Developing a Method for Simulating Trip Tours in Urban Areas

Peter.R.Stopher[1], Tony.Bertoia[1]
[1]The Institute of Transport and Logistics Studies, University of Sydney, NSW, Australia.

1 Background

The purpose of this study was to determine whether analyses based on tours rather than trips could be used to characterise household travel patterns successfully for use in synthetic travel data simulation. This study used the 2001 US National Household Travel Survey (NHTS) Day Trip file to analyse household travel and to determine whether tours could be derived intelligently from the available data. The 2001 NHTS data is the most recent national travel survey dataset from the U.S. and is a very large data set (containing 642,292 trips). The size of the data set provides a unique challenge and opportunity to examine and refine trip-to-tour conversion procedures on a large scale.

This paper outlines the operational and procedural steps involved in cleaning the data set and formulating strategies for categorising the trips into tours. Some basic statistical analyses of the tour data are presented, and a summary of important considerations when dealing with tour-based data are suggested.

1.1 Definition of a tour

There has been some debate over the most appropriate definition of a tour, with differences often reflecting varied research goals (see O'Fallon and Sullivan, 2005, for a review). Some authors (e.g., Axhausen, 2000) suggest that tours should be considered a series of linked trips that begin and end at the same location. This characterisation of tours could therefore include the possibility that tours may begin and end at work (for example, an employee taking a lunch break and returning to work). On the other hand, others employ the definition first articulated by Adler and Ben-Akiva (1979): *"a set of consecutive trip links which begin and end at an individual's home"* (p.244). This definition has been employed by many researchers in the field, including Bhat et al. (2001), Kitamura (1984) and Golob (1986). For the purposes of this study, it was deemed appropriate to use the home-based definition of a trip tour as described by Adler and Ben-Akiva (1979) above.

Furthermore, there has been little consensus regarding terminology describing each trip within a tour, with terms such as 'trip leg', 'trip link', 'activity stop', and 'segment' being variably used (O'Fallon and Sullivan, 2005). These authors used the term 'segment' to refer to each trip within a tour, arguing that it implies that a trip is part of a greater whole (i.e., a tour). However, we argue here that the word 'stage' is a more appropriate term to use, particularly given the specific meanings of the terms 'stage' and 'tour' in the context of travel. The online Oxford English Dictionary (www.oed.com) defines 'stage' as 'a division of a journey or process', which is more specifically defined as both 'a period of a journey' as well as 'the distance travelled between two places of rest'. Similarly, the term 'tour' (defined as 'a turning round, a circular movement, a revolution' and 'an excursion or journey including the visiting of a number of places in a circuit or sequence') also involves a very specific and appropriate definition specific to travel. Hence, the versatility of the term 'stage' and the similarity in meaning to the intended definition of a tour in the context of travel survey research, suggest the use of 'stage' to complement the use of the term 'tour' in defining a series of trips beginning and ending at home.

2 Data Cleaning Procedure

The original NHTS 2001 day trip file contained a total of 642,292 trips. From this original trip file, two major deletions were necessary. First, we deleted adult proxy records (i.e., records of trips made by adults, which were recorded by another adult proxy). The reason for excluding these adult proxy records is that they could be inaccurate and thus lead to under-reporting of trips made (Greaves, 2000). Second, weekend trips were deleted to reduce the variability of trips and hence increase the chances of correctly specifying the nature of tours. It also served to increase the comparability of the results with other studies using weekday travel only. After these deletions were made, the total number of trips in the database was 382,361. A number of logical checks were performed on the data prior to any work being done on the major anticipated problems regarding, for example, missing trips to home, incorrect time information, etc.

2.1 Logic checks

Logic checks were conducted on the data to identify problems associated with certain variables. Problems were flagged and dealt with manually where appropriate. Examples of some of these logic checks and their solutions are outlined below:

- *Logic check:* The driver on a trip should not be younger than 16.
  Outcome: 301 such cases were found and codes were modified.
- *Logic check:* The start time of a trip should not be later than the end time of a trip.
  Outcome: 529 problems existed in the database. Some problems were due to the coding of time-related variables. Such variables were recoded to allow for more correct calculation of time.
- *Logic check:* The start time of a trip should not be earlier than the end time of the previous trip for the same person.
  Outcome: 1,348 problems existed in the database. Inconsistencies after recoding time-related variables were adjusted manually to reflect a correct chronological sequence of trips.
- *Logic check:* The number of drivers in the household should not be greater than the number of residents in the household.
  Outcome: No such problems were found.
- *Logic check:* If the individual was a driver on the trip, they must have a driver's licence.
  Outcome: No such problems were found.
- *Logic check:* The number of household members on a trip should not be greater than the total number of household members.
  Outcome: 23 problems existed in the database. Values of the number of household members on the trip were altered to reflect values found in other trips for the same person or members of the same household.

2.1.1 Missing to-home trips

Once the logic problems in the database were dealt with, it was necessary to deal with unreported, but likely, trips to home that were not included in the database. This was a crucial step in terms of setting up the trip database for eventual conversion to a tour file, because by definition, a tour involves a chain of trips beginning from home and ending at home. Missing trips back home at the end of the day are a common problem in diary surveys, and many cases were found in this data set where such trips were not reported. Missing trips to home at the end of each day were inserted automatically using a loop function in SPSS syntax.

