the population and thus a county is either over- or under-represented in the sample. In addition, transit users were over represented in the sample since additional transit users from other sources were added to the sample. Statistical inferences based on this multi-stratified sample would be biased toward the over-represented population segments.

Pre-Wave Self-Selection

Prior to the wave 1 data collection, households that denied to participate in panel survey and those that agreed to participate but failed to return complete wave 1 travel diaries may share common observed, unobserved, and unobservable characteristics. Earlier empirical analysis (2) revealed that there were noticeable differences between the groups that agreed and denied to receive travel diaries. For instance, it was found that those who accepted travel diaries were younger, had more young children, lived for short time in their county, and used bus more often. Without consideration for the differences between these two groups (participants and non-participants), statistical analyses would be invalid because of systematically missing observations.

Panel Attrition & Refreshment

Households that returned complete travel diaries but failed to do so in the subsequent surveys may also share some common characteristic (resulting in systematic attrition). Although additional households with similar sociodemographic characteristics were recruited to resemble the dropout households, the representativeness of the replacement sample is unknown. In addition, the replacement households add another dimension of potential non-randomness to the sample as well.

To obtain valid statistical inferences, the three sources of non-randomness in the sample evolution need to be taken into account in every data analysis (e.g., models need to include an array of variables that reflect self-selection and simple descriptive statistics need to be reported by specific categories with the proper cautionary remarks). This is cumbersome. A more appropriate approach is to adjust the sample composition such that it represents the true population. This approach first requires a BENCHMARK sample of the true population against which the PSTP can be adjusted and second a model system to reshape the distribution of the PSTP to consider the missing or partially missing observations.

IV. ALGORITHM COMPONENTS

The weighting scheme devised here reflects our-recent findings on PSTP sample composition and past studies on panel attrition. Figure 2 presents a flow chart of this new weighting scheme. PUMS data from the CENSUS 19-90 are used as the benchmark sample. First, weights are derived to account for county of residence stratification. These weights are based on the county frequencies in the CENSUS 1990. Then, income is imputed for a small portion of the sample based on number of children and employment composition in the household. This was deemed necessary because income is missing in PSTP in a systematic way. Income imputation is followed by a "loop" structure (Figure 2) in which the joint distribution in PUMS (X_{ijk}), based on household income, household size, and car ownership, is used as benchmark to adjust with weights that transform the PSTP counterpart distribution (Y_{ijk}) into one that resembles PUMS (the weights are simply $w_{ijk} = X_{ijk} \, I \, Y_{ijk}$). The loop is needed to identify the optimal number of categories to use for the three variables in the joint frequencies based on which weights are created. Once an acceptable similarity, in a statistical sense, between PUMS and weighted PSTP is achieved (tested using the chi-square on the marginal distributions for the three categorical variables used here) the issue of non-response is addressed. For initial non-response weights are created (using the inverse of probability of selection) for the initial refusal to fill-in the travel diaries and the failure to return diaries for RDDs jointly. The transit sample is also adjusted for failure to return diaries. The subsequent step is similar to the model system derived for the wave 1 to wave 2 data and presented in Pendyala *et al* and Kitamura (1,3) modified to account for multiwave attrition.

V. CREATION OF WEIGHTS FOR THE PSTP

The maintained hypothesis here, for the weighing process, is that Public Use Microdata Samples (PUMS) represent the population (PUMS is treated as if it was the population). The first step is to correct for any bias of the initial RDD contacts. As the initial contacts were recruited by stratifying the population according to the county of residence, after applying the weight of the county of residence, the initial RDD sample should resemble the distribution of the PUMS. If this is not satisfied, the weighted initial RDD sample needs to be further adjusted. Once this is done coefficients of the double self-selection model are estimated and related probabilities of wave 1 RDD participants are computed (this is done by specifying a bivariate probit model with correlated choices). Comparison between the distributions of the weighted sample and the PUMS serves as a criterion of choosing one weight over others. In parallel to the initial RDDs, another portion of the initial contacts is the initial transit group. The initial transits were drawn through some special recruitment method, which may not necessarily result in a random sample. Prior to the first wave data collection, unlike the initial RDDs, initial transits face one choice alone, i.e., to return complete travel diaries or not. Based on the same assumption of the initial RDDs, i.e., mode choice decisions and the self-selection (i.e., returning travel diaries or not) are independent, the self-selection can be modeled as a univariate probit model. The inverse of the selection probability is the weight that can be used to correct for biases resulting from households' self-selectivity. After applying the above weights to initial RDDs and initial transits separately, those with complete travel information selected through RDD and from special transit groups were pooled together to form wave 1 participants.

Since the sample of the initial transits was stratified on the travel mode, the pooled sample of the-wave 1 participants is an overlapping stratified sample. The third step in weighting attempts to adjust the composition of the pooled sample to counterbalance the overcounting of the transit users. It should be noted that ideally the weight for mode 'choice should have been applied to the sample prior to self-selection. Due to unavailability of mode information of the initial contacts, this was postponed to the stage after the initial self-selection. The multiwave attrition model used here is an ordered probit model and it is applied to the pooled wave 1 participants. Statistical inference regarding each subsequent wave can be drawn by properly applying weights to the corresponding stayer sample.

Based on the methodology proposed in the previous sections derivation of the weights for the various sample components in PSTP are discussed in this section. First the "true population" sample PUMS is briefly described. This is followed by the creation of a weighted sample of initial RDDs to resemble the PUMS. Then, the weights to account for self-selectivity are presented. A discussion on weights for choice-based stratification is provided next. Finally, the ordered probit model used to create weights for panel attrition is illustrated.

V.1 PUMS as Population

In this study, the 5% Public Use Microdata Samples (PUMS) was used as the benchmark sample. PUMS is from the "long form" of the 1990 Census questionnaire. It is primarily based on individual county, groups of counties, and places. It contains two distinct datasets, containing household and person records, including the full range of population and housing information collected in 1990 Census. A household and its household members (persons) are linked by an assigned unique serial number. Household and person records have separate weights provided by the Census. When the weights are applied, the sample is expanded to the total population. Therefore, it is a sample of microunits that comes the closest to the true population in the Puget Sound region.

