E. Expansion and Re-weighting of the PSRC Household Travel Survey

Report to Puget Sound Regional Council

Expansion and Re-weighting of the PSRC Household Survey Data

January 15, 2007

Mark Bradley Research and Consulting
Cambridge Systematics
MORPACE International

1. Introduction
This memo describes the method and results for expanding the 2006 Household survey data and then re-weighting the data to match Census target data along several key dimensions.

2. Preliminary expansion by sampling region and area type
The first round of expansion only takes account of sampling rates within the five sampling regions (the four counties, with King County split into City of Seattle and Rest of King), and whether or not a household was inside one of the ZIP+2 high transit access areas used for sample enrichment ("oversampling").

The numbers used for this expansion are shown in Table 1. The first section shows the estimated number of actual households located inside and outside the high transit access areas in each region. This was done by using a GIS overlay of ZIP+2 and Census Block boundaries and 2000 Census data to determine the fraction of households inside and outside the areas in each region in 2000, and then factoring up those numbers to match the actual number of households in 2005 from the 2005 American Community Survey (ACS) data—the most recent Census data available.

Next in Table 1, the survey households are broken down into this same 5 region x 2 area type classification, using the same GIS ZIP+2 overlay with the geocoded XY coordinates of each survey household. The numbers are shown separately for households that were contacted via random digit dialing (RDD sample) and households that were contacted via the ZIP+2 are dialing lists for selected high transit access areas. Note that out of the 699 oversample households, 25 are found to be located outside of the transit areas. This indicates that the ZIP+2 phone lists contained some slight inaccuracies, but were very successful overall in finding households in high transit access areas.

The first set of expansion factors in Table 1 were calculated for the RDD sample only, by dividing the first set of numbers by the second set. As one would expect from a randomly distributed sample, the computed factors are very similar inside and outside the transit areas. The

similarity between the numbers in the two rows gives us some confidence in the GIS method used to classify households in and out of the transit areas.

Table 1: Expansion by Sampling Region and Area Type

Sampling area	Seattle	Rest of King	Snohomish	Pierce	Kitsap	Total
2005 Census data						
1- Out of Transit Areas	6917	308692	194788	250733	91674	852805
2- In Transit Areas	254457	175978	59223	32244	2722	524623
Total	261374	484670	254011	282977	94396	1377428
RDD sample HH						
1- Out of Transit Areas	28	752	525	542	576	2423
2- In Transit Areas	915	332	155	65	15	1482
Transit oversample HH						
1- Out of Transit Areas	3	11	9	2	0	25
2- In Transit Areas	443	158	40	31	2	674
Total enriched sample HH						
1- Out of Transit Areas	31	763	534	544	576	2448
2- In Transit Areas	1358	490	195	96	17	2156
Expansion factors for RDD sample only						
1- Out of Transit Areas	247.04	410.49	371.03	462.61	159.16	351.96
2- In Transit Areas	278.10	530.05	382.08	496.06	181.47	354.00
Expansion factors for enriched sample						
1- Out of Transit Areas	223.13	404.58	364.77	460.91	159.16	348.37
2- In Transit Areas	187.38	359.14	303.71	335.88	160.12	243.33

The final rows in Table 1 show the expansion factors for the total enriched sample, calculated by dividing the numbers in the first rows by the number of HH in the total enriched sample. As one would expect, the expansion factors outside of the of the transit areas remain very similar to those for the RDD sample only, while the expansion factors for households inside the transit areas are much lower due to the fact that there are many more such households in the enriched sample.

Note that 142 survey households were not included in the expansion, and are given an expansion weight of 0 for any weighted analyses. The households are:
- 92 households that were recruited at park and ride lots
- 18 households that were recruited on-board ferries
- 32 RDD households whose addresses could not be geocoded

3. Re-weighting the data to match Census-based targets
Household travel surveys typically tend to obtain lower contact rates and/or response rates for certain types of households. The types of households that are typically under-represented in the data are:
- Households that only contain young adults
- Very low income households
- Households that do not own vehicles
- Very large households

To adjust for the possibility of such differences in contact and response rates, the survey households were re-weighted to simultaneously match Census-based target data for:
- Household size distribution
- Household vehicle ownership distribution
- Household number of workers distribution
- Household income distribution
- Household lifecycle distribution

The data used for these targets was provided by PSRC staff, and is shown in Table 2.

