WORKING PAPER
ITS-WP-04-05

Monte Carlo Simulation of Sydney Household Travel Survey Data with Bayesian Updating using Different Local Sample Sizes

By

Graham Pointer, Peter Stopher & Philip Bullock

March, 2004

ISSN 1440-3501

INSTITUTE OF TRANSPORT STUDIES

The Australian Key Centre in Transport Management

The University of Sydney and Monash University

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITS-WP-04-05

TITLE: Monte Carlo Simulation of Sydney Household Travel Survey Data with Bayesian Updating using different local sample sizes

ABSTRACT: There is increasing interest in the potential to simulate household travel survey data as an alternative to collecting large sample household travel surveys, or as a means to augment sample sizes, well beyond what can usually be considered. In prior research on simulating such data, it has been shown that it is possible to reproduce, within reasonable bounds of accuracy, an actual household travel survey, It has also been found that the procedure of updating the distributions of the simulated variables, using Bayesian updating with subjective priors, can provide significant improvement in the accuracy with which an actual household travel survey can be simulated. In work performed to date, it has not been determined what the optimal size would be for the update sample to be used in the Bayesian updating. Rather, prior work has used a sample of approximately 500 households, largely as a matter of convenience and cost. In this paper, we report on further research that compares different sample sizes for the local update data. It was found that a reasonable updating could be obtained from a sample as small as 300 households, chosen through a stratified sampling procedure, and that results improved substantially when the update sample was increased to 500. However, an increase in the sample to 750 did not produce very much additional improvement, suggesting that sample sizes of this size and larger may not be economically justified. At the same time, the research suggests that there may be room for a more targeted sampling procedure which could allow smaller samples to be more cost-effective.

KEY WORDS: ***Household Travel Surveys, Simulation Bayesian Updating, Sample Size.***

AUTHORS: Graham Pointer, Peter R Stopher, and Philip Bullock.

CONTACT: Institute of Transport Studies (Sydney & Monash)
The Australian Key Centre in Transport Management, C37
The University of Sydney NSW 2006, Australia

Telephone: +61 9351 0071
Facsimile: +61 9351 0088
Email: itsinfo@its.usyd.edu.au
Internet: http://www.its.usyd.edu.au

DATE: March 2004

1. Introduction

Household travel surveys (HTSs) are increasingly expensive to undertake *[1]*. Problems with response rates, mobile phones, and the increasing difficulty of conducting CATI surveys make it likely that the household travel survey as we know it will have to change significantly in the future. For modeling purposes, sample sizes of 3,000 households and upwards are required, irrespective of the size of the region, putting an adequate sample size increasingly beyond the reach of many urban areas. There is also an increasing need to gain area specific knowledge within the metropolitan area along transport corridors and within subregions. This degree of information, due to the cost of additional completed samples, is generally not feasible to obtain from HTSs.

In an effort to develop an alternative to the large-scale HTS, recent research *[2], [3], [4], [5], [6], [7], [8]* has developed a method to produce simulated HTS data. This method commences by using a set of distributions of travel characteristics obtained from a nationwide sample, which are updated to a specific locality, using a small local update sample and Bayesian updating with subjective priors. A source of data on local households, with full socio-demographic data is required, from which a sample is selected, much in the same way that a HTS would be sampled. Using the demographic data of each sampled household, a Monte Carlo simulation is performed of specific travel attributes, namely, the number of trips by purpose, the main mode of travel for each trip, the time of departure of each trip, and the trip duration in minutes for each trip.

Distributions of key variables influencing travel behavior were derived by Greaves (1998) from the Nationwide Personal Transportation Survey (NPTS), conducted in the U.S. in 1995. The U.S. distributions were used in the research for Adelaide and now Sydney, because no recent nationwide HTS exists in Australia. These U.S. distributions have already been used in prior research in Australia *[6], [8]* and have been shown to produce reasonable simulated travel data.

The present research uses data from the Sydney HTS, which is widely regarded as a clean, representative sample. The Sydney HTS began in 1997 and is a continuous survey, of about 3,000 households each year *[9]*.

2. Data Preparation

The simulation of socio-travel characteristics requires a large sample of local social-demographic data. The 1996 Household Sample File (HSF) is a one percent sample from the 1996 census household data collected by the Australian Bureau of Statistics *[10]*. For the purposes of this research the HSF was limited to households in Sydney. At the time of this research the 2001 HSF was not available. The HSF for Sydney was then randomly reduced to the region covered by the Sydney HTS, so that the simulation could be compared to that survey. Sampling from the HSF was done using a stratified sampling procedure by household size and number of vehicles available to the household, so as to replicate this joint distribution in the Sydney HTS. The comparison base was the pooled Sydney HTS from 1998, 1999, and 2000. The data were modified so that such items as trip purposes and modes matched the definitions used in the NPTS of 1995, and the distributions developed by Greaves *[1]*. Also, households surveyed on a weekend and part response households were deleted, because both of these were excluded from the distributions developed from the NPTS. The area variables depicted in the HSF did not correspond to the Sydney HTS. Areas included in the Sydney HTS that were deleted were: the Blue Mountains, Newcastle and Wollongong. These are areas not traditionally referred to as part of metropolitan Sydney. The Sydney HTS sample size used for this research was 5,124 households.