A fundamental question arises when attempts to compare PSTP to the PUMS are made. Are the time frames of the two samples comparable since the PUMS was collected in Spring of 1990 whereas the PSTP in the Fall 1989? This difference may cause the two samples to have different population characteristics' distributions. PUMS, however, can be used as benchmark sample for PSTP for three reasons. First, the time span between the Census data collection and PSTP is very short for some sociodemographic variables such as household size, car ownership, employment, etc. Second, the income variable in the Census 1990 is household income in 1989 coinciding with the first wave of PSTP. Third, PUMS is the only large sample available with a sufficient number of anchors, i.e., common variables in both datasets, and the ability to produce multivariate joint frequencies to use as a reference sample. However, with the availability of 1980 Census, one way to adapt to the time difference between Census and other samples would be to "interpolate" or "microsimulate" the population at a given time point and compare the sample at hand (in this case PSTP) to the interpolated year in PUMS. Such an approach was not needed here due to the time closeness between PUMS and PSTP.

V.2 Initial Contacts and County of Residence

The sample of initial contacts is composed of initial RDDs -and initial Transit. For the initial RDDs, which were recruited based on the county of residence, weights that correct for unequal probability of selection (exogenous stratification based on county of residence) are applied prior to any data treatment. The weight is simply the inverse of the ratio of the selection probability in the sample over the probability being drawn from the population. To do this one needs the household proportion in the region residing in each of the four counties in the Puget Sound. In the past the number of households in each county were obtained from the Census projections (Census 1990 data were not available yet). In this paper Census 1990 information is used. These proportions are reported in Table 1 where the number and observed frequency (or probability) of households in the initial RDD sample and in the Puget Sound region obtained from the PUMS are presented. The derived weight is shown in the last column of the table.

Table 1. Household Weight by County of Residence

County	PSTP		PUMS		Weight
	Count	%	Count	%	$[PSTP\%/PUMS\%]^{-1}$
King	1954	0.4015	601960	*0.579*	1.341779
Kitsap	491	0.1009	66920	0.0634	0.603601
Pierce	1148	0.2359	208981	0.201	0.980177
Snohomish	1274	0.2618	161798	0.1556	0.605636
Total	4867	1.0000	1039659	1.0000	

After the weight for county of residence is applied, a weighted initial RDD sample is obtained. If the initial RDD is truly random, the resulting weighted initial RDD sample should be representative of the entire population, which means that the distribution of any population segment in the weighted RDD should be identical to that of the PUMS. This hypothesis is tested by comparing distributions of some representative variables from the weighted initial RDD sample to those from the PUMS. The control variables (or anchors) used are household income, household size (HHSIZE), and the number of vehicles owned by a household (NUMVEH).

In PSTP, an additional complication arises, about 14.6% of households have missing or incomplete income information whereas only 0.06% in the PUMS. - If the missing data are simply unavailable for unknown reasons to the analyst and unrelated to the fact that other households in the sample contain complete information, we can assume that income-is missing randomly. This type of missing data is called the ignorable case (4) and ignoring these households would result in inefficient regression coefficient estimates. However, if missing data are systematically related to the phenomenon being modeled, inference drawn exclusively from the households with complete information would be qualitatively different. This kind of missing data can be due to self-selection (this is the case encountered in Section V.3 and V.4). To diagnose potential non-randomly missing income in PSTP a binary logit regression model was used for PSTP and PUMS by pooling the two datasets. The missing income logit model uses a dichotomous dependent variable with 0 showing missing income. In addition to commonly used demographic variables, a dummy variable that differentiates records from the PSTP and PUMS is also included in the binary response model. T-tests show that the coefficient of the dummy variable is significant at a 5% significance level suggesting that the income is missing in a different way in the PUMS and PSTP datasets. For the PUMS data we assume that missing income occurs randomly. The model estimates show that employment status and the number of children are the major determinants of PSTP missing income.

The comparison of distributions of both samples, is shown in Table 2. Households with very low (less than $7,500) and very high incomes (over $70,000) are under represented in the PSTP. χ^2- tests of the distribution in which missing income is treated as an additional category is also given in Table 2. This phenomenon is very well known in telephone interviews that are unable to capture very low income strata that may not have a telephone and very high income that may not answer the telephone call at all. This hypothesis is tested by including and excluding missing income as a category in Table 2 and repeating the same test for equal frequencies between PSTP and PUMS. Since, the χ^2 contribution is extremely large for the category of missing income, we also provide the χ^2 - tests without considering missing income households in Table 2. This dramatically decreases the χ^2. Note, however, that the χ^2 is still large enough to reject the hypothesis that the weighted initial RDDs and the PUMS are from the same distribution.

Table 2 Comparison of Income Distributions of PUMS and the Weighted Initial RDDs

Category	Income ($)	PUMS		PSTP		X^2 *Contribution*	
		Count	Percentage	Count	Percentage	W/ missing	W/o missing
1	<7,500	77163	7.2	124	2.5	146	103
2	7,500-14,999	109353	10.2	414	8.5	14	0
3	15,000-24,999	179136	16.7	699	14.4	16	0
4	25,000-29,999	90006	8.4	440	9	2	23
5	30,000-34,999	90153	8.4	561	11.5	55	125
6	35,000-49,999	218559	20.4	1090	22.4	9	67
71	50,000-69,999	168960	15.8	520	10.7	80	29
8	~70,000	135288	12.6	310	6.4	151	88
	Missing	612	0.06	709	14.6	82943	N/A
Total		1069230	100.0	4867	100.0	83418	436

V.3 Imputation of Missing Income

To fully use the information available in PSTP we need to impute income. After detecting that missing income occurs in a systematic way, a logit model for missing income was estimated indicating that household employment status and the number of children are the major contributors in determining missing income. A common approach to replace the missing values in income is by using a model to predict income. The predicted values may be obtained from a linear regression or an ordered discrete response model (e.g., an ordered probit) which depends on the employment status and ~he number of children. The coefficients in the regression models may be unbiased and appear to bring a gain in efficiency. However, Greene (4, pp. 275) suggests that "the gain in efficiency from using these fitted values may be illusory**.**" On the other hand, in practice, it is very difficult to generate ordered discrete response models that predict the dependent variable well when the discrete outcomes are more than three. As an alternative, a pseudo "Monte Carlo simulation/replication" is used to impute missing income based on employment status and the number of children (the two determinants of missing income). This requires that the variables of employment status and the number of children have complete information[1].