Approximately 17 percent of the inserted to-home trips involved a visit to friends or relatives. This was expected because many return trips to home are often unreported if the previous destination was a social occasion. Problems with inserted trips were manually fixed to ensure they made intuitive sense. The total number of inserted missing to-home trips was 9,914, which comprised 2.5 percent of all trips.

2.1.2 Other data cleaning issues:

There were many problems related to the coding of start times and end times for trips which prevented accurate calculations of trip time. To overcome these problems, it was essential to conduct a 2-stage process of cleaning the start and end time variables. First, start and end times were recoded to reflect the fact that the travel day in the NHTS begins at 4 a.m. Hence, start and end times with values from 0 to 59, for example, which reflect times from midnight to 12:59 a.m., were recoded to times with values from 2400 to 2459. This process was also undertaken for start and end times from midnight to 3:59 a.m. (i.e., up to 2759, which corresponds to 3:59 a.m.). Times of 4 a.m. onwards retained their original values.

Second, it was essential to identify the problematic start and end times after they had been recoded to this new time structure. These variables were manually altered to reflect common sense within the context of an individual's travel patterns and to conform to the logic checks carried out prior to this step.

2.1.3 Preparing the trip file for conversion to a tour file

A number of steps were undertaken to prepare the trip file for conversion to a tour file. First, the first stage of each tour was identified. Each subsequent stage of the tour was then identified and assigned a numerically increasing 'tour stage' value. This process was continued until a trip to home was identified, indicating the end of the tour. After trips were identified as belonging to a tour, each tour was numbered within each person in the trip file. This would allow us to derive information later about how many tours each individual made.

Further tour information was obtained and assigned to each trip within that tour so that this information would be retained upon conversion to a tour file. The tour information included the start time of each trip; the end time of each trip; the purpose of each trip; the travel time of each trip; the travel mode of each trip; and the length of stay at each destination, and whether the tour was simple or complex. Tours were classified as either simple or complex on the basis of categorisations made in the previous step. Simple tours were defined as tours with only two trips (Home-Destination and Destination-Home), whereas complex tours were defined as tours involving more than two trips (e.g., Home-Work, Work-Shop, Shop-Home). Once all of the cleaning and data preparation was completed, the final trip file contained a total of 391,973 trips, representing an increase in the number of trips relative to the pre-cleaning database, mainly as a result of adding trips back to home that were omitted in the original data.

3 Results

The cleaned NHTS 2001 Day Trip file was converted to a tour file using SPSS. Relevant demographic information was retained, and important tour-related information was aggregated to the tour level. The tour file contained a total of 135,847 tours. A similar process of aggregation was also conducted on this tour file to create a 'Person tour file' which contained aggregated information about tours at the person level. The Person tour file contained a total of 85,099 persons.

3.1 Tour File Data Analysis

To understand characteristics of these tours, correlation analyses were conducted at the person level (N=85,099 persons). The results showed significant correlations between the number of tours per person and a number of key factors (see Table 1). A positive correlation between the number of tours and the number of trips per person (including non-tour trips) suggests that the likelihood of chaining trips into tours is related to the total number of trips. Non-tour trips included chains of trips which either did not have a home starting point (likely to occur in the first trip) or which did not have a home destination (such as visiting a friend and staying overnight at the end of the travel day). Failure to include such trips in any comparison of trip and tour rates would provide an inaccurate representation of each person's travel throughout the travel day.

The findings also showed a modest negative correlation between the number of tours per person and total tour duration. This suggests that the greater the amount of time spent away from home, the lower the likelihood of embarking on further trips in that tour. This makes intuitive sense because a greater amount of time spent at activities decreases the amount of time available during the rest of the day to embark on further tours. Other results showing, for example, a significant negative correlation between the number of tours and total tour travel time confirmed the importance of availability of time throughout the day.

Table 1: Correlations between Number of Tours per Person and Key Variables

Variable	Pearson Correlation (with number of tours)	N
Total Trips (Tour and Non-Tour Trips)	.721**	85,099
Total Travel Time During Tours	-.059**	84,721
Total Activity Duration[a]	-.063**	82,968
Total Number of Trips in All Tours	.721**	85,099
Total Tour Duration[b]	-.472**	85,035
Number of Single Mode Tours	.937**	76,330
Number of Multi-Mode Tours	.250**	15,098
Number of Simple Tours	.750**	55,541
Number of Complex Tours	.459**	48,904
Household Size	.012**	85,099
Age	.097**	83,942
Number of vehicles in household	.028**	85,099
Number of licensed drivers in household	.061**	84,795

[a] Total time spent at each destination throughout each tour
[b] Time from start of first trip to end of last trip in each tour
** p<.01

Given the importance of availability of time during the day, one might expect a positive correlation between age and the total number of tours. One could argue that, as people enter retirement age, they may gain some freedom from the time constraints of regular full-time work, thus resulting in a greater amount of time available to embark on a greater number of tours. However, one might also expect that mobility issues associated with ageing may restrict opportunities to embark on a greater number of unnecessary tours. Hence, one might also expect a negative correlation between age and the total number of tours because of the likelihood that either older adults might chain more trips together within the one tour, or that older adults may be likely to make less trips in general.