For the update samples, three sets of subsamples were drawn from the Sydney HTS, to represent the likely situation that would have existed if stratified random samples of households been drawn and surveyed in Sydney for the purposes of updating. Arbitrary sample sizes of 300, 500 and 750 were used in this research. The samples were again drawn using a stratified random sampling procedure, with the stratification being on household size and number of vehicles, and being designed to replicate the relative frequencies of each cell as found in the full Sydney HTS. Due to rounding, final update sample sizes of 303, 500 and 751 households were drawn.

Summary statistics were compared for the Sydney HTS; the HSF sample used during the simulation procedure; the 303, 500 and 751 household update samples, and, the NPTS data set. The aim of replicating the characteristics of the Sydney HTS in the update samples and the HSF sample was found to have been achieved. Some major differences between the NPTS data set, which provides the basis of the trip characteristics distributions used in the simulation procedure, and the Sydney data sets were evident: the percent of non-car owning households in the NPTS is roughly half that of the Sydney data sets. The NPTS reported approximately 0.3 more vehicles per household than the Sydney data sets.

3. Updating and Simulating Travel Survey Data

It was necessary to recategorize the update samples drawn from the Sydney HTS into the categories set out by Greaves *[1]*, to generate the distributions for the simulation. In line with the research undertaken by Kothuri *[7]* and Stopher and Pointer *[8]*, Bayesian updating with subjective priors was used. This approach updates with a predetermined weight between the original data set and the update data set irrespective of sample size and variance. The aim of this updating approach was to ensure that the update would influence the distributions towards the local data and away from the U.S. NPTS distributions. The standard method of Bayesian updating, shown below, would necessarily reduce substantially the effect of a small update sample of a few hundred observations, when used to update an original parameter derived from thousands of observations.

This approach is outlined in the following equation:

$$q_2 = \frac{\left(\frac{q_1}{s_1^2} + \frac{q_S}{s_S^2}\right)}{\left(\frac{1}{s_1^2} + \frac{1}{s_S^2}\right)}$$

where: q_1 = original statistic,
q_S = local statistic,
q_2 = updated statistic,
s_1 = standard deviation of the original statistic,
s_S = standard deviation of the local statistic

The update sample sizes ranged between 303 and 751 households whereas the NPTS consisted of a sample of 42,033 households *[11]*. To enable the local samples to influence the distributions used in the simulation procedure, weights were applied during the Bayesian updating procedure. The NPTS and local samples were equally weighted. This special case of Bayesian updating with subjective priors can be simplified to the following equation:

$$q_2 = \frac{q_1 + q_S}{2}$$

where: q_1 = original statistic,
q_S = local statistic,
q_2 = updated statistic.

The resulting distributions were transformed into cumulative frequency tables in readiness of the simulation procedure.

It is important to discuss how this research should be judged. It is essential to understand the role the simulation procedure would play in the real world. The job of the simulation procedure is to produce a data set that adequately resembles the travel behavior of the target area. The resulting data must be able to support the estimation and application of travel demand models. The principal area of judgment must, therefore, be on travel data expanded to the size of the population: the real world situation. The focus of this research was on comparing the expanded simulations to the expanded Sydney HTS data set. A comparison of average trip rates by purpose per household was also performed.

Throughout the paper the simulation data set gained from the NPTS distributions is referred to as the original simulation, the simulation from NPTS distributions with Bayesian updating by a 303 household local sample is referred to as the 303 simulation, and similarly the 500 simulation and the 751 simulation.

4. Simulation of Trip Rates

Table 1 compares the household trip rates by purpose between the Sydney HTS data set and the simulations. The table also exhibits the z-test results for equal population means. All simulations compare favorably statistically to the Sydney HTS data set. The most frequent trip purpose per household for all data sets was home other trips. The simulations underestimate the home-based other (HBO) trips undertaken by Sydney households. The 303 and 500 simulations estimate HBO trips better than the original simulation. The improvement in estimating the HTS trip rates from the original simulation to the 303 simulation is marked. The 303 simulation was closer to the Sydney HTS than the original simulation for six of the seven trip purposes while the 500 and 751 simulations were closer for five and four purposes respectively. Compared with the other simulations, the 303 simulation produced the closest overall trip rates to the Sydney HTS for three of the seven purposes, and was equally close to the 500 simulation in the estimation of home based other trips. Compared with the 751 simulation, the 500 simulation was closer to the Sydney HTS for three of the seven trip purposes and both simulations produced the same estimate for home-college trips. Statistical tests were also performed to compare trip rates between the 303, 500 and 751 samples. Overall, it appears that the simulations have been responsive to the different update sample sizes. Three of the seven trip rates (home-work, other-work and other-other) were significantly different between the 303 and 500 simulations. Differences between the 500 and 751 and 300 and 751 simulations were slightly less apparent, with two of the seven trip purposes showing significant differences. Between the 500 and 751 simulations, home-shopping and other-work trips were significantly different, while home-school and other-work trips were significantly different between the 300 and 751 simulations. It is clear that Bayesian updating of NPTS distributions with a local sample produces results that are closer than without updating.