Suppose there are K income categories, indexed by k. Further assume that there are I employment status categories and J categories for the number of children, indexed by i and j, respectively . Employment status is expressed as a binary variable with 1 indicating employed and 0 unemployed. The number of children is censored when it is more than 3, i.e., the variable has four discrete outcomes: 0 children, 1 child, 2 children, and 3 or more children. Using this notation, for income category k, we can crossclassify the income by employment status and the number of children. Table 3 illustrates such a crossclassification in a general way.

1 In the PSTP, six out of 5175 households do not have employment status. Income of these six households is missing too. In addition, all of the six households are two-adult households without any child and none of the six households agreed to participate in the panel survey. The employment status is imputed in such a way that they are assigned to one of the three employment categories (0, 1, and 2, representing no employed, one employed, and two employed members in a household) proportionally to the sample portions consisting of households that have two adults but without any child. The number of such households are 477, 381, and 848 for employment category 0, 1, and 2, respectively. This is done by randomly drawing a number from a uniform distribution and assigning them to one of the three categories according to the probabilities that equal the proportions of the sample, which are 0.2796, 0.2233, and 0.4971 for employment category 0, 1, and 2, respectively

.**Table 3. A General Notation of Crossclassification by Employment and Children**

Income k		Number of Children				
Employment Status		0	1	2	3+	Total
	0 (Unemployed)	X_{11k}	X_{12k}	X_{13k}	X_{14k}	$X_{1.k}$
	1 (Employed)	X_{21k}	X_{22k}	X_{23k}	X_{24k}	$X_{2.k}$
	Total	$X_{.1k}$	$X_{.2k}$	$X_{.3k}$	$X_{.4k}$	$X_{1.k}$

For each income category, a similar crosstabulation table can be produced. The ratio X_{11k}/X_{11} is an estimate of the probability that a missing income household would belong to the income category k if the household is unemployed and has no child. In general terms, for a missing income household with employment status i and number of children category j, the probability ~ that the household is in income category k is

$$P_{ijk} = \frac{X_{ijk}}{X_{ij}} \qquad i=0,...,I;\ j=0,...,J;\ \mathrm{k}=1,...,\mathrm{K}$$

$$X_{ij} = \sum_{k=1}^{k=8} X_{ijk}$$

where $\mathbf{X}_{ijk}$ = Number of households with complete income information which are in *i*th category of employment status, *j*th category of children, and income category k;

$\mathbf{X}_{ij}$ = Total number of households with complete income information that have employment status i and children category j.

Let $\mathbf{Y}_{ij}$ denote a missing income household that has employment status i and in children category j. $\mathbf{Y}_{ij}$ is assigned to income category k according to the probability $\mathbf{P}_{ijk}$. This is done in the following three steps:

(1) transform probability $\mathbf{P}_{ijk}$ for all k = 1,...,K, to a cumulative distribution function $\mathbf{P}'_{ijk}$ i.e.,

$$P^{/}_{ikj} = \sum_{k=1}^{k} P_{ikj}$$

(2) draw a random number z from a uniform distribution-over interval [0,1];

(3) if $\mathbf{P}'_{ij(k-1)} \leq z <$ then is as signed into income category k.

Once this is applied all observations in PSTP have complete income information. A comparison of frequencies between PSTP and PUMS is presented in Table 4 that contains the contribution from each category of the variables of interest. The total χ^2 of income, including imputed income, is 361.42 (compared with PUMS) with 7 degrees of freedom. This leads to the rejection of the hypothesis that households from the PUMS and the PSTP are from the same income population. Similar χ^2-tests of the frequencies of HIHSIZE and NUMVEH are *596.62* and *767.56,* with degrees of freedom 9 and 7, respectively. These values also lead to the conclusion that HHSIZE and VEHNUM in the PSTP are significantly different from those in the PUMS.

Table 4. Comparison of Distributions of Income, HHSIZE, and NUMVEH

category	Income[2]		HHSIZE		NUMVEH[3]	
	PSTP	PUMS	PSTP	PUMS	PSTP	PUMS
1	3.81	7.2	16.17	26.3	3.7	7.8
2	10.09	10.23	*35.55*	34.2	23.14	31.6
3	16.81	16.76	18.84	16.3	39.96	38.9
4	10.29	8.42	18.22	14.4	19.95	15.2
5	12.76	8.44	7.4	*5.7*	7.97	4.6
6	25.21	20.45	2.53	2	2.81	1.3
7	12.78	15.81	0.76	0.7	1.19	0.4
8	8.22	12.66	0.41	0.2	1.27	0.3
9	-	-	0:06	0:1	-	-
10	-	-	0:06	0:1	-	-
Total χ^2	361.42		596.62		767.56	

V.4 PUMS-Based Weights

Since χ^2 suggest the rejection of the hypothesis that the initial RDD and the PUMS are from the same distribution, we decide to weight the households in the PSTP such that the modified distribution resembles that of the PUMS. The rationale for such weighting is that the resulting sample should be representative of the PUMS in "every" aspect. This requires considering all variables simultaneously. For instance, considering a sample with M variables, indexed by i and each having N_i categories. The sample can be crossclassified into $N_1 \times N_2 \times \ldots \times N_M$

[2] Definition of each category is the same as in Table 2.

[3] Category 1 represents the category without any vehicle. Category 2 denotes .1 vehicle, and so on.

categories. The weights for each category will simply be the ratio of the PUMS cell frequency over that in the PSTP. Households belonging to the same category take the same weight.

