Method documentation **Vehicle occupancy in the 1997/98 New Zealand Travel Survey**

Date: **12 June 2003**

1. OCCUPANCY ANALYSIS PREPARATIONS: MATCHING VEHICLE PASSENGER TRIP LEGS TO DRIVER TRIP LEGS FROM SAME HOUSEHOLD

1.1 Initial dataset

The starting point for this analysis was a subset of the Land Transport Safety Authority (LTSA) 1997/98 Travel Survey database restricted to key urban centres only (Auckland cities, Wellington cities, Christchurch). In the interests of data integrity, we also restricted analysis to those cases where respondents had satisfactory completed all questionnaires (PEFORMS=1).

Additional restrictions for analysis of occupancy were:

1. Only households with complete responses from all eligible household members (RESPSTAT=1). This was done to ensure the validity of the matching procedure (otherwise passenger trips might remain unmatched simply because a driver from the same household had not completed the questionnaires, etc.) This restriction reduced the total number of trip legs analysed from 41,479 to 37,774.
2. Only trip modes *vehicle driver, vehicle passenger*. This restriction reduced the total number of trip legs substantially further to 28,188.
3. Removed a small number of trip legs (68) with vehicle types Motorbike, Taxi, and Other. Motorbikes don't really have "occupants" and also make minimal contribution to congestion; Taxis are more comparable to public transport which is necessarily excluded, and Other potentially includes drivers of buses. The remaining three vehicle types were: Car/station wagon, Panel van/Van/Ute/4 wheel drive, Truck[1].
4. Removed three apparently duplicate trip legs, and also many responses (36) from an outlier: one apparent bus driver (regularly driving large numbers of people, up to 33, in a non-household vehicle albeit with non-specified vehicle type).
5. Removed 230 trip legs which were 60+ km long. First, most long trips will be on roads outside our urban scope of interest. Second, results from small sub-groups with a few very long trips could be severely affected by extreme values in weighting (to be consistent with occupancy measures taken on the road and a

[1] We considered also eliminating the 744 trip legs with vehicle type Truck. However, these trips were recorded for a wide range of reasons including social/recreation and did regularly include household members as occupants (even when reportedly being driven "on employer's business"). In addition, in around a quarter of trip legs with vehicle type Truck the driver did not claim to have a current truck licence, which suggests the description of the vehicle as a truck may have been relatively informal. Indeed, the trip destinations/purposes recorded by such drivers (i.e. "truck" drivers without truck licences) were very similar to drivers of Panel vans/Vans/Utes/4 wheel drives (respectively 27% vs 29% home; 35% vs 35% work and employer's business; 10% vs 9% shopping). The pattern for truck trip legs driven by those with truck licences was quite different (12% home; 83% work and employer's business; 2% shopping).

good international practice, each trip in occupancy analysis here is usually weighted in proportion to its length).

In summary, the base (unweighted) has:

- 27,851 trip legs -- 18,150 trip legs by drivers, and 9701 trip legs by passengers. The trip legs by drivers comprised 11,427 driving alone, and 6,501 driver trip legs with passengers present (222 of driver trip legs did not record the number of vehicle occupants).
- 3467 people (1759 in Auckland; 705 in Wellington; 1003 in Christchurch).
- 1396 households.

1.2 Matching

To split total occupancy (TRPEOPL) into major components (household versus non-household passengers; children versus adults) some awkward preliminary data processing was necessary. This preliminary work was required because such fine and time-consuming distinctions were not directly recorded by drivers in the survey. Respondents were (often) independently recording key aspects of trips such as the time of departure, hence it was unrealistic to expect matches to the exact minute between different respondent's records of the same trip. In addition, it seemed quite possible that different occupants on a trip might well remember different minor stops or describe the same address differently. Hence, the matching was done in a way that allowed for some imprecision.

The 15 different match types used are documented in Table 1. Note that absolutely exact matches were dealt with first and hence given priority over all less precisely recorded matches. It was unrealistic to expect perfectly exact matches only in such a database -- this matching task is extraordinarily demanding, and not at all what the survey process was designed to achieve.

Table 1: Match types ("Y" below means exact match required)

Note: All match types concern people from the same household (SAMNO) and on the same day (DAYWK)