Table 1: Comparison of the Sydney HTS and Simulated Data Household Trip Rates by Purpose

Purpose	Sydney HTS	Simulation (NPTS)	Simulation (NPTS & 303 Update sample)	Simulation (NPTS & 500 Update sample)	Simulation (NPTS & 751 Update sample)
Home-Based Work	1.14	1.58**	1.40**	1.32**	1.38**
Home-Based School	0.59	0.69*	0.60	0.62	0.67**
Home-Based College	0.10	0.19**	0.16**	0.17**	0.17**
Home-Based Shop	1.10	1.31**	1.17*	1.23**	1.14
Home-Based Other	3.94	3.61**	3.69**	3.69**	3.56**
Other-Work	1.15	1.12	0.97**	1.23	1.12
Other-Other	1.90	2.00	1.96	2.13**	2.01
TOTAL TRIPS	9.91	10.49**	9.94	10.39**	10.04

Note: * indicates statistically significant difference in trip rates at the 95 percent confidence level
***** indicates statistically significant difference in trip rates at the 99 percent confidence level***

The experience of Stopher *et al.* *[6]* when simulating travel data for Adelaide was that the simulation invariably produced trip rates that were greater than the equivalent HTS. An explanation given was the higher proportion of households included in the HTS, compared to the simulation, which did not make a trip on the survey day. This was also the case with the original simulation based on Sydney demographic data. The original simulation produced 3 percent of households that did not make a trip on the survey day compared to 8.2 percent of households in the Sydney HTS. After updating, the mobility factors remain low for the updated simulations: between 2.8 percent and 3.2 percent.

Table 2 exhibits those trips taken by mobile households only in Sydney and the simulations. A similar pattern can be found to that reported above concerning all households. Compared with the original simulation, the 303 simulation was closer to the Sydney HTS for five of the seven purposes while the 500 simulation was closer for six purposes and the 751 simulation was closer for four purposes (and at least equally close for two other purposes). Overall the 303 simulation again produced the closest rates to the Sydney HTS out of all the simulations, having the closest rates for four out of seven of the trip types. The mobile household 500 simulation was better than the mobile household 751 simulation for four trip purposes.

Table 2: Comparison of Household Trip Rates for the HTS and Simulations for Mobile Households Only

Purpose	Sydney HTS	Simulation (NPTS)	Simulation (NPTS & 303 Update sample)	Simulation (NPTS & 500 Update sample)	Simulation (NPTS & 751 Update sample)
Home-Based Work	1.24	1.62**	1.45**	1.36**	1.43**
Home-Based School	0.64	0.71*	0.62	0.64	0.69
Home-Based College	0.11	0.20**	0.16**	0.18**	0.17**
Home-Based Shop	1.20	1.35**	1.21	1.27*	1.18
Home-Based Other	4.29	3.72**	3.81**	3.82**	3.68**
Other-Work	1.25	1.15*	1.00**	1.28	1.15*
Other-Other	2.07	2.06	2.03	2.21*	2.08
TOTAL TRIPS	10.80	10.81	10.28**	10.75	10.39**

Note: * indicates statistically significant difference in trip rates at the 95 percent confidence level
***** indicates statistically significant difference in trip rates at the 99 percent confidence level***

Stopher and Pointer *[8]* found that, in Adelaide, the use of U.S. travel characteristics (from the NPTS) in the simulation procedure contributed to higher home-work trips than was actually the case. It was found that through Bayesian updating, the impact of travel behaviors peculiar to the U.S. was lessened. This is also the case in Sydney with respect to average trip rates per household. Travel characteristics peculiar to Sydney have been incorporated through the Bayesian updated simulations for almost all trip purposes.

5. Data Expansion

The simulations were expanded by household size and vehicle ownership (size-vehicle), and mobility factors. As mentioned previously, the simulation procedure does not adequately account for nonmobile households. Therefore, it is necessary to provide separate weights for mobile and nonmobile households during expansion. The mobility factors within each household size and vehicle category of each update sample were used during the expansion of each corresponding simulation. These distributions within the HTS are used throughout the simulation procedure, so that comparison is possible between the HTS and the simulations. The expansion procedure therefore uses these same distributions to expand the simulations and the HTS. The population total was gained from the HSF data. The HSF is a simple one percent sample of census data. The HSF file was multiplied by 100 to gain a population of 1,418,000 households. Obviously, in a real world application, there would not be an HTS to which to compare the simulation data set; the available HSF data would be used during the simulation and expansion procedures. This research is not intended to evaluate and compare the survey methodology of the HSF and the HTS.

It was not appropriate to expand the simulations using solely the U.S. distributions. With the absence of mobility factors with which to expand the original simulations, the only option would have been to expand the data set subject to household size and vehicle representation within the HTS. The previous section showed that the original simulation was outperformed by the Bayesian updated data sets. Therefore, analysis of an expanded original simulation would have achieved little. It has been assumed that the expanded simulations using Bayesian updated travel distributions estimate local travel characteristics better than an expanded simulation using solely NPTS distributions.

The mobility factors differ for each data set and are shown in Table 3. The 751 household sample mobility rates approximate the HTS better than the other sample sizes, as expected. The lower the sample size the more lumpy the data would be.

Table 3: Comparison of Mobility Rates of the HTS and the Update Samples.