In fact, the findings showed that age was slightly positively related to the total number of tours per person, which suggested that an important factor was the greater amount of time available to embark on tours. This again is reflected in a significant positive

correlation between age and the number of trips in all tours (R=.158, N=83,942, p<.05), suggesting that older individuals in the NHTS 2001 survey embarked on a greater number of trips in each tour than younger adults. This finding implies that once older adults leave home, they generally get more done than younger adults before they return home. This may reflect greater economy of travelling among older relative to younger adults. However, given the relatively weak correlation, it is important to note that such conclusions are merely conjecture and in need of further study.

Further correlation analyses of tours at the household level were conducted (see Table 2). The findings revealed that the number of household tours was more strongly correlated with household size than with the number of adults, suggesting that household size may be more useful in characterising household tour behaviour. The findings also showed a stronger correlation between the number of household tours and the number of vehicles than between the number of tours and the number of licensed drivers. This suggests that, at the household level, the number of available vehicles may be more important than the number of drivers in determining the number of household-level tours.

Table 2: Correlations between Number of Tours per Household and Key Variables

Variable	Pearson Correlation (with number of tours)	N
Number of vehicles in household	.477**	46,818
Number of licensed drivers in household	.206**	47,005
Household size	.593**	47,005
Number of adults in household	.249**	46,770

** p<.01

Correlations at the person level (Table 1) and at the household level (Table 2) showed that the number of vehicles and the number of licensed drivers in a household are more strongly correlated with the number of household tours than the number of person tours. This suggests that characteristics of household tour behaviour may not be inferred from person tour behaviour, and that it can be beneficial to analyse them separately.

3.2 Travel Mode

Correlation analyses involving the number of tours by travel mode category revealed an interesting picture of the travel modes employed in tours. A very high positive correlation with the number of single mode tours and a moderate positive correlation with the number of multi-mode tours were found. The strength of the correlations, as well as the relative number of single and multi-mode tours (see Table 1), suggest that by far the most common tour mode category is the single mode tour. Furthermore, given that the total number of tours is quite closely related to the number of single mode tours, this suggests that people who embark on a higher number of tours during the day are more likely to use the same mode throughout a tour than people who are less likely to embark on many tours during the day.

Similar analyses of simple and complex tours showed that the total number of tours was significantly related to the number of simple tours as well as complex tours. This suggests that people who undertook more tours during the day generally opted for simple tours, both in terms of travel mode used and in terms of number of stages in the tour. Further analysis of the purpose of each tour is described in Section 3.3.

3.2.1 Tour Mode Categorisation

The main mode for a tour can be categorised in a number of ways. O'Fallon and Sullivan (2005) argued that basing categorisations on measures relying on time estimates of

individual trips, including speed estimates, can lead to reliability issues due to underestimation on, for example, walk trips. Instead, a better approach is to categorise the main mode of tours based on some measure of distance, particularly if geo-coded trip information is available. Unfortunately, distance estimates in the NHTS 2001 day trip file database were reported by respondents, and are likely to involve some reporting biases such as underestimating the distance on long trips. However, we argue that although it is not ideal to use respondent estimates of trip distance, it is nevertheless sufficiently reliable to serve the purpose required in allowing for tours to be categorised according to the mode used on the longest trip during the tour.

Consider an individual who estimated that in the morning they travelled by car for 5km to drop a child off at day care, and then drove 2km to a train station, where they caught a train that travelled 15km to their workplace. After work, they travelled home in the reverse manner. Now suppose that the train trip (which incorporates the change mode trip at the train station) took approximately fifteen minutes, while the morning car trip to the childcare centre took 10 minutes. However, suppose that the car trip home from the childcare centre in the evening took twenty-five minutes due to unusually heavy traffic. In this case, using a time-based rule for assigning main mode to this tour would result in this tour being classified as a 'car driver' tour. This is because the total trip time for the train trip would be thirty minutes, while for the car driver trip to and from childcare it would be thirty-five minutes. Instead, using distance-based criteria, this tour will always be classified as a 'train' trip even if the distance on the train trip is underestimated. This simple example illustrates how unusual circumstances such as traffic congestion can alter the assignment of main mode to tours, and reinforces the argument that even accounting for potential underestimation in the distance of long distance trips, it is preferable to use distance-based criteria to assign main modes to tours.

Another issue involves whether to assign travel mode based on the total distance in the tour related to each travel mode or whether another criterion can be applied. In this study, we categorised main mode by identifying the stage in the tour which involved the greatest amount of distance travelled. The mode of travel at this stage was assigned as the main mode for the tour. We reasoned that focusing on the trip (stage) covering the greatest distance in the tour is preferable to adding the distances for all trips using the same mode and choosing the mode covering the greatest distance. This is because in some situations there is the potential to choose a travel mode that is not representative of the purpose of the tour. Hence, a tour involving a 20km return trip to work on a train, as well as a series of car driver shopping trips in the evening totalling 21km could characterise this tour as a car driver work-related tour. Information relating to the main mode is more likely to be appropriately attributed to a tour if the main mode of the trip comprising the greatest distance covered is used. This criterion does not allow for a series of smaller and relatively unimportant trips to dominate the selection of the main mode for the tour. In this study, this maximum trip distance criterion allowed for classification of main mode to 97.2% of all tours. Table 3 shows that the majority of tours were classified as 'vehicle driver' tours, while almost a quarter of all tours were 'vehicle passenger' tours.