Ideally, in order to have a perfect match between PSTP and PUMS all possible variables with all possible categories should be cross-classified in both samples and then proper weights derived. However, this results in a great number of empty cells in both PSTP and PUMS. Empty cells (zero frequency) in the PUMS can be ignored as they simply mean that such categories do not exist in the population. However, cells that are empty in the PSTP but non-empty in the PUMS may be problematic because there are no sample observations in PSTP to replicate and match PUMS. As a matter of fact, there is a very large chance that a cell is empty in the PSTP but not in the PUMS as the PUMS has considerably larger sample size than PSTP. On the other hand, decreasing the number of categories, either reducing the number of variables used in the classification or aggregating the values of the variable, will reduce the number of empty cells. Such a remedy to the problem of empty cells will decrease the accuracy of the resulting weighted PSTP sample due to lack of sufficient dimensions to properly characterize the population features. Selection of control variables and the level of aggregations are determined on the basis of experimental trials.

Two sets of experimental trials are designed. One is to examine the effect of different number of control variables. The other is to investigate the effect of various degrees of aggregation. Restricted by the number of variables in both data sets, we initially select four variables, i.e., income, HHSIZE, NUMVEH, county of residence (COUNTY)[4] as the control variables. Since county of residence has been considered in the stratified sampling, this variable is not as important as the other three variables at this stage. So its effect will examined after the effect of collapsing categories in the other three variables.

The original Income, HHSIZE, and NUMVEH in the PSTP were classified into eight, ten, and eight categories, respectively. When the original categories are used we get 8 x 10 x 8 (640) cells. Of these, 163 cells in PSTP are empty, which leads to 14,340 (1.34%) unused observations in PUMS. If the number of the unused observations is very small (e.g., when the sample size in PUMS is scaled down to the same sample size as the PSTP, the number of unused observations becomes less than 1), there is no problem of encountering empty cells in the PSTP as these empty cells simply indicate that such population segment is too rare to be included in the PSTP. But on the other hand, if this number is large (e.g., larger than 1 when the PUMS is scaled similarly), it implies that population segments are not presented in the PSTP, which would cause the same problems as if they were missing systematically from PSTP. To solve this problem, control variables are aggregated (collapsing their categories). Table 6 provides various degrees of aggregation with the associated χ^2-tests. In the table, the notation a-b-c stands for the combination of Income, HHSIZE, and NUMVEH which are aggregated to a, b, and c categories, respectively. The column named "Aggregation" is the χ^2 of the aggregated categories that are used to create the weight, which accounts for the

[4] Whenever COUNTY is involved, the PSTP is referred to as the original PSTP, i.e., the sample without applying the weight for the county of residence. Similarly, income is imputed based on the original PSTP in this case.

different distribution in the PSTP. After creating the weight, the distributions of the weighted initial RDD are compared to that of the PUMS, using the same variable definitions as those in the original PSTP. These χ^2s are shown in the rows labeled as Original.

From Table *5*, it can be seen that when the χ^2 tests are based on the original category, χ^2 increases dramatically compared with those based on the categories used to create the weight. This contributes to the errors resulting from the aggregation. It is worth to note that increases in the number of categories and decreases in χ^2 do not follow a linear relationship. This is the direct result of the effect of the degrees of aggregation and missing cells in the PSTP. In addition to the effect of the degrees of aggregation, the effect of COUNTY is investigated for the cases *6-5-5* and *7-5-5*, which are labeled as *6-5-5-4* and *7-5-5-4*, respectively (There are four counties in the PSTP). The results of χ^2-tests are shown in the last two columns in Table *5*. Based on the χ^2 tests the *7-5-5* combination is the best performing and it is used here.

Table 5. χ^2-tests for Various Level of Aggregation

		Combinations of Control Variables				
		5-5-4	6-5-5	7-5-5	6-5-5-4	7-5-5-4
Aggregated	INCOME	0.71	0.08	0.16	0.25	2.88
	HHSIZE	0.43	0.07	0.15	0.46	1.28
	NUMVEH	13.19	0.24	*0.59*	1.82	3.95
	COUNTY	N/A	N/A	N/A	0.53	0.46
	Total χ^2	14.33	0.39	0.90	3.06	8.57
Original	INCOME	79.29	79.29	5.84	60.13	7.00
	HHSIZE	20.97	23.87	21.41	20.96	24.43
	NUMVEH	146.04	54.54	48.76	52.07	-
	COUNTY	-	-	6.14	0.53	0.46
	Total χ^2	246.30	157.70	82.15	133.69	86.40

By applying the weight of county of residence, missing income, and the PUMS-Based weight, the initial RDD contacts are corrected for all non-randomness emerging in the sampling process. Initial transit households (recruits that are not RDDs) are not weighed by either county of residence or the PUMS-Based weights because they are a “set-aside” sample.

V.5 Pre-Wave Self-Selectivity

Prior to the data collection of the first wave survey, the initial contacts were given opportunities to choose whether or not to become wave 1 participants. However, initial RDDs and initial transits had quite different selection processes. Initial RDDs were facing two sequential decisions: first receiving travel diaries and second returning complete travel diaries. Initial transits had only one decision to return or not to return complete travel diaries. Due to the difference in this "self-selection" initial RDDs and initial transits are dealt with separately. In this section, weight creation for initial RDDs is first presented, followed by that of initial transits.

As the self-selection of initial RDDs involves two sequential choices, a bivariate probit model is used. Table 6 shows the definitions of the variables used in the models. The model results, including two independent univariate probits and a bivariate probit (for each decision separately), are provided in Table 7. In the first self-selection (i.e., accepting or rejecting to receive travel diaries), it appears that households with a large number of adults and employed household members are more likely to choose to receive travel diaries. Households with more children, less than two cars, and lower income tend to reject receiving travel diaries. In addition, residence and attitudes have significant influence on households' first self-selection. Households that reside in Pierce county seem to be less likely to be involved in panel surveys. However, if households live in their current home for less than five years, they are more likely to participate. Households that perceive good living environment in their residence county appear to be likely to participate in the travel diary portion of the survey. Apparently, when they face serious traffic congestion they show less interest (this is somewhat surprising).

In the second self-selection equation, households that have more children, more than three cars, more male household members, less employed members, and low incomes are more likely not to return complete travel diaries. Households that live in the current residence for less than 1 year, needing cars to go to work or pick up children, are within two blocks to the nearest bus stop, and have at least one student in the households are less likely to fill out complete diaries as well. Only the households with a large number of drivers tend to be more likely to fill out complete travel diaries. It is interesting to note that attitudes do not influence household decisions at this level. This may be due to the fact that the second self-selection is more dependent on the individual members instead of the entire household.