Size vehicle groupings		Percent of mobile households			
		Sydney HTS	303 Sample	500 Sample	751 Sample
1-2 Persons, 0 Vehicles	1	76.7%	78.0%	76.5%	78.4%
1 Person, 1+ Vehicles	2	88.2%	89.5%	87.3%	85.3%
2 Persons, 1 Vehicle	3	90.8%	79.6%	88.9%	91.8%
2 Persons, 2+ Vehicles	4	93.3%	91.9%	96.7%	92.3%
3+ Persons, 0 Vehicles	5	96.2%	100.0%	100.0%	93.8%
3 Persons, 1 Vehicle	6	95.8%	90.0%	97.0%	95.9%
3 Persons, 2 Vehicles	7	96.5%	100.0%	100.0%	95.7%
3 Persons, 3+ Vehicles	8	96.0%	100.0%	100.0%	90.9%
4+ Persons, 1 Vehicle	9	96.4%	95.8%	95.0%	96.7%
4 Persons, 2 Vehicles	10	98.3%	100.0%	100.0%	98.4%
4 Persons, 3+ Vehicles	11	99.4%	100.0%	100.0%	100.0%
5+ Persons, 2 Vehicles	12	99.0%	100.0%	100.0%	97.6%
5+ Persons, 3+ Vehicles	13	98.5%	87.5%	100.0%	100.0%
Total		91.8%	90.1%	92.4%	91.5%

The expanded Sydney HTS and simulations were compared on a daily basis for: number of trips by purpose, mode share by trip purpose, departure time by trip purpose, and trip duration by purpose.

6. Comparisons of Numbers of Trips by Purpose

Table 4 shows the number of trips by purpose for the expanded HTS and simulations and the percent difference of each trip purpose for each expanded simulation compared to the expanded HTS. Overall, the 500 simulation is closest to the HTS, with an error on total trips of 1.6 percent. Each of the 303 simulation and the 500 simulation produce the best estimates for three trip purposes, while the 751 simulation is best on only one purpose – other-other trips. The 500 simulation is best on home-work, home-other and other-work trips. The 303 simulation is best on home-school, home-college, and home-shop. On overall trips, the 500 simulation is best, followed by the 751 and then the 303.

The number of home-college trips from the expanded Sydney HTS seems questionable. The 2001 census data from the Australian Bureau of Statistics (ABS) reports that there are 105,844 full time and 63,788 part time university students in Sydney *[12]*. This means that the Sydney HTS reports 0.81 trips per student, while the simulations report from 1.24 to 1.37 trips per student from the 303 and 500 simulations, respectively). During the cleaning of the Sydney HTS data set, part response households were removed. This was to ensure the Sydney HTS was compatible with the NPTS. The removal of part response households from the Sydney HTS data may have contributed to lower counts of home college trips than was actually the case. Alternatively, there may be an issue on how home-college trips are defined in the HTS, compared to the NPTS, from which the simulations are drawn.

Table 4: Comparison Total Trips by Purpose for the Expanded HTS and Simulations

Purpose	**Sydney HTS**	**Simulation (NPTS & 303 Update sample)**		**Simulation (NPTS & 500 Update sample)**		**Simulation (NPTS & 751 Update sample)**	
	Count	**Count**	**Diff. as % of Sydney HTS**	**Count**	**Diff. as % of Sydney HTS**	**Count**	**Diff. as % of Sydney HTS**
Home Work	1,620,018	1,877,947	15.9%	1,811,694	11.8%	1,880,881	16.1%
Home School	832,978	824,492	-1.0%	866,628	4.0%	919,972	10.4%
Home College	137,815	209,692	52.2%	232,570	68.8%	227,081	64.8%
Home Shop	1,563,287	1,556,768	-0.4%	1,676,383	7.2%	1,555,231	-0.5%
Home Other	5,582,061	4,961,321	-11.1%	5,071,073	-9.2%	4,869,672	-12.8%
Other-Work	1,632,748	1,292,862	-20.8%	1,701,669	4.2%	1,519,293	-6.9%
Other-Other	2,692,650	2,623,189	-2.6%	2,928,916	8.8%	2,752,979	2.2%
TOTAL TRIPS	14,061,557	13,346,270	-5.1%	14,288,933	1.6%	13,725,109	-2.4%

The simulation procedure produced between 11.8 and 16.1 percent more home-work trips than reported by the expanded Sydney HTS. Increasing the update sample from 303 to 500 households improved the expanded simulated data regarding home-work trips. Home-other trips were underestimated by the expanded 500 simulation by 9.2

percent compared to the expanded Sydney HTS. Through increasing the update sample to 500 households the expanded simulation registered only 4.2 percent more other work trips than found in the expanded Sydney HTS.

7. Mode Share Comparisons

Figure 1 shows that the simulation procedure did not account for the high levels of public transport and walking undertaken in Sydney. As noted earlier, the NPTS data set, which provided the travel distributions updated by small local samples, is more auto-dominated than the Sydney HTS. This was reflected by car ownership per household and the percent of non-car owning households. The expanded 751 simulation is closer to the expanded HTS than the other expanded simulations for five trip purposes for the bike/walk mode. The differences between the expanded HTS and the expanded simulations were most notable for the home-other and other-other trip purposes.

The most frequent trip purpose was home-other. The expanded simulations accounted well for three of the four modes used to undertake this trip purpose. Home other trips undertaken by the bike/walk mode were substantially underestimated by the expanded simulations, which produced roughly half the number of trips reported by the expanded HTS. The simulations underestimated the bike/walk mode in Sydney for all trip purposes, except home college.

The most frequently-used mode of travel was driving by privately owned vehicle (POV). Figure 2 compares the number of trips undertaken by POV drivers by trip purpose for the expanded HTS and simulations. All the simulations compare favorably to the expanded HTS. The expanded simulations overestimated the number of other-other trips undertaken by the POV driver mode. This relates to the simulation procedure not accounting for the high use of the bike/walk modes in the HTS.