Table 3: Main Mode of Tour based on Longest Distance Travelled

Mode	Tours	Percent
Vehicle Driver	78797	58.0
Vehicle Passenger	33179	24.4
Public Transport	7536	5.5
Bicycling	1415	1.0
Walking	11091	8.2
Total of classified tours	132018	97.2
Missing	3829	2.8
Total	135847	100.0

3.3 Tour Purpose

Tour purposes were classified according to a tripartite hierarchical model of activities based on variability in location, frequency, duration and scheduling of activities (Stopher, Hartgen and Li, 1996). ***Mandatory*** activities are said to be those which have largely fixed frequency and location, such as work, school and day care. ***Flexible*** activities are those which may have some fixed characteristics, but at least one characteristic that is variable. For example, while grocery shopping may be done at varying frequencies, it nevertheless is a necessary activity for households to conduct at least once in a while. Similarly, other flexible activities include banking, shopping, some work-related activities and medical care. The key defining factor in flexible activities is the combination of flexibility in at least one characteristic and the necessity of the activity for survival. On the other hand, ***Optional*** activities are considered to be activities that are not necessary for the biological survival of an individual. Such activities include most social, recreational, cultural and community activities.

The key feature of an optional activity is that the activity itself is not necessary – that is, one can make a choice not to be involved in that activity without affecting survival. While the use of such a needs-based classification of activities in time-use research is not new (see Bhat and Koppelman, 1999), the difficulty in applying decision rules to the complexities involved in tours dictated that we rely on relatively simple tour purpose classifications prior to either applying or indeed possibly developing more complex needs-based theories in tour-based analyses. Table 4 shows the specific trip purposes that were used to categorise tours into the three tour purposes outlined above. Tours that contained at least one mandatory trip were classified as mandatory tours. Of the remaining unclassified tours, those that contained at least one flexible trip were classified as flexible tours. Finally, all remaining tours which contained at least one optional trip were classified as optional tours. The tour purpose categories shown in Table 4 are those used in the 2001 NHTS. For example, school/religious activity is a single trip purpose in the data set, however it is often characterised as two distinct trip purposes in other data sets.

Table 5 shows the frequencies for each tour purpose categorisation in the tour file. Only 0.3 percent of tours were unclassifiable due to a lack of any trip purpose information in any of the trips in the tour (aside from ‘going to home’). Further analyses of expanded tour purposes (designed to further distinguish tour purposes according to whether the tours were simple or complex) can also be seen.

The findings show clearly that the majority of tours made by persons in the NHTS day trip database (58.4%) are simple tours. Of these simple (2-stage) tours, it appears that there were a similar proportion of mandatory (22.9%), flexible (16.6%) and optional (20.8%) tours. This pattern of simple tours suggests that people are almost equally likely to embark on a simple tour specifically to engage in recreational activities as they are to embark on a simple tour involving, for example, work-based activities.

However, when analysing complex tours, Table 5 shows that the percentage of optional tours (3.9%) is dramatically smaller than either mandatory (17.7%) or flexible tours (19.7%), while both complex mandatory and flexible tours are relatively comparable to simple mandatory and flexible tours respectively. These findings suggest that people are less likely to embark on anything other than simple tours when engaging in an optional (e.g., social) activity. This makes sense when one considers the very low likelihood that, for example, one might engage in grocery shopping before or after a social engagement.

Clearly, these findings demonstrate a distinct pattern of tour travel based on the tour purpose classifications used in this study. Although simple tours are relatively equally likely to involve either mandatory, flexible or optional activities, complex tours are rather unlikely to involve optional activities. A clear hierarchy in terms of people’s travel patterns

based on activities has emerged. Tours involving only optional activities clearly are more likely to involve two trips (or stages) only, suggesting that people are unlikely to chain optional activities, such as going to the gym and visiting a friend. On the other hand, tours involving mandatory and flexible activities at any stage are equally likely to be simple or complex, suggesting that people setting out to accomplish mandatory or flexible activities are more likely to look for the opportunity to combine another mandatory, flexible or optional activity along the way, such as going to work and dropping a child at day care or school.