Imprecision in match	Vehicle*	Address leave	Address arrive	Time leave	Time arrive	Additional checks
0. Perfect match	Y	Y	Y	Y	Y	
1. 1 Time only	Y	Y	Y	Y	-	
2. 1 Time only	Y	Y	Y	-	Y	
3. Time	Y	Y	Y	Approx2**	-	Duration or distance***
4. Time	Y	Y	Y	-	Approx2	Duration or distance
5. 1 Place & Time	Y	Y	-	Y	-	Duration or distance
6. 1 Place & Time	Y	-	Y	-	Y	Duration or distance
7. 1 Place, Time	Y	Y	-	Approx2	-	Duration or distance
8. 1 Place, Time	Y	-	Y	-	Approx2	Duration or distance
9. Places	Y	-	-	Y	Y	Duration or distance
10. Places, Time	Y	-	-	Approx1	-	Duration or distance
11. Places, Time	Y	-	-	-	Approx1	Duration or distance
12. Vehicle (matches restricted to other **household vehicles****** only)	-	Y	Y	Y	Y	Duration or distance
13. Vehicle, Time	-	Y	Y	Approx1	Approx1	Duration or distance
14. Vehicle, 1 Time & Place	-	Y	-	Approx1	-	Duration or distance
15. Vehicle, 1 Time & Place	-	-	Y	-	Approx1	Duration or distance

* The LTSA variable names corresponding to the columns above were: VEHICLE, TRSTADD, TRADDNO, LEAVEHR, ARRIVHR.

** "Approx2" when referring to Times means up to 2 hours difference; Approx1 means up to 1 hour difference. Initially, we allowed only a 1 hour difference; but inspection of unmatched cases showed several cases where trip legs with over 1 hour difference in recorded times clearly should be matched (particularly in the weekends).

*** "Duration and distance" as an Additional check means that we have checked that the duration or distance of the trip legs recorded by the passenger is broadly consistent with the duration recorded by the driver. By "broadly consistent", we mean that durations up to 5 min are matched with durations no longer than 15 minutes, durations of 5 to 10 minutes are matched with durations no more than 2 1/2 times as long, and durations over 10 minutes are matched with durations no more than twice as long. (Slight adjustments were made for consistency at the "joins" between these 3 criteria.) We also checked that the passenger distance was not more than twice as long (nor less than half as long) as the matched driver distance. Given the vagaries of responses to these questions, we only deleted matches which failed in **both** the time and the distance consistency checks. This resulted in deletion of 39 matches.

**** "Household vehicles" are registered vehicles used by the household and usually parked there overnight (whether private or company-owned). It seemed reasonable to assume that passengers recording travel in such vehicles were usually being driven by members of their household, and if a reasonable match in terms of time and place existed it seemed possible that either driver or passenger did not accurately record the particular vehicle they were in. Clearly passengers travelling in **non-household** vehicles were much more likely to have non-household drivers (and hence excluded from these matches where the recorded vehicle differed between driver and passenger).

Table 1 above lists the different matches in a logical order, albeit an order in which some matches may well be better evidence of a good match than ones a few lines higher. The final hierarchy of matches (reflected by the order in Table 2) was decided only after

examining the crucial cases where different match types resulted in passenger legs being matched to different driver legs.

Table 2 shows that **96% of matches found were quite precise** (match types are 0 through 6). It is very arguable whether or not we should use the least accurate 5 match types at the very bottom of the table; but their inclusion will have little effect overall because they account for only around 1% of the total matches. Note also that the most common match types (0, 1,2,5,6) were separately produced from two different matching procedures and then reconciled to minimise the risk from errors in data processing.

Table 2: Final hierarchy of matches

Match type and imprecision		Number	%
0	Perfect match	5265	73.6
1,2	1 Time only	388	5.4
3,4	1 Time & 1 approx Time	169	2.4
5,6	1 Place & Time	1038	14.5
9	Places	32	0.4
12	Vehicle (but still a household vehicle)	38	0.5
7,8	1 Place, 1 Time & 1 approx Time	164	2.3
13	Vehicle, 2 approx Times	11	0.2
10,11	Places, 1 Time & 1 approx Time	27	0.4
14,15	Vehicle, 1 Time & Place	21	0.3
Total		7153	100.0

In addition to the types of imprecision summarised above, allowance was made for drivers incorrectly recording the number of vehicle occupants. In particular, trip legs where the driver recorded no passengers (TRPEOPL = 1) were initially excluded from possible matching to passenger trip legs. However, examination of cases where passengers claimed to be travelling in household vehicles but had not been matched showed that matches to such trip legs should be allowed (but we restricted this relaxation to match types 0 through 11; that is, excluding the match types where the vehicle was also wrongly recorded). This change resulted in around 100 extra passenger trip legs matched.

Table 3 shows that only 5% fell into the category indicating likely failure of matching ("Not matched, household vehicle"). **Hence this process successfully assigned around 95% of passenger trip legs.** This is not ideal, but seems quite sufficient given that the main analysis of occupancy will be based on the number of occupants directly recorded by the driver. The matching process described here provide an extra level of detail in **subsidiary** analyses only. The bias from failed matches will probably be minimal for two kinds of subsidiary analyses (household children, household adults), but greater for the third kind of subsidiary analysis (non-household passengers). Non-household passengers can only be estimated for each driver trip leg by subtracting the matched passengers from the total number of occupants recorded by the driver (TRPEOPL). This estimate of non-household passengers will be contaminated by the unmatched trip legs using household vehicles (around 1/5 of the combined total), and to a lesser extent by the small number of not matched cases where a non-household vehicle was actually been driven by a household member.