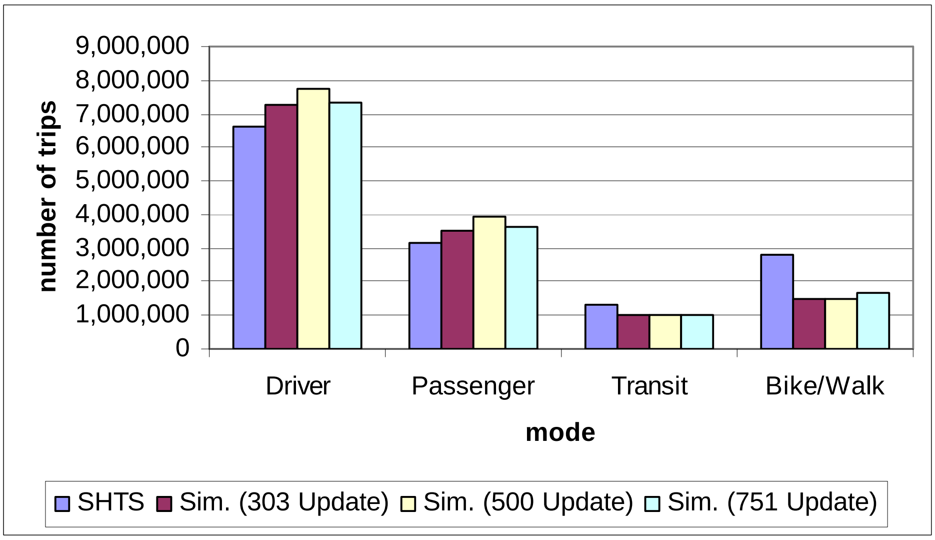

Figure 1: Comparison of the Expanded HTS and Simulations by Mode

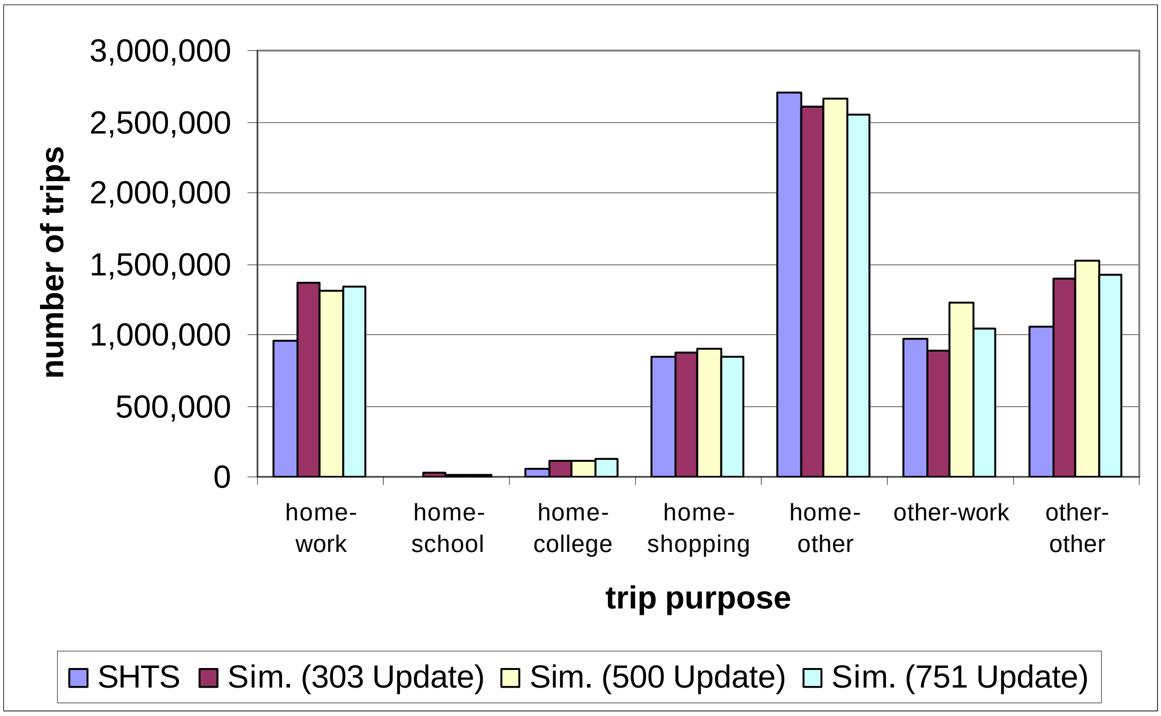

Figure 2: Comparison of POV Driver Trips for the Expanded Sydney HTS and Simulations

8. Departure Time Comparisons

The three expanded simulations were compared to the expanded HTS data set across four time periods by trip purpose. The departure time periods used were: 6:01 a.m. – 9 a.m., 9:01 a.m. – 4 p.m., 4:01 p.m. – 7 p.m., and 7:01 p.m. – 6 a.m. The expanded 500 simulation was closer to the expanded HTS data than the other simulations with respect to departure time for twelve of twenty eight pairings. The expanded 751 simulation was closer than the other simulations for nine pairings, while the 303 simulation was closer for seven pairings. All the expanded simulations were close to the expanded HTS data with respect to departure time; approximately 45 percent of pairings across all expanded simulations were within ten percent of the expanded HTS count. The closeness of the expanded simulations to the expanded Sydney HTS data set can be seen in Figure 3 which presents the total number of trips undertaken by trip purpose.