The univariate probit models assume independence between the two decisions (to participate in the travel diary portion and to return the travel diary). The bivariate model attempts to include possible dependency between the two decisions and this is reflected in the correlation coefficient, ρ, between the two participation equations. However, p is not significantly different than zero as the t-test attests. When the model specification is changed, however, the model results indicate ρ to be significantly different from zero, implying a strong correlation between the two self-selection equation and related decision making. The results of the model are presented in Table 8. This model specification excludes the effects of the number of employed household members, the distance to the nearest bus stop, large households, and car requirement for work. Instead, it adds the number of adults and the county of residence. All the variables in this model are significant at a 90% confidence level. However, this model does not fit the data as well as the previous model since the χ^2-test is about 170 for 10 degrees of freedom, comparing to 192 for 12 degrees of freedom in the previous model. Two additional variables in the previous model increase the value of χ^2 by 20, which is a clear indication that the previous model specification is better than that of the second. To evaluate how the weights are effected by the correlation coefficient, both model specifications are included in the following analysis. To differentiate between the two approaches these two model specifications are named as independent univariate and bivariate probit models, respectively.

Table 6. Description of variables

Variable Name	Description
TOTADULT	Number of adults
TOTi_17	Number of children under 17
CAR2	Indicator, 1 = have 2 cars, 0 = otherwise
LIVEGOOD	Indicator, 1 = county livability rate is at least good, 0 = otherwise
YRHM5LES	Indicator, 1= live in current residence less than 5 years, 0 = otherwise
NUM_EMP	Number of employed people in the household
TRAFGOOD	Indicator, 1 = traffic congestion is not a serious problem in the residence area
DIST2B	Indicator, 1 = the nearest bus stop is within two blocks
CAR1	Indicator, 1 = have one car, 0 = otherwise
KITSAP	Indicator, 1 = the household lives in Kitsap county, 0 = otherwise
YRHM1	Indicator, 1 = live in current residence less than one year
DIST1/4	Indicator, 1 = the nearest bus stop is within one-forth mile
MALEN	Number of males who are 15 years or older
CAR REQD	Indicator, 1 = car required at work, 0 = otherwise
CAR_CHLD	Indicator, 1 = car required to pick up children
STUDENT	Indicator, 1 = student, 0 = otherwise
LICENSE	Indicator, 1 = have valid driver license 0 = otherwise

Table 7. Model Estimation Case I (Insignificant Correlations)

	Univariate		Bivariate	
	Coefficient	T-ratio	Coefficient	T-ratio
Constant	*0.40575*	*4.532*	0.41217	*4.561*
TOTADULT	-0.22934	-7.004	-0.23236	-7.024
TOTi_17	0.04391	*2.354*	0.04388	2.368
CARO	-0.45188	-5.383	-0.45564	-5.315
CAR1	-0.17451	-3.687	-0.17756	-3.717
PIERCE	-0.12681	-2.685	-0.12699	-2.696
LIVEGOOD	0.16861	2.871	0.17000	2.923
YRHM5LES	0.18426	4.890	0.17894	4.743
NUMEMP	0.04525	1.745	0.04627	1.796
TRAFGOOD	-0.13788	-3.029	-0.13531	-2.992
LOWINC	-0.30674	-5.437	-0.30634	-5.404
$\chi^2(10)$=229				
Constant	0.77485	8.473	0.88740	*5.562*
TOT117	-0.07267	-1.979	-0.08142	-2.152
CAR3	-0.12443	-1.829	-0.12115	-1.808
YRHM1	-0.15759	-2.406	-0.16388	-2.511
NUM_EMP	-0.13752	-3.459	-0.14049	-3.508
DIST1/4	-0.12301	-2.161	-0.12166	-2.139
MALEN	-0.17049	-3.137	-0.16136	-2.978
CAR_REQD	-0.13420	-2.434	-0.13757	-2.528
CAR_CHL	-0.25581	-3.356	-0.25175	-3.333
STUDENT	-0.17178	-2.678	-0.16951	-2.678
LICENSE	0.12844	2.490	0.13371	2.618
LOWINC	-0.40821	-4.968	-0.35149	-3.140
LARGEHH	-0.23026	-1.876	-0.21051	-1.688
$\rho(\varepsilon_1,\varepsilon_2)$	-	-	-0.18902	-0.792
$\chi^2(12)$=192			LL= -5389.206	

$\rho(\varepsilon_1,\varepsilon_2)$ = Correlation between the error terms

LL = Log-likelihood function at convergence

χ^2 (12) = Measure of goodness-of-fit measure

Table 8. Model Estimation Case II (Significant Correlations)

	Univariate		Bivariate	
	Coefficient	T-ratio	Coefficient	T-ratio
Constant	0.40575	4.532	0.41504	4.738
TOTADULT	-0.22934	-7.004	-0.24148	-7.428
TOT117	0.04391	2.354	0.04277	2.317
CARO	-0.45188	-5.383	-0.42193	-5.178
CAR1	-0.17451	-3.687	-0.17413	-3.754
PIERCE	-0.12681	*-2.685*	-0.11195	-2.479
LIVEGOOD	0.16861	2.871	0.14445	2.605
YRHMSLES	0.18426	4.890	0.18648	5.079
NUM_EMP	0.04525	1.745	0.06614	2.646
TRAFGOOD	-0.13788	-3.029	-0.13370	-3.077
LQWINC	-0.30674	-5.437	-0.30203	-5.360
χ^2 (10)=229				
Constant	0.71073	8.473	0.31844	1.951
TOTADULT	-0.16148	-2.232	-0.23458	-3.355
TOT117	-0.11131	-4.097	-0.08423	-2.937
CAR2	0.11191	2.118	0.13622	2.861
SNOHOMISH	-0.12326	-1.780	-0.10146	-1.617
YRHM1	-0.16844	-2.593	-0.11459	-1.808
MALEN	-0.15785	-2.731	-0.13496	-2.620
CAR_CHL	-0.30463	-4.212	-0.25932	-3.713
STUDENT	-0.20403	-3.208	-0.17874	-3.036
LOWINC	-0.29040	-3.593	-0.42401	-5.270
LICENSEN	0.11317	1761	0.13624	2.312
$\rho(\varepsilon_1,\varepsilon_2)$	-	-	0.58202	3.235
χ^2 (12)=192			LL= -4959	