Table 4: Tour Purpose Classifications

Tour Classifications	Trip Purposes	Destination Purpose Code
Mandatory	Go to work	11
	Return to work	12
	School/religious activity	20
	Go to school as student	21
	Day care	24
Flexible	Attend business meeting/trip	13
	Other work related	14
	Go to library: school related	23
	Medical/dental services	30
	Shopping/errands	40
	Buy goods: groceries/clothing/hardware store	41
	Buy services: video rentals, dry cleaner/post office/car service/bank	42
	Buy petrol	43
	Use professional services: attorney/accountant	61
	Use personal services: grooming/haircut/nails	63
	Pet care: walk the dog/vet visits	64
	Meals	80
	Get/eat meal	82
Optional	Go to religious activity	22
	Social/recreational	50
	Go to gym/exercise/play sports	51
	Rest or relaxation/vacation	52
	Visit friends/relatives	53
	Go out/hang out: entertainment/theatre/sports event/ go to bar	54
	Visit public place: historical site/museum/park/library	55
	Family personal business/obligations	60
	Attend funeral/wedding	62
	Attend meeting: PTA/home owners association/local government	65
	Transport someone	70
	Pick up someone	71
	Take and wait	72
	Drop someone off	73
	Social event	81
	Coffee/ice cream/snacks	83
	Other reason	91

Table 5: Simple and Complex Tours by Purpose

		Tour Type		
Tour Purpose Category	**Measure**	**Simple**	**Complex**	**Total**
Mandatory	Frequency	28588	24062	52659
	(Percent)	(35.9%)	(42.9%)	(38.8%)
Flexible	Frequency	22586	26716	49305
	(Percent)	(28.4%)	(47.6%)	(36.3%)
Optional	Frequency	28213	5310	33526
	(Percent)	(35.5%)	(9.5%)	(24.7%)
All categorised tours	Frequency	79387	56088	135490
	(Percent)	(99.8%)	(99.97%)	(99.7%)
Missing/Misclassified	Frequency	162	18	357
	(Percent)	(0.2%)	(0.03%)	(0.3%)
Total Tours		**79549 (100%)**	**56106 (100%)**	**135847 (100%)**

NB: 192 tours were not included in the above analyses because they were incomplete and consisted of only 1 tour leg.

3.3.1 Analyses of Complex Tours by Purpose

Specific analyses of complex tours were deemed appropriate to illuminate the findings outlined above. Descriptive statistics for each *complex* tour purpose type are presented in Table 6. The results indicated that although a greater average number of trips were observed among flexible and optional relative to mandatory tours, the differences in total travel time were very small. The most striking findings were the large differences in total elapsed time of tours as a function of tour purpose. As expected, mandatory tours involved a much greater elapsed time relative to the others, and flexible tours were shorter in duration than optional tours.

Table 6: Descriptive Statistics for Complex Tours by Purpose

		Complex Tour Type		
Statistic		**Mandatory**	**Flexible**	**Optional**
Number of trips in tour	Mean	5.66	6.77	6.68
	S.D.	2.515	2.964	3.185
Total travel time in tour (mins)	Mean	72.47	72.02	69.31
	S.D.	61.061	75.602	82.314
Total elapsed time in tour (mins)	Mean	536.00	214.29	239.68
	S.D.	185.179	173.584	200.913

The findings suggest that while mandatory complex tours generally involve fewer trips than either flexible or optional tours, they also involve a much greater amount of elapsed time (over twice as long). The number of trips in flexible and optional tours may be greater than the number of trips in mandatory tours because flexible and optional tours may comprise a greater number of different activities than mandatory complex tours, which are likely to involve a limited number of activities. The length of mandatory tours (almost 9 hours on average) suggests that the likelihood of adding extra activities onto such tours is reduced because of the amount of time people are already spending away from home on such tours.

3.3.2 Modal Split of Tours by Purpose

Analyses of the modal split of tours by purpose (see Table 7) revealed that approximately the same number of vehicle driver tours were classified as mandatory or flexible, and that these constituted more than sixty percent of the tours for each of these two purposes. In contrast, vehicle driver was used for less than fifty percent of optional tours, and the number of such tours was about half the number for each of mandatory and flexible tours. Vehicle passenger accounts for the largest share of optional tours at almost thirty percent, while accounting for twenty seven percent of flexible and less than twenty percent of mandatory tours. In number, more vehicle passenger tours are flexible, and the smallest number of vehicle passenger tours by purpose is still the optional tours. Public transport accounts for over twelve percent of mandatory tours, but just over one percent of flexible and less than one percent of optional tours. In numeric terms, overwhelmingly the largest number of public transport tours are mandatory tours, while few public transport tours are optional. Bicycling and walking account for numerically larger numbers of optional tours, and also account for almost three percent and over fifteen percent of these tours. Bicycling is used very little for mandatory or flexible tours, with shares of these tours of less than one percent. Walking accounts for a higher number and percentage of flexible tours than mandatory, but both are only about one third of the share for optional tours.

Table 7: Modal Split of All Tours by Purpose

	Mandatory		Flexible		Optional	
Mode	**Tours**	**Percent**	**Tours**	**Percent**	**Tours**	**Percent**
Vehicle Driver	31593	60.0	31054	63.0	16027	47.8
Vehicle Passenger	10385	19.7	13121	26.6	9549	28.5
Public Transport	6532	12.4	696	1.4	297	0.9
Bicycling	302	0.6	204	0.4	898	2.7
Walking	2708	5.1	3068	6.2	5282	15.8
Total	51520	97.8	48143	97.6	32053	95.6
Missing	1139	2.2	1162	2.4	1473	4.4
Total	52659	100.0	49305	100.0	33526	100.0

