Data Inconsistencies Detection and Correction: an Application to Identifying Joint Household Travel from the Household Travel Survey

Chinh Ho*, Corinne Mulley

Institute of Transport and Logistics Studies,

The University of Sydney Business School, New South Wales, Australia

Kerry Shaz

Bureau of Transport Statistics,

Transport for NSW, New South Wales, Australia

***** Email: mailto:quoc.ho@sydney.edu.au

1 Aims and objectives

Intra-household decisions mean the travel decisions of each household member are not necessarily independent of the travel behaviour of other household members. The existence of intra-household interactions giving rise to joint engagement of household members in activities and travel is widely acknowledged. Intra-household interactions and travel have received limited attention, although the body of research is growing, as seen by special issues of Transportation [1] and Transportation Research [2]. Quantitative investigations are limited by the absence or restricted availability of information on participating household members in activity based travel surveys [3].

Research to date has identified intra-household joint activities and travel by connecting the trip records of each household member, based on a set of matching criteria. Inconsistencies in reported times and locations between/among household members questions how effective sets of matching criteria can be in identifying joint household travel.

This paper has three aims: first, to examine the ability of flexibly defined sets of matching criteria to identify joint household travel; second, to offer a method for detecting and correcting data inconsistencies; and finally, using the Sydney Household Travel Survey, to provide a detailed analysis of inconsistent cases and thus to suggest improvement to activity-travel surveys.

2 Methodology

Using three waves of the Sydney Household Travel Survey data (2007/08, 2008/09 and 2009/10) and choosing only fully responding households, this paper identifies joint household travel for 7,286 households. To test the best method of identifying joint household travel, including those reported inconsistently by household members, four sets of matching criteria are developed using algorithms based on reported travel mode, origin and destination, number of household members in the car, and a flexible joint-in-time condition respectively. This produced four datasets for further evaluation.

Information on the number of household members travelling together is available for car trips only and this required separate evaluation for car and non-car trips. For car trips, an algorithm method is used to test the reliability of information on the number of persons travelling together. Four measures were used to evaluate and identify the best set of matching criteria: percentage of identifiable household joint drivers and passengers, households where members report inconsistent trip details, and misidentified trips. The first two measures are adapted from Shaz and Corpuz [4].

Assuming the best matching criteria for car trips applies also to non-car trips, a graphical method is used to identify misidentified joint household non-car trips. By graphically illustrating household members' travel diaries, the travel party size can be identified which can then be compared by the travel party size identified by the algorithm method.

In correcting for inconsistencies, three assumptions about joint household trips are made:

1. If two household members jointly travel and there is a difference in trip details, the details reported by the car driver are more reliable than the car passenger;
2. If more than two household members travel together, the trip details consistently reported by more members are assumed to be the consistent details;
3. If more than two household members travel together and there is no larger group of consistently reported details on trips, the trip details made by the group including the car driver is assumed to be more reliable.

'Inconsistent' in this context means inconsistent reporting of trip details between household members. These assumptions allow the trip details of the joint travel to be aligned, according to rules which are reflective of reality. Inconsistencies are corrected by updating inconsistent records with the values reported by the co-traveller(s) who, according to the assumptions above, are considered consistent. The best matching criteria were applied again to the corrected dataset to identify the improvement in correctly identifying joint travel.

3 Results

Table 1 shows the performance of the four sets of matching criteria in identifying joint household travel. These successfully identified more than 93% and 92% of the expected number of car drivers and car passengers respectively. The more flexible the matching criteria, the lower the level of under-identified joint household trips, but this was found to be at the expense of independent trips being over-identified as joint trips. The set of matching criteria (2) is selected as the best set following trade-offs between the performance measures where the ability to avoid over-identified trips is heavily weighted because their correction is more difficult.

Table 1 Performance of the Four Sets of Matching Criteria

Matching criteria with a joint-in-time condition using	Performance of matching criteria in terms of four measures				
	Measure 1[a]	Measure 2[b]	Measure 3[c]	Measure 4a[d]	Measure 4b[e]
(1). 5-min intervals, starting and ending	93.2%	92.2%	10.0%	56	1,966
(2). 10-min intervals, starting and ending	93.9%	92.9%	9.3%	79	1,807
(3). 10-min intervals, starting	94.8%	93.9%	8.2%	104	1,560
(4). 10-min intervals, ending	95.1%	93.8%	8.6%	129	1,569

[a] Percentage of identifiable number of household joint driver trips

[b] Percentage of identifiable number of household passengers

[c] Percentage of households for which members have reported inconsistent trip details

[d] Number of joint household trips being over-identified

[e] Number of joint household trips being under-identified

Total number of person trips is 81,731

The graphical method was applied to a 1% 'randomly' selected sample of households. Figure 1 illustrates how joint household travel is identified using a household with five members. Travel mode and travel party size identified by the best set are shown in the bottom right-hand corner box as comparison. For this household, the two methods provide the same results. Overall, a comparison of the results between the best matching criteria and the graphical method show joint non-car trips significantly less likely to be under-identified and more likely to be over-identified than car trips.