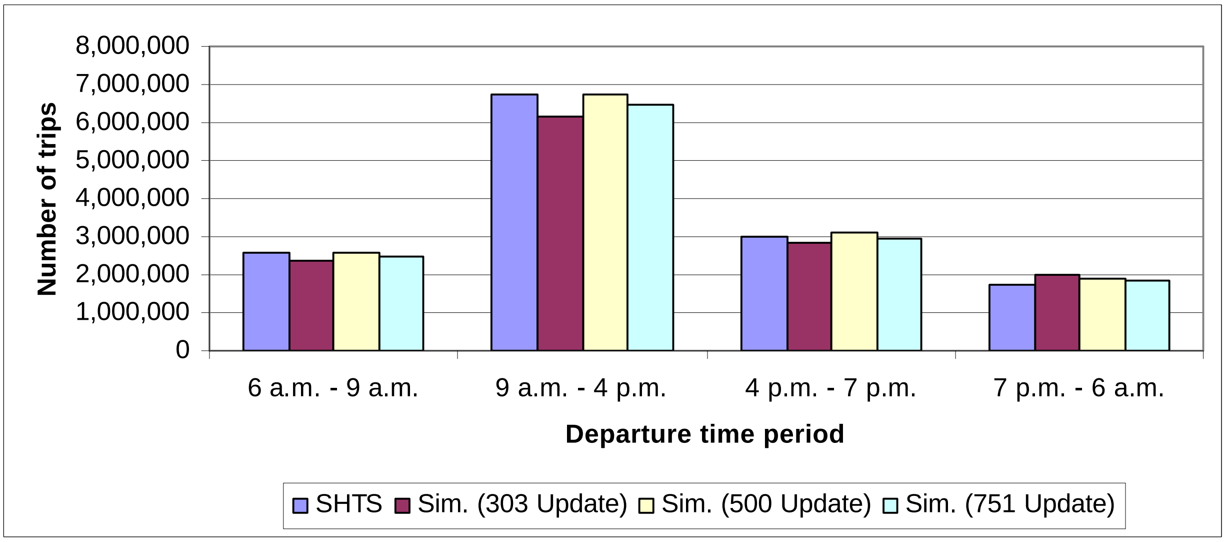

Figure 3: Comparison by Departure Time Period for the Expanded HTS and Simulations

In order to categorize the departure-time distributions for the simulation procedure, the NPTS and sample data were spread over 23 socio-demographic categories and 24 hourly categories. It would be expected that by spreading the update sample over this number of categories, lumpy data would result. This was the case in a few of the data category pairings. It would be expected that as the update sample size increases so would the similarity of the corresponding simulation to the expanded Sydney HTS. However, the expanded 500 simulation performed better in replicating the departure time by trip purpose of the expanded HTS than the other two expanded simulations.

9. Trip Duration Comparisons

All of the expanded simulations approximate the expanded HTS data. Figure 4 shows a comparison of the total number of trips by five-minute steps in trip duration for each expanded simulation. The expanded 303 simulation was remarkably close to the expanded HTS, and was closer than the other two expanded simulations for eight of eleven total trip duration categories. The three expanded simulations were very similar in their overall resemblance to the expanded HTS, and were within approximately ten percent of the number of total trips recorded by the expanded HTS for trip durations up to 25 minutes long. For other trip durations the difference was closer to 30 percent. The level of accuracy corresponds to the limited number of data points provided by the update samples for the longer trip durations. Lack of data points translates to the local data being unable to exert enough influence on the NPTS distributions that drive the simulation procedure.

It was noted earlier that the expanded simulations recorded a lower number of home-work trips than the expanded HTS. Figure 5 shows the comparison of the expanded HTS and simulations in relation to the home-work trip purpose by five-minute trip durations. The main shortfall in the number of home-work trips between the expanded Sydney HTS and the expanded simulations can be attributed to the number of people that spent over one hour traveling. The simulations mirror the movement of the expanded Sydney HTS for all trip duration categories. The expanded simulations overestimated the number of home work trips between one and thirty minutes in duration compared to the expanded Sydney HTS data set. The expanded simulations underestimated the number of trips reported by the expanded Sydney HTS data set for the remaining trip duration categories. These general differences between the expanded simulations and the expanded Sydney HTS data set can be attributed to the size of Sydney. Sydney is a large city. Consequently the population of Sydney travel for long periods of each day to get to work. In addition, Sydney has very few urban freeways, compared to most North American cities of comparable size, which may also lead to longer trip durations. Also, the NPTS, from which the travel distributions were derived, combines data from a number of large, medium and small urban areas. The updating procedure successfully resulted in expanded simulations where home-work trips mirrored the movement, but not the scale, of the expanded Sydney HTS data. The updating of the NPTS travel distributions with a local update sample with respect to home-work trip lengths was successful. As the size of the update sample increased the resulting expanded simulations moved closer to the expanded Sydney HTS data.