A better understanding of the behavioural implications of these findings is obtained by separately analysing simple and complex tours. Table 8 shows the modal split of tours by tour purpose for simple and complex tours. Assuming that mandatory tours have the least variability in duration, departure time, and frequency, and optional tours have the most, we can draw some interesting conclusions about tour patterns. For example, the modal split of walking tours increases from mandatory through flexible and to optional tours. This effect is more pronounced for complex rather than simple tours. It appears that among simple tours, the modal share increase for *walking* is accompanied by a modal share increase for *vehicle passenger* tours. The increase in the percentage of *vehicle passenger* tours is more dramatic among complex tours, however, where the modal share of vehicle passenger optional tours is almost double the modal share of vehicle passenger mandatory tours. A striking finding is the dramatic modal share difference in *public transport* tours, both simple and complex, for mandatory relative to either flexible or optional tours. This is particularly so for simple tours. Also the findings suggest that mandatory simple tours (e.g., work or education-related tours) served by public transport are more likely to be simple (2-stage) tours rather than complex tours. This makes sense because perceptions of the inconvenience and time-consuming nature of using public transport (Bertoia et al., 2005) may discourage the use of public transport to append extra trips onto simple mandatory tours (such as doing some minor shopping at the end of a work day).

Table 8: Modal Split of Simple and Complex Tours by Purpose

		Mandatory		Flexible		Optional	
	Mode	**Tours**	**Percent**	**Tours**	**Percent**	**Tours**	**Percent**
Simple Tours	Vehicle Driver	14744	51.6	13878	61.4	13415	47.5
	Vehicle Passenger	5658	19.8	5215	23.1	7535	26.7
	Public Transport	4965	17.4	280	1.2	198	0.7
	Bicycling	222	0.8	128	0.6	819	2.9
	Walking	2166	7.6	2316	10.3	4954	17.6
	Total	27755	97.1	21817	96.6	26921	95.4
	Missing	833	2.9	769	3.4	1292	4.6
	Total	28588	100.0	22586	100.0	28213	100.0
Complex Tours	Vehicle Driver	16849	70.0	17176	64.3	2612	49.2
	Vehicle Passenger	4727	19.6	7906	29.6	2014	37.9
	Public Transport	1567	6.5	416	1.6	99	1.9
	Bicycling	80	.3	76	.3	79	1.5
	Walking	542	2.3	752	2.8	328	6.2
	Total	23765	98.7	26326	98.5	5132	96.6
	Missing	306	1.3	393	1.5	181	3.4
	Total	24071	100.0	26719	100.0	5313	100.0

3.3.3 Complex tour activities

We conducted further analysis of the constituent activities for each complex tour purpose type. The aim was to investigate how many complex mandatory tours involved any flexible or optional trip purpose activities, and how many complex flexible tours involved any optional activities. Table 9 displays the frequencies of each trip activity for each complex tour purpose type. Note that flexible tours, by definition, cannot include any mandatory activities, and optional tours cannot include either flexible or mandatory activities.

Table 9: Activity Analysis of Complex Tours by Purpose

	Mandatory		Flexible		Optional	
Mode	**Tours**	**Percent**	**Tours**	**Percent**	**Tours**	**Percent**
All mandatory activities	1388	5.8	-	-	-	-
At least one flexible activity and no optional activities	10795	44.8	11383	42.6	-	-
At least one optional activity and no flexible activities	6781	28.2	-	-	-	-
Combination of flexible and optional activities	5107	21.2	15336	57.4	-	-
Total	24071	100.0	26719	100.0	5313	100.0

The findings show that about 45 percent of the complex mandatory tours involve at least one flexible activity and no optional activities, while about 49 percent involve either flexible or optional or both activities. A little less than 30 percent of mandatory tours involve at least one optional activity and no flexible activities, while slightly more than 20 percent involve both flexible and optional activities. A small percentage of complex mandatory tours involves only mandatory activities, suggesting that almost all individuals embarking on complex mandatory tours include at least one other type of activity. On the other hand, almost sixty percent of flexible tours involve at least one optional activity, suggesting that a majority of individuals embarking on complex flexible tours also engage in optional activities.

3.4 Household and Person characteristics

The 2001 NHTS contains a variety of household and person demographic information which may be used to illuminate some of the behavioural characteristics associated with certain types of tour classifications described thus far in this paper. The relative benefits of using household or person information to characterise travel behaviour is an important issue in understanding trip tours. Clearly, some needs are household needs, which may or may not be shared among members of the household (e.g., the same individual may be responsible for all the shopping needs of the household whereas the breadwinning may be shared among three householders). This suggests that an understanding of the interrelationships within the household is important. The difficulty in achieving an appropriate characterisation of such household trip interrelationships suggests that applying this debate to tours may provide even more obstacles for transport modellers and researchers. Regardless of this difficulty, it is important to begin to attempt to characterise both person and household tour behaviour before even considering whether interrelationships within the household can be modelled successfully. In this section, we outline some preliminary and ongoing work in developing a better understanding of person- and household-level tour behaviour.

3.4.1 Person-level analyses

Person-level analyses of gender differences and differences between workers/students and non-workers/non-students are reported in this paper. Analyses of gender differences in tours (Table 10) revealed that while males embarked on significantly more mandatory tours than females, the reverse was found for both flexible and optional tours. Analyses of tours by mode revealed gender differences in all mode categories. Males embarked on significantly more vehicle driver, public transport and bicycling tours than females, whereas the reverse was found for vehicle passenger and walking tours.