$\rho(\varepsilon_1,\varepsilon_2)$ = Correlation between the error terms
LL = Log-likelihood function at convergence
χ^2 (12) = Measure of goodness-of-fit measure

The weight that accounts for self-selection for a given household is the reciprocal of the probability to participate in the subsequent stage of the survey. This probability is computed directly from the model estimates by applying the usual probit equation. Applying the weights that correct for county of residence, non-randomness in the RDD, and the self-selection, the wave 1 RDD participants *(1545)* should become representative of the true population. Table 9 provides the results of χ^2 -tests, which test the hypothesis that the wave 1 RDD participants are from the same distribution in PUMS.

Table 9. χ^2-tests of Weighted Wave 1 Participants

		RDD		RDD and Transit		Choice-based weight	
		Univariate	Bivariate	Univariate	Bivariate	Univariate	Bivariate
Aggregated	Income	36.99	35.32	48.61	48.48	64.84	63.72
	HHSIZE	7.82	10.49	6.45	13.88	11.32	16.58
	NUMVEH	5.46	8.66	8.06	5.25	43.84	54.16
	Total	50.27	54.47	63.13	67.61	120.00	134.46
Original	Income	49.95	48.04	60.09	60.20	75.64	73.86
	HHSIZE	56.86	54.00	46.57	54.11	55.46	62.97
	NUMVEH	6.97	9.83	9.38	6.49	46.23	56.48
	County	28.66	22.68	17.95	25.54	7.02	7.59
	Total	142.44	134.55	133.99	146.34	184.36	200.90
Weighted Sample Size		4860	4853	5161	5154	4736	4739
Original Sample Size		4867		5175		5175	

The initial transits face only one decision (between returning complete travel diaries or not) and the usual binary probit model is used for them (model results are shown in Table 10). It is found that households with a large household size, living in King and Snohomish counties, having more employed members and more male members, tend not to return their travel diaries. Households with more drivers and with people having bus-pass are likely to fill out complete travel diaries. It is very clear then that the factors influencing self-selection Within the RDD sample are very different from the transit sample. This conforms with our effort to create weights for initial RDDs and transits separately. The χ^2-tests of weighted wave 1 transits, together with weighted wave 1 RDD, are shown in Table 9.

Table 10. The Binary Probit Model for Initial Transits

	Coefficient	T-ratio
Constant	0.86218	3.511
HHSIZE	-0.27880	-4.732
KING	-0.81842	-4.047
SNOHOMISH	-0.56784	-2.927
NUM_EMP	-0.25039	-2.131
MALEN	-0.24113	-1.541
LICENSEN	0.58739	4.315
BUSPASS	0.21453	1.353
$X^2(10)=229$		

V.6 Choice-based Weight

Choice-based sample stratification is taken into account after considering county of residence, non-randomness in RDD, and self-selection. The weighted RDDs are assumed to represent the true population, based on which the choice-based sample is corrected. The weight for SOV and Carpool is calculated in a similar way as was done for County of residence. A comparisons of distribution of the weighted wave 1 participants and the PUMS are also provided in Table 9. Table 11 provides the choice-based distributions for each mode in the sample and the Wave 1 RDDs as well as the weight used.

Table 11. Weight for Choice-based Sampling Stratification

Travel Mode	Sample %	Wave 1 RDD %	weight
SOV	3339/5163	3329/4861	1.062127
Transit	1244/5163	953/4861	1.062127
Carpool	580/5163	5794861	0.460722[5]
Total	1.00	1.00	

V.7 Multi-wave Panel Attrition

The final stage of sample adjustments is the multi-wave panel attrition, which is modeled by an ordered probit model. The dependent variable is the number of waves that a household

5 The weight is for both sampling strata. It is the inverse of the sum of the ratios of SOV or carpool and the Transit. The ratio of SOV or carpool is the SOV and carpool percentage in the PSTP over that of the PUMS. Similarly, the ratio of Transit is the Transit percentage in the PSTP over that of the PUMS

missed in the panel surveys, i.e., wave 1, 2, and 3 dropouts missed 3, 2, and 1 surveys, respectively, while the four-wave stayers missed none. The variables used in the model are presented in table 12 and the model estimates are shown in Table 13. All of the variables included in the model significantly contribute explaining attrition behavior. Households without any cars, more employed members, and more household members are less likely to continue participating in the panel. The household type is also significantly affecting the process of attrition. Specifically, households with young adults (single and non-single) tend to leave the panel. As found in other early studies, households recruited by RDD appear to drop out of the panel. However, households with a p9sitive attitude toward their living environment are more likely to continue staying in the panel. The incentive method used also affects the attrition behavior.

Table 12. Variable Definition in the Ordered Probit Model

Variable Name	Definition
CAR0	Indicator, 1 = 0 car, 0 = otherwise
NUMEMP	Number of employed members in the household
YRHM5LES	Living in the current residence less than 5 years
LOW	Indicator, 1 = incomes less than $15,000, 0 = otherwise
535	Indicator, 1 = single adult household under 35 years old, 0 = otherwise
M35	Indicator, 1 = multi-adults household under 35 years old, 0 = otherwise
HHSIZE	Number of persons in the household
RDD	Indicator, 1 = recruited by Tele-RDD, 0 = otherwise
LICENSEN	Number of driver licensee in Household
LIVEGOOD	Indicator, 1=household's perception toward the living environment is good in the county of residence; 0=otherwise
INCENT2	Indicator, 1=Incentive method 2; 0=otherwise
MU(1)	Estimate, a cutoff point that separates second and third category
MU(2)	Estimate, a cutoff point that separates third and forth category