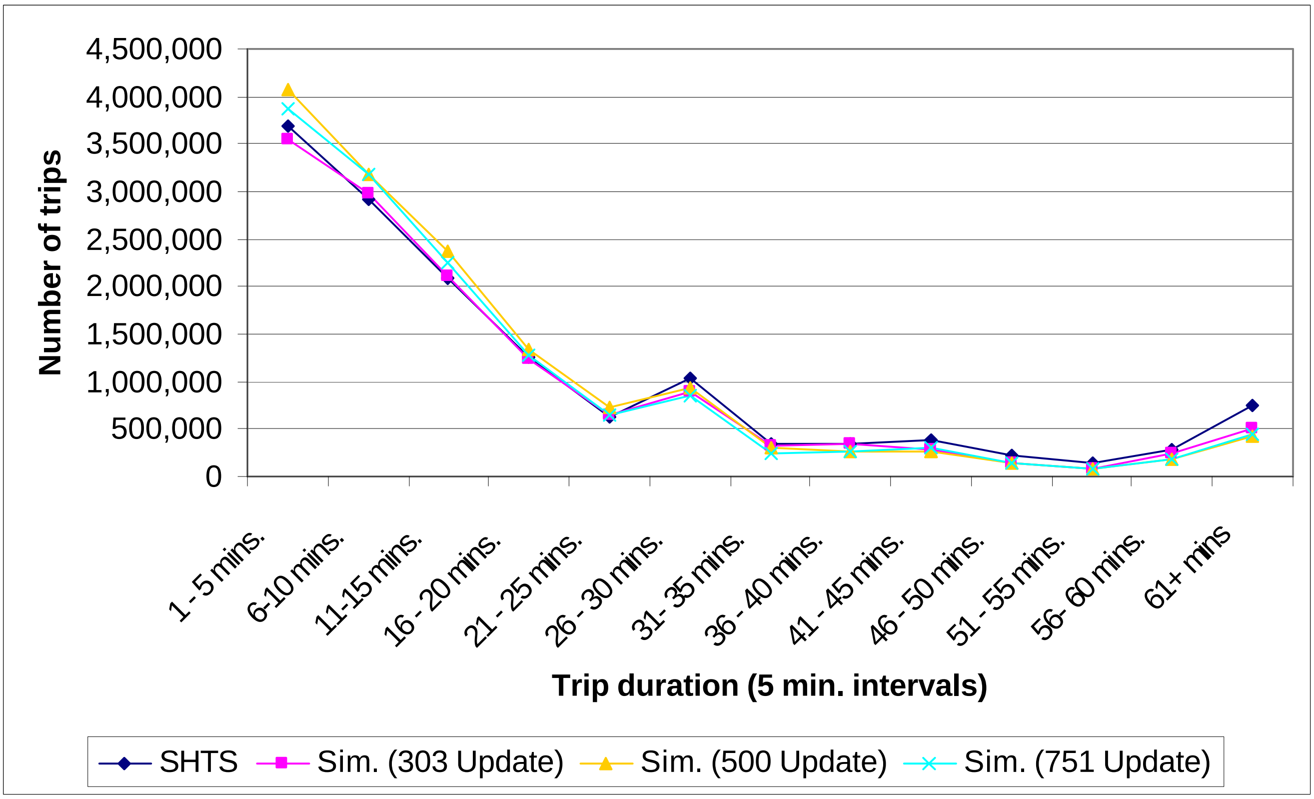

Figure 4: Comparison of Trip Durations for the Expanded HTS and Simulations

10. Conclusion

The simulation of travel data with Bayesian updating of NPTS travel-characteristic distributions with local data from a small update sample, has been shown again to produce data that are similar to actual surveys. The aim of this research was to build on the success gained by using this procedure in Adelaide by learning more about the ideal update sample size. This research has found that there was considerable benefit from using the 500 household update sample as opposed to the 303 household update sample. The added benefit of increasing the household sample size to 751 does not appear to be commensurate to the expense needed to obtain such a sample. If expense was not an issue the organisation would obtain a full HTS anyway. Having said that, the experience of this research is that an update sample of 500 does a comparable job to an update sample of 751 households. The 303 simulation, although inferior to the 500 simulation in terms of performance, performed quite well, and better than expected. The 303 simulation showed that the simulation procedure produces representative data when limited data points are available. The principle application of the simulation procedure where a smaller update sample may be a necessity is in the simulation of smaller areas where limited data points are available, e.g., the simulation of transport corridors and within subregions.