Table 10: Descriptive Statistics for Tours by Purpose and Mode – Gender Differences

Category	Type of Tour/Mode	Males			Females			
		N	Mean	Std. Dev	N	Mean	Std. Dev	t
Tours by Purpose	Mandatory	25548	0.66	.577	27108	0.58	.586	19.26**
	Flexible	20792	0.54	.761	28508	0.61	.758	-14.55**
	Optional	14557	0.38	.630	18961	0.41	.684	-7.05**
Tours by Mode	Vehicle Driver[a]	36501	0.94	.982	42296	0.91	.982	4.94**
	Vehicle Passenger	13483	0.35	.629	19687	0.42	.662	-16.95**
	Public Transport	3749	0.10	.302	3786	0.08	.280	7.67**
	Bicycling	946	0.02	.171	469	0.01	.107	14.35**
	Walking	4782	0.12	.383	6309	0.14	.406	-4.49**

[a] equal variances assumed (Levene's test for equality of variances was not significant at α=.05)
** p < .01

Analyses of differences between workers/students and non-workers/non-students revealed a number of significant differences in the number of tours by purpose and mode (see Table 11). As expected, workers/students embarked on a significantly greater number of simple and complex mandatory tours than non-workers/non-students, while the reverse was found for simple and complex flexible and optional tours. This suggests that working people are less likely to be involved in simple flexible and simple optional tours, with the majority of the latter being accomplished by those who are not working. The likely consequence of this appears to be an increased tendency by workers/students to embark on more complex mandatory tours, whereby flexible and optional activities are included in a mandatory tour, than non-workers/non-students. In contrast, non-workers/non-students embark on significantly more complex flexible and optional tours than workers/students, presumably due to lifestyle factors such as increased leisure time.

One of the problems with the categorisation of workers/students vs. non-workers/non-students is that, so far as we are aware, there is no variable in the NHTS explicitly stating an individual's current study status. Instead, study status must be inferred from trip purpose categories such as 'go to school as student'. The likely effect of this is an underestimation of the number of students in the database, and, conversely, an overestimation of the number of non-workers/non-students. This leads to the counterintuitive situation where we have 2,586 mandatory tours that are made by non-workers and non-students, which simply cannot be true if the tours have been classified correctly and the survey participants responded truthfully. The likely explanation for this result is the underestimation of the number of students in the database.

Analyses of tours by mode revealed that the number of vehicle driver and public transport tours were significantly greater among workers/students than non-workers/non-students, while the reverse was found for vehicle passenger, bicycling and walking trips.

Table 11: Descriptive Statistics for Tours by Purpose and Mode – Workers and Non-Workers

		Workers/Students (N = 58,269)			Non-Workers/Non-Students (N=26,830)			
Category	**Type of Tour/Mode**	**N**	**Mean**	**Std. Dev**	**N**	**Mean**	**Std. Dev**	**t**
Tours by Purpose	Simple Mandatory	27017	0.46	.002	1571	0.06	.001	147.325**
	Simple Flexible	11865	0.20	.002	10721	0.40	.004	-43.539**
	Simple Optional	15021	0.26	.002	13192	0.49	.004	-47.802**
	Complex Mandatory	23056	0.40	.002	1015	0.04	.001	147.325**
	Complex Flexible	12612	0.22	.002	14107	0.53	.004	-73.826**
	Complex Optional	2588	0.04	.001	2725	0.10	.002	-26.551**
Tours by Mode	Vehicle Driver	58335	1.00	.004	20462	0.76	.006	32.357**
	Vehicle Passenger	17176	0.29	.002	16003	0.60	.005	-58.628**
	Public Transport	6720	0.12	.001	816	0.03	.001	48.482**
	Bicycling	909	0.02	.001	506	0.02	.001	-3.021**
	Walking	6915	0.12	.002	4176	0.16	.003	-11.982**

** $p < .01$

3.4.2 Household-level analyses

Given that some of the relevant household information available in the NHTS 2001 databases was represented as continuous variables, it was deemed appropriate to analyse the relationship between the number of tours, per household, by purpose and mode and the appropriate household variable using correlation analyses. Analyses of two important household variables, household size and the number of available vehicles, are presented in **Error! Reference source not found.**. Household size was positively related to all purpose and mode tour categorisations, although particularly strong correlations were found for the number of mandatory and vehicle passenger tours. Correlations between the number of vehicles and the number of tours were generally small, with the exception of correlations with the number of mandatory and vehicle driver tours. Clearly, the relatively strong positive relationship between household size and the number of vehicle passenger tours is to be expected. Further analysis of the nature of the modest positive relationship between household size and the number of household-level optional tours may be important in understanding why individuals in larger households embark on more optional tours, and whether these tours could be incorporated into other tours to reduce household travel.

4 Conclusions

This study raised a number of important issues in dealing with tour-based data from household travel surveys. The most important and time-consuming aspects of this study involved cleaning the trip data and ensuring that trips were appropriately categorised into tours. A number of key data cleaning issues were encountered in this study, including logic checks, missing to-home trips, inefficiently coded variables and missing trip data.