Table 13. Model Estimate of the Ordered Probit Model

	Coefficient	T-Ratio
Constant	-0.64178	-6.273
CAR1	0.32397	10.082
NUM_EMP	0.04616	3.607
YRHM5LES	0.36064	20.636
LOW	0.22812	8.722
535	0.52365	12.383
M35	0.36199	9.007
HHSIZE	0.10303	12.290
RDD	0.29123	3.006
LICENSEN	-0.08040	-5.288
LIVEGOOD	-0.19907	-8.053
INCENT2	0.07486	4.094
MU(1)	0.20261	38.583
MU(2)	0.95702	93.588
LL = -5379		

VI. Wave 2 Data Weighting

After all the weights have been applied the weighted stayer sample should be representative of the true population. Similar to what has been done previously, the χ^2-tests are employed to test the hypothesis that the weighted stayer sample and the PUMS are from the same distribution. Four such tests are performed: weighted wave 2 stayers, weighted wave 2 stayers plus wave 2 refreshments, unweighted wave 2 stayers, unweighted wave 2 stayers plus wave 2 refreshments. The test results are provided in Table 14, along with the original and weighted sample sizes.

Although there are dramatic decreases in χ^2 when weights are applied, χ^2-tests reject the hypothesis that weighted samples and the PUMS are from the same distribution. Recalling that a PUMS adjustment was made before the survey participation weights were applied (to adjust the initial RDDs), the results here are a measure of the partial success probabilistic weights have. Indeed, in this case the attrition model succeeded in closing the gap between PUMS and Wave 2 stayers but only by approximately *50%* (compare the chi-square values between weighted and unweighted and compare the results here with the results in Table 9). This is the reason that weights are remedial measures and not a true fix for non-randomness in samples.

The large χ^2 values lead also to the conclusion that the attrition model lacks the ability to truly explain attrition behavior. This may also imply that a more appropriate model selection is needed. Indeed, ordered probit models implicitly assume that all wave 1 stayers (1712 households) have -chosen the outcome prior to wave 2 survey (via the estimated propensity to participate in the survey). A variety of other models are currently estimated to determine if a better solution can be achieved.

Table 14. Comparisons Between Wave 2 Stayers and the PUMS

		Weighted		Unweighted	
		Wave 2 stayers	Wave 2 stayers & refreshments	Wave 2 stayers	Wave 2 stayers & refreshments
Aggregated	Income	140.95	154.09	191.99	214.29
	HHSIZE	29.19	28.49	80.63	92.85
	NUMVEH	30.10	42.41	118.30	148.79
	Total	200.24	224.99	390.92	455.93
Original	Income	159.77	184.18	199.42	239.16
	HHSIZE	134.54	137.61	168.09	190.15
	NUMVEH	35.97	45.58	119.09	155.88
	COUNTY	40.34	61.01	252.79	403.04
	Total	370.62	428.38	739.39	988.23
Weighted Sample Size		7770	8198	1452	1950

VII. CONCLUSIONS

Creation of the weights that account for the bias introduced by sample stratification, pre-wave self-selection, missing data, and multi-wave panel attrition is presented in this paper. In general, the weighting scheme shown here performs fairly well. However, due to sampling and non-sampling errors, possible incorrect model selection, and limited number of explanatory variables, there exist noticeable differences between the weighted PSTP and PUMS data, especially after applying the weight that takes into account the multi-wave' panel attrition. Further improvements may be made using a variety of more complex model systems and weighting mechanisms. The true population distribution remains a hard target to achieve. This is the reason that during survey administration one needs to recover as much participation as possible and to avoid the thorny selective participation in the survey. In addition, in panels households that move out of the region and thus are dropped out of the panel may need to be included so that panel attrition behavior can be differentiated from out-of-the region residential relocation.

REFERENCES

(1) Pendyala, R. M., K. G. Goulias, and R. Kitamura (1993) Developing of Weights for a Choice-Based Panel Survey Sample with Attrition. *Transportation Research A,* Vol. 27A, No.6. pp.477-492.

(2) Murakami E. and W. T. Watterson (1990) Developing a Household Travel Panel Survey for the Puget Sound Region. *Transportation Research Record,* 1285, pp. 40-46.

(3) Kitamura R., R. M. Pendyala , and K. G. Goulias. (1993) Weighting Methods for Choice-Based Panel Correlation and Initial Choice. In Transportation and Traffic Theory. Elsevier Science Publishers, pp.275-294.

(4) Greene, W. H. (1993) Econometric Analysis. Macmillan Publishing Company.