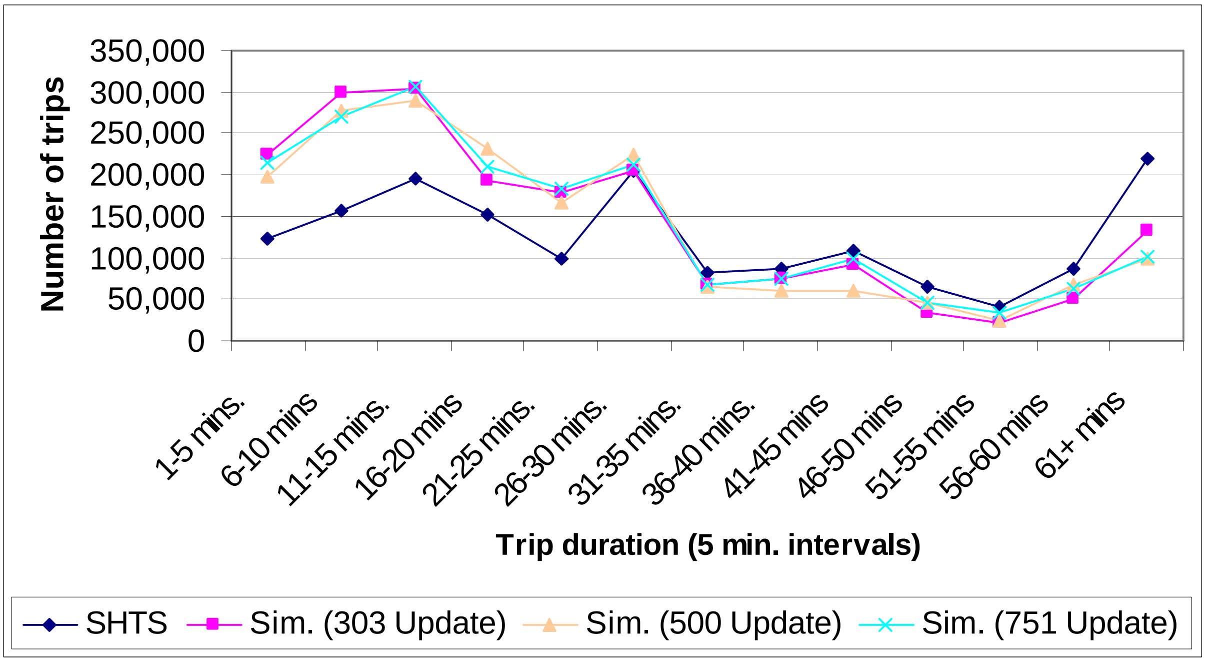

Figure 5: Comparison of Home-Work Trips for the Expanded Sydney HTS and Simulations

It was evident from the results of the various simulation procedures that it was not possible to remove all of the idiosyncrasies of the U.S. data, nor was this expected. The updating procedure reduced or negated travel characteristics peculiar to the U.S. Where major idiosyncrasies existed between the source travel-characteristic distributions, the NPTS, and the local situation, as evidenced in the Sydney HTS, muted differences were carried through to the simulations. The updated simulations could not totally account for the lower number of home work trips, the high level of bike/walk use (and, therefore, lower POV driver and passenger trips), and the greater length of Sydney trips. After the experience of Adelaide these can be regarded as general differences between Australia and U.S. The simulation procedure worked well at mirroring changes in the level of use among categories; however, the relative volumes varied. An interesting byproduct of the research was the finding that the simulations did not routinely move closer to resembling the Sydney HTS as the update sample size increased. There were some situations where the 303 simulation performed better than the 751 simulation. The reason for this was twofold:

- The varying travel characteristics of the update samples. The update samples from the Sydney HTS were random subject to Sydney HTS household size and vehicle representation. Within these categories, households exhibited a range of other behaviors relating to trip mode, duration, and departure time.
- The different mobility factors applied to the simulations during expansion. The 751 simulation had the benefit of a larger sample size than the other simulations from which to assign mobility factors. However, this did not stop the resulting expanded data set from being lumpy. Therefore, the expanded 751 simulation failed in some areas to perform well compared to the other two expanded simulations.

This suggests the possibility for further research. It is evident that different household types have a disproportionate influence on different trip characteristics. Research into the oversampling, or targeted sampling, of specific household types to counteract the U.S. penchant for privately-owned vehicles and short trip lengths would be of value. This raises the question of whether to oversample specific household types within the update sample or within the socio-demographic data set used to drive the simulation, or both. This is not a failure of the simulation procedure. This research has shown that with a small local sample, it is possible to produce a data set similar to that of a large household travel survey. This suggestion for research, if successful, would produce simulated data that builds on this achievement. Two other possible directions for future research are to consider alternative weights, such as weights that are proportional to the size of the update sample; and to use a hierarchical Bayesian approach, to take into account the biases in the NPTS data as applied in Australia.

References

Schofer, Joseph L. (2002) Letter report – National Household Travel Survey (NPTS), Committee to Review the Bureau of Transportation Statistics (BTS) Survey Programs, Washington D.C., June 18.

Greaves, Stephen P. (1998), *Applications of GIS Technology in Recent Household Travel Survey Methodologies*. Report prepared for the Travel Model Improvement Program, Federal Highway Administration.

Greaves, S.P. (2000), *Simulation Household Travel Survey Data for Metropolitan Areas*, Unpublished Ph.D. Dissertation, Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, Louisiana.

Greaves, S.P. and P.R. Stopher (2000) Creating a Simulated Household Travel/Activity Survey – Rationale and Feasibility Analysis, *Transportation Research Record No. 1706*, pp. 82-91.

Stopher, P.R., Greaves, S.P. and Bullock P. (2001) Synthesizing Household Travel Survey Data: Application to Two Urban Areas, Unpublished Paper.

Stopher, P.R., Bullock, P. and Rose J. (2002) Simulating household travel data in Australia: Adelaide case study, Unpublished paper.

Kothuri, S. (2002), *Bayesian Updating Of Simulated Household Travel Survey Data For Small/Medium Metropolitan Areas*, Master's Thesis, Louisiana State University, Baton Rouge, Louisiana, U.S.A.

Stopher, Peter R. and Graham Pointer (2003), Monte Carlo Simulation of Household Travel Survey Data with Bayesian Updating, paper presented to the REAAA/ARRB Joint Meeting, Cairns, Queensland, May.

TDC (2001) Household Travel Survey: Survey Methodology, Transport Data Centre, New South Wales Department of Transport, Sydney, August.

Australian Bureau of Statistics Census: 1996 Census Household Sample File (HSF).

Federal Highway Administration (FHWA) (1999) 1995 NPTS User's guide, August.

ABS (2002) 2001 Census Basic Community Profile and Snapshot: 11 Sydney (Major Statistical Region) New South Wales, 19th November, Canberra.