The survey households were classified into these same categories. The categorization for the number of persons, vehicles and workers is straightforward. For income, there were about 6% of the survey households whose income is only known within ranges of $50,000 (0-50K, 50K-100K, >100K) and a further 6% whose income is not known at all. For expansion purposes, those households were pseudo-randomly classified into one of the possible income categories based on the final digits in their serial ID number. The classification for lifecycle is based on the number of adults in the household, the age of the oldest adult, and the ages of any children under 18.

Table 3 shows the further percentage adjustment that is needed to match each of these targets when only the first-pass expansion factors are applied. By definition, the targets for total in each region and for inside and outside the transit oversample areas were matched exactly, as those targets were used to calculate the first-pass expansion. The table shows in bold type all cells that need to be adjusted by more than one third (33.3%).

The largest adjustment is need for the young adult households. Both lifecycle categories for young adult households with no children need to be adjusted upward by about 200% (in other words, the expansion factors need to be tripled). The largest underrepresentation is in Kitsap County, but it is true in all counties. Households with children also need to be adjusted upwards outside of King County. Low income households, households with 3+ workers, and households with no vehicles also need to be adjusted upwards in all regions.
Larger households need to be adjusted upwards outside of King County. There are very few cells that need downward adjustment by more than 33%.

Table 2: Targets for (Re)Weighting the Data, Based on 2005 ACS Data

	Seattle	Rest of King	Snohomish	Pierce	Kitsap	Total
Total	261374	484670	254011	282977	94396	1377428
Household size						
1- 1 person	114274	126065	64792	73690	24929	403750
2- 2 persons	84450	170541	88800	97849	32361	474001
3- 3 persons	28529	72929	37909	43138	14770	197275
4- 4+ persons	34121	115135	62510	68300	22336	302402
Car ownership						
1- No vehicles	42623	28966	12350	15728	2965	102632
2- 1 vehicle	110666	147334	70617	85901	29396	443914
3- 2 vehicles	77034	194314	97700	109717	35228	513993
4- 3+ vehicles	31051	114056	73344	71631	26807	316889
Workers in household						
1- No workers	60766	97030	54616	68234	24082	304728
2- 1 worker	111534	207104	95137	116738	36650	567163
3- 2 workers	78187	149909	87515	82368	28947	426926
4- 3+ workers	10887	30627	16743	15637	4717	78611
Household income						
1- Less than $10,000	24357	22120	15355	21851	5643	89326
2- $10,000 - 19,999	29314	39494	22101	26008	9643	126560
3- $20,000 - 29,999	25995	41784	23468	30856	9941	132044
4- $30,000 - 39,999	27606	43188	27757	33176	10011	141738
5- $40,000 - 49,999	24178	49631	29327	31340	8980	143456
6- $50,000 - 59,999	19483	34558	17863	29120	10669	111693
7- $60,000 - 69,999	18136	38336	21878	24061	8211	110622
8- $70,000 - 79,999	13990	30383	16261	18702	5658	84994
9- $80,000 - 89,999	11436	29230	17885	12643	5648	76842
10- $90,000 - 99,999	12002	26100	12997	10396	4609	66104
11- $100,000 or more	54877	129846	49119	44824	15383	294049
Household lifecycle						
1- Young children, 1-5	24874	63165	37787	45018	15882	186726
2- School children, 6-17	28774	104572	51192	59342	17118	260998
3- Young adult, 18-34	32331	22315	9886	14133	4603	83268
4- Mid - adult, 35-64	56661	73810	38283	37629	13558	219941
5- Older- adult, 65 +	25282	29940	16398	21928	6768	100316
6- Young 2+ adult, 18-34	30962	30472	17137	16059	5338	99968
7- Mid 2+ adult, 35-64	42761	108270	59894	59882	20148	290955
8- Older 2+ adult, 65 +	19729	52126	23434	28986	10981	135256