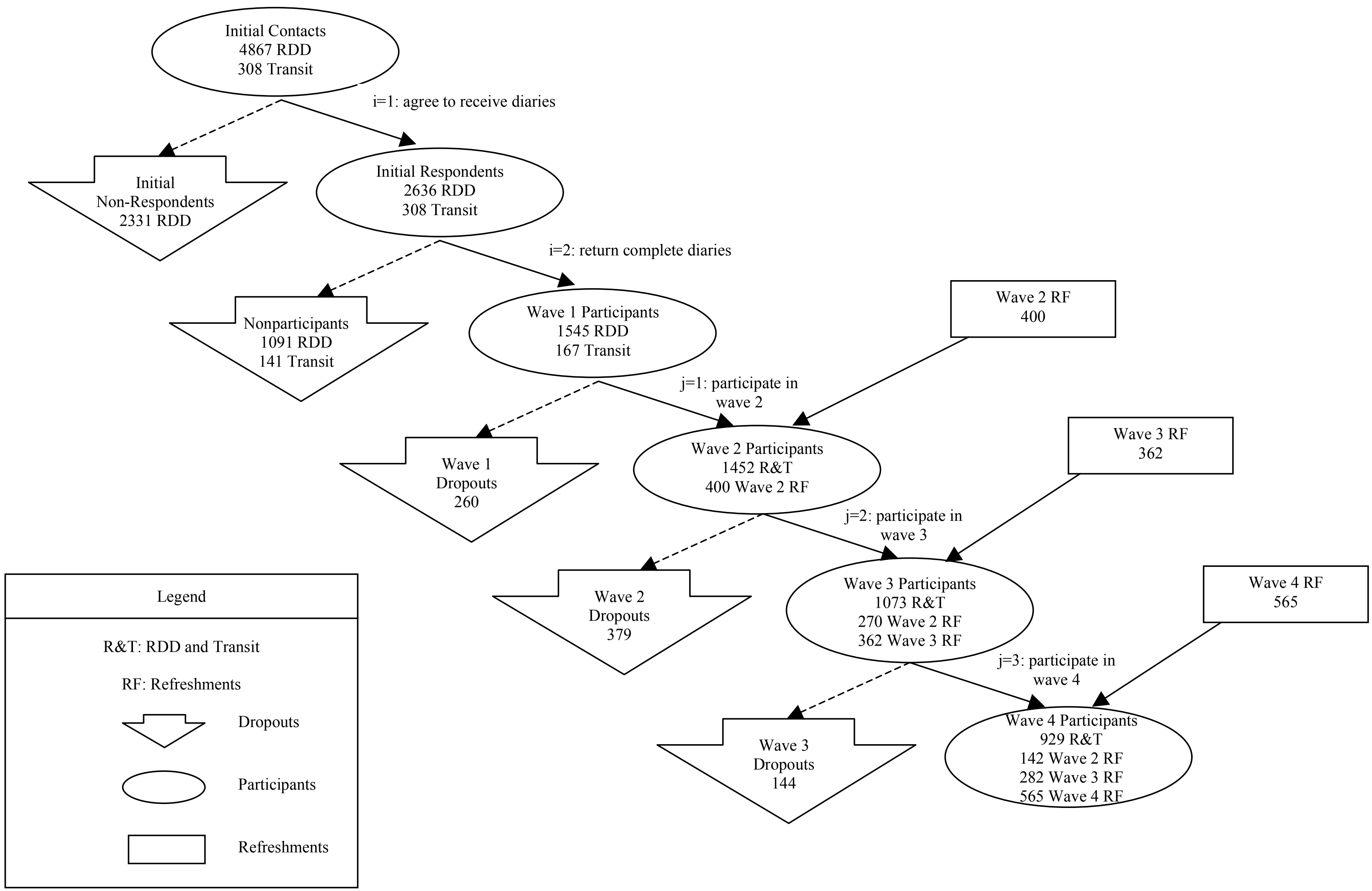

Figure 1. Time sequence of PSTP participation

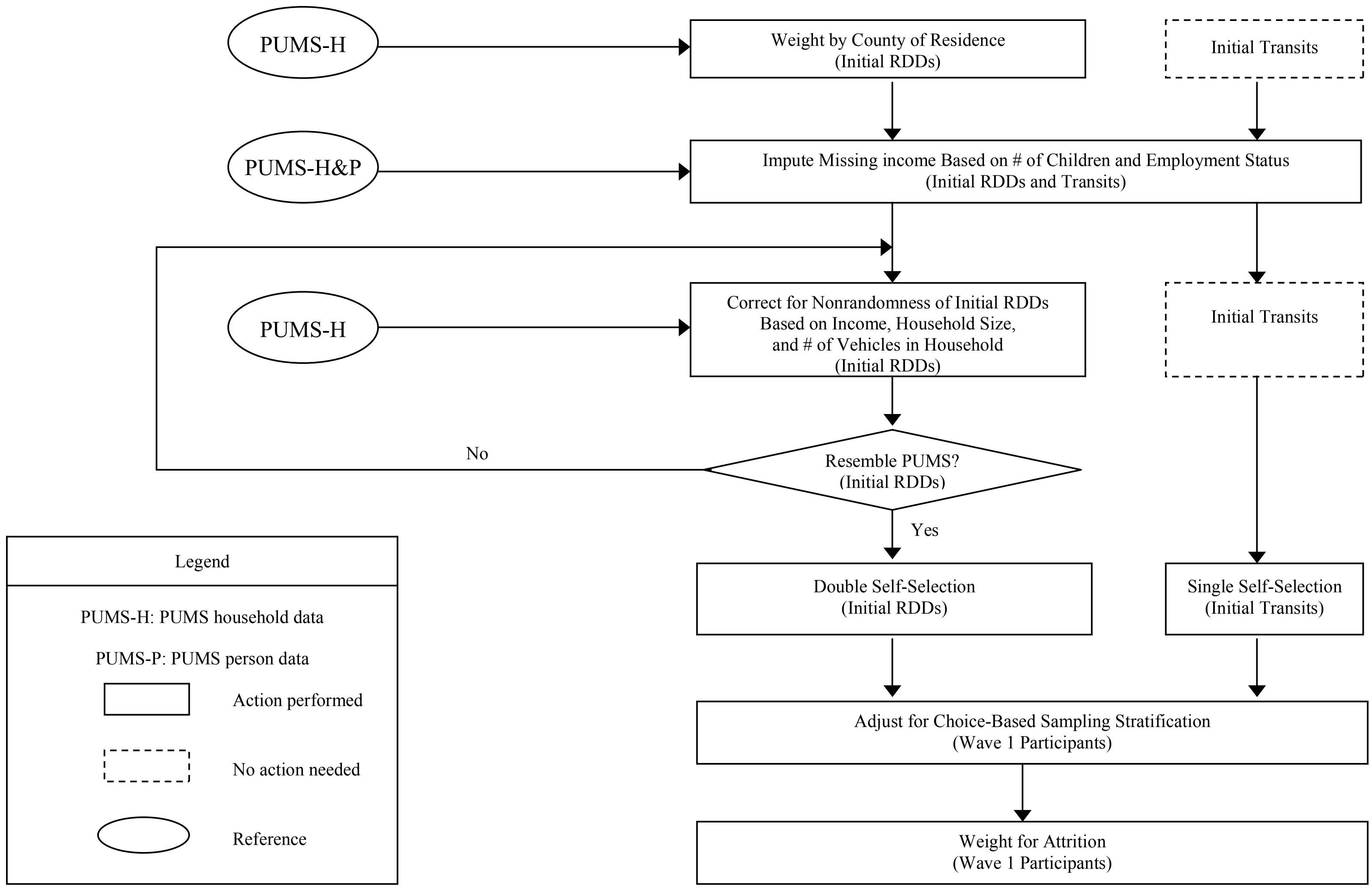

Figure 2. PSTP weighting scheme flowchart