One of the most important lessons arising from this study was that a series of clearly documented logic checks was necessary to ensure that latent inconsistencies in the relationships between key variables were revealed and repaired. In order to reveal underlying inconsistencies effectively in the data, a period of thorough data mining and understanding the data was essential. Another important lesson involved dealing with missing data, including missing trips to home. This is a particularly important aspect of any study investigating tour-based travel, because failure to identify actual trips to home will result in failure to appropriately categorise home-based tours. Furthermore, missing data in many key variables were identified. This required a great deal of time invested in developing automated procedures to infer the missing information from the available data as well as assessing each problematic situation on a case-by-case basis.

Table 12: Correlation Analyses of Number of Tours and Household Characteristics

		Household Size	Number of Vehicles
	Tour Type	**Pearson Correlation**	**Pearson Correlation**
Tours by Purpose	Mandatory	.523**	.230**
	Flexible	.178**	.051**
	Optional	.366**	.089**
Tours by Mode	Vehicle Driver	.254**	.312**
	Vehicle Passenger	.523**	.103**
	Public Transport	.319**	-.028**
	Bicycling	.117**	.019**
	Walking	.176**	-.073**
	N	47005	47005

** p < .01 (2-tailed).

A final major hurdle in this study was the existence of inefficiently coded variables such as trip purpose, travel mode and start and end time of each trip. Some time was spent ensuring that, for example, start and end times were coded to ensure accurate time calculations. These issues were particularly important with respect to trip purpose, because trip categorisation into home-based tours depended on the availability of accurate information regarding trip purposes. Similarly, further categorisation of tours depended on the availability of accurate information regarding travel mode, trip distance and start and end time of each trip.

The results of the tour-based analyses revealed expected relationships between a variety of tour measures (such as the total number of tours, the number of stages in tours, total travel time etc) as well as demographic variables such as household size and age. The results also demonstrated that categorisations based on tour purpose and tour mode, even at a relatively simplistic level, resulted in meaningful differences between key person- and household-level variables (such as gender, household size etc) in terms of frequencies and mean number of tours. It is clear, for example, that categorising tour purpose along needs-based criteria, and categorising the main mode of tours along

distance-based criteria, are the preferred options in making appropriate characterisations of tours.

Analyses of modal split by tour purpose revealed discernible travel patterns and interdependencies. For example, public transport tours were much more likely to be mandatory rather than either flexible or optional tours, especially for simple tours. The findings suggested that analyses of modal share for tours necessitate simultaneously considering the purpose of the tour. Tour purpose was also related to the duration of tours, which may be an important consideration for modelling purposes. Analyses of simple and complex tours by purpose revealed that while simple tours were relatively equally likely to involve either mandatory, flexible or optional activities, complex tours were rather unlikely to involve solely optional activities (see Table 9).

This study also showed that it may be useful to conduct separate analyses of the number of tours aggregated to persons and households. The findings indicated the existence of tour patterns based on gender and work status. Relatively strong relationships between the number of tours and both household size and the number of vehicles suggests that these variables may be particularly useful in characterising tour-based travel.

In summary, the findings suggest that deconstructing tours based on the main purpose, main mode, number of trips (simple or complex), person-level variables such as gender, and household-level variables such as household size, is a useful method of characterising tour-based travel. Once such characterisations were made, the results were relatively easy to understand, and travel patterns were quite easily discernible. The next step now is to deconstruct tour-based travel even further, and to examine the relative importance of information about activities, persons, modes and time in explaining tour-based travel.

5 References

Adler, T and Ben-Akiva, M (1979) A theoretical and empirical model of trip chaining behaviour *Transportation Research Part B 13* (3), 243-257

Axhausen, K W (2000) Definition of movement and activity for transport modelling. D.A. Hensher and K. Button (Eds.) *Handbooks in Transport: Transport Modelling* Elsevier: Oxford

Bertoia, T Tideman, J and Stopher, P (2005) Community Perceptions of 'TravelSmart' Behaviour in South Australia *Paper presented at the 28th Australasian Transport Research Forum* Sydney Australia

Bhat, C R and Koppelman, F S (1999) A retrospective and prospective survey of time-use research *Transportation 26,* 119-139

Bhat, C Srinivasan, S and Guo, J (2001) Activity-based travel demand modeling for metropolitan areas in Texas: model components and mathematical formulations. Center for Transportation Research, The University of Texas at Austin *Accessed from: www.utexas.edu/depts/ctr/program/ctr_pubs_2002.html* (June 2003)

Golob, T F (1986) A nonlinear canonical correlation analysis of weekly trip chaining behaviour *Transportation Research (A) 20A*, 385-399

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

Kitamura, R (1984) A model of daily time allocation to discretionary out-of-home activities and trips *Transportation Research (B) 18B*, 255-266

O'Fallon, C and Sullivan, C (2005) Trip chains and tours: definitional issues associated with household travel surveys *Paper presented at the 28th Australasian Transport Research Forum* Sydney Australia

Stopher, P R Hartgen, D T and Li, Y (1996) SMART: simulation model for activities, resources and travel *Transportation 23,* 293-312