Table 3: Extent of matching for passenger trip legs

Match result	Number	%
Matched, household vehicle	7037	72.6
Matched, non-household vehicle	116	1.2
Not matched, household vehicle	448	4.6
Not matched, non-household vehicle	2093	21.6
Total	9694*	100.0

* This total is slightly different to the base number of passenger legs described earlier, because a handful of cases have been deleted (because of inconsistency becoming evident during the matching process) and because we changed the mode from Drive to Passenger in a few cases (because 2 drivers were recorded for the same short trip leg).

Near the end of the development of the matching procedure, 121 passenger trip legs then described as "not matched, household vehicle" were visually checked. The most common conclusions were:

- Twenty-four matches missed because driver incorrectly reported no passengers. Subsequently as described above, matches were allowed to driver trip legs where the driver had specified that they were the only occupant (i.e., TRPEOPL = 1).
- 20 legs reflected apparent forgetting (usually by the driver) of an extra leg in a journey remembered by the passenger. Given that correction of these cases would have required artificial creation of entirely new trip legs, no correction was attempted.
- In only a handful of cases the visual inspection did suggest a match to an existing driver leg missed by the existing automated matching procedures.
- 14 legs appeared plausibly to have a non-household driver in a household vehicle (because householders lacked drivers licences).
- 11 legs may plausibly have had a non-household driver in a household vehicle, because the householder was elderly.

2. STATISTICAL WEIGHTING FOR OCCUPANCY

2.1 Total km driven as base for weights

Not all trips are equal, that is, some are distinctly longer than others. Taking account of this in occupancy analysis appears important because of:

- Confounding. Some are trip types are consistently longer than others (e.g., some trip purposes may consistently be associated with short trip lengths), and some types of driver make longer trips on average. In particular, similar to results from other countries, men in our sample reported driving much further on average during the previous year (17,310 km) than women (9263 km). Such patterns could have confounding effects on analysis of occupancy by such variables unless distance is properly accounted for as part of weighting.
- Consistency. It is desirable that analysis of occupancy here be consistent conceptually with total occupancy as commonly observed from the roadside. Random observation of occupancy from the roadside will count trips in proportion to the distance, not equally for each trip leg or person (other things being equal).

For example, consider a sample frame consisting of only 2 vehicle trips; the first trip was 2 km with 5 occupants, and the second trip was 100 km with only 1 occupant. Simple analysis with trip legs as the base, would result in calculated **occupancy of 3,** i.e., (5+1)/2. Clearly, a superior measure of occupancy on the roading system (e.g. as would be observed from observation points randomly chosen in time and space) would take account of the distance and would result in an **occupancy close to 1**, e.g., (5*2+ 1*100)/102 =1.08.

Hence, our statistical weight multiplies the usual weight[2] by the distance recorded for each trip (by the driver). Similarly, in later analyses based on results summarised for each driver, the usual weight is multiplied by the total distance driven by the person.

The main risk resulting from the inclusion of distance in the weight is that dangerously extreme values may be created. For example, the occupancy results for a small subgroup may be largely determined by only one or two cases that happen to have extremely high distances and hence high weights. In the present study is risk was reduced by restricting the analysis to trip legs shorter than 60 km long (longer trip legs will generally involve travel outside the urban area of interest here). In addition, parallel unweighted results for major findings were also examined to check that results were unlikely to be a misleading artefact of weighting.

[2] The usual weight is based on the person post-stratification weights supplied by the LTSA with the database. This aligns the weighted sample to the estimated population data (derived from Statistics New Zealand Census results) for December 1997 by the same region, age group, and gender classifications. Consistent with LTSA practice for the nationwide sample, we adjusted the person weight so that there was an even distribution of weighted person travel days across the days of the week (our adjustment for this was somewhat different insides although identical in principle because we are analysing a sub-sample only from the three major urban areas).

Note that including distance in the measurement of vehicle occupancy is general practice in the major US travel study:

> "For NPTS data, vehicle occupancy is generally computed as person miles of travel per vehicle mile (referred to as the travel method). Note that the other commonly-used definition of vehicle occupancy is persons per vehicle trip (referred to as the trip method).... Because longer trips often have higher occupancies, the travel method generally yields a higher rate (1.59 for the 1995 NPTS) than the trip method (1.50)." Hu & Young (1999, p. G-4).

Hu, Patricia S. and Jennifer R. Young. **Summary of Travel Trends: 1995 Nationwide Personal Transportation Survey.** Washington, D.C.: Federal Highway Administration, 1999.

e:\client.cs\frst\occupancy\report\method1a.doc
12 June 2003