Table 3: Further percentage adjustment needed after first-pass expansion

	Seattle	Rest of King	Snohomish	Pierce	Kitsap	Total
Total	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sampling area						
1- Out of Transit Area	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2- In Transit Area	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Household size						
1- 1 person	6.6%	-7.5%	-16.8%	-10.7%	0.4%	-5.8%
2- 2 persons	0.6%	7.5%	-13.9%	-14.3%	-30.6%	-6.4%
3- 3 persons	-14.0%	-3.5%	**34.1%**	**39.7%**	25.4%	9.8%
4- 4+ persons	-8.1%	0.8%	**39.9%**	23.5%	**100.5%**	15.2%
Car ownership						
1- No vehicles	**89.6%**	**193.1%**	**160.5%**	**89.4%**	9.4%	**113.3%**
2- 1 vehicle	-1.9%	10.0%	-9.1%	2.4%	11.2%	2.1%
3- 2 vehicles	-18.9%	-7.0%	-6.0%	-13.7%	-7.4%	-10.3%
4- 3+ vehicles	0.1%	-13.5%	8.4%	12.5%	-1.5%	-1.4%
Workers in household						
1- No workers	-3.9%	-12.3%	-16.9%	-9.9%	-32.5%	-13.2%
2- 1 worker	2.0%	3.7%	-8.1%	-3.9%	14.0%	0.2%
3- 2 workers	-3.9%	1.7%	19.9%	6.9%	23.7%	6.2%
4- 3+ workers	**44.6%**	13.8%	**42.0%**	**79.4%**	**48.2%**	**35.2%**
Household income						
1- Less than $10,000	**176.6%**	**88.7%**	**145.2%**	**154.3%**	**61.0%**	**129.7%**
2- $10,000 - 19,999	**83.6%**	**104.4%**	33.2%	5.9%	**51.3%**	**52.8%**
3- $20,000 - 29,999	27.3%	**56.6%**	-4.0%	29.4%	**45.3%**	29.2%
4- $30,000 - 39,999	29.9%	**39.5%**	27.6%	10.7%	-3.2%	24.0%
5- $40,000 - 49,999	-3.8%	3.4%	-9.3%	-9.6%	-22.7%	-5.5%
6- $50,000 - 59,999	-23.5%	-33.3%	-26.7%	-16.3%	17.6%	-23.2%
7- $60,000 - 69,999	-18.4%	-29.4%	-7.7%	-7.7%	-9.5%	-18.3%
8- $70,000 - 79,999	-31.9%	-28.1%	-25.7%	**-34.9%**	-27.5%	-29.9%
9- $80,000 - 89,999	**-35.6%**	-6.1%	-2.9%	-13.3%	10.9%	-11.7%
10- $90,000 - 99,999	-23.2%	-29.0%	-33.2%	-24.6%	**-35.6%**	-28.8%
11- $100,000 or more	-19.5%	-1.5%	9.6%	3.3%	-12.1%	-3.8%
Household lifecycle						
1- Young children, 1-5	-15.0%	3.2%	**41.6%**	26.7%	**112.3%**	16.6%
2- School children, 6-17	-14.1%	6.5%	**35.6%**	**36.9%**	**36.1%**	15.8%
3- Young adult, 18-34	**238.3%**	**129.9%**	**212.7%**	**193.2%**	**621.5%**	**198.7%**
4- Mid - adult, 35-64	-17.5%	-12.8%	-21.1%	-31.7%	-11.3%	-19.2%
5- Older- adult, 65 +	-12.7%	-28.7%	**-37.5%**	-2.9%	-24.1%	-22.1%
6- Young 2+ adult, 18-34	**267.2%**	**334.7%**	**147.3%**	**150.4%**	**738.0%**	**239.8%**
7- Mid 2+ adult, 35-64	-23.4%	-4.2%	-16.4%	-17.1%	-30.8%	-14.9%
8- Older 2+ adult, 65 +	-27.5%	-24.4%	-29.1%	-32.5%	**-44.4%**	-29.5%

The procedure for matching all of these targets simultaneously was to use iterative proportional fitting (IPF), also commonly referred to as "Fratar". A program was written to cycle through the six sets of targets in Table 3, adjusting the sample expansion factors to match each target in each region. After 10 iterations through the targets, the summed expansion weights across the sample match all of the targets simultaneously (i.e. all of the cells in Table 3 become zero).

The final adjusted expansion factors wee provided to PSRC as variable EXPFAC2 in SPSS data file HHLDEXP.SAV.