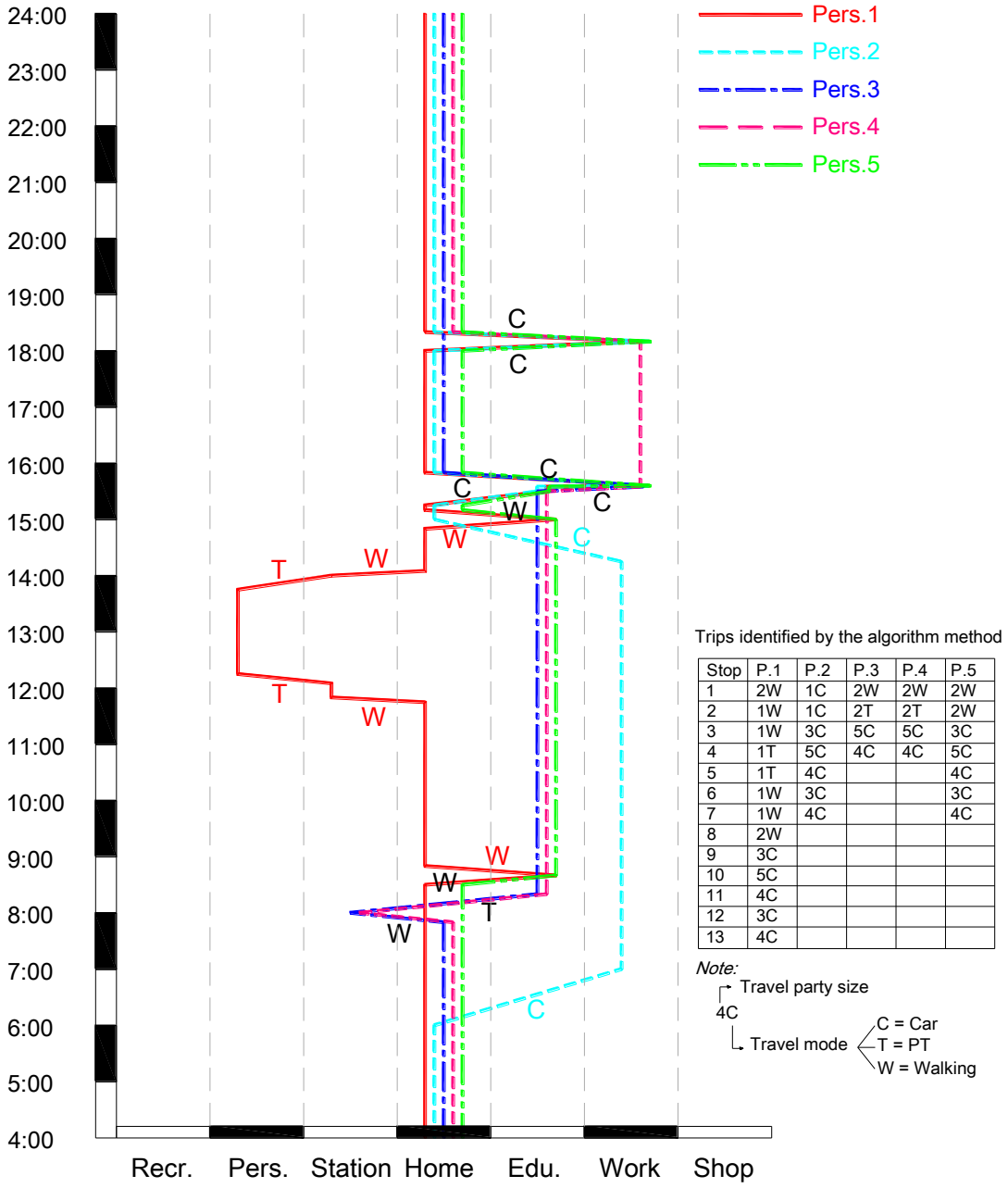

Figure 1 Visualisation of Travel Patterns of an Example Household

Analysis of inconsistent cases showed departure times were reported more consistently than arrival times. Inconsistent reporting of trip origins and destinations each accounts for about 35% with 30% of inconsistent cases arising from discrepancies in departure and arrival times. Segmentation analysis revealed trip details by females, non-workers, and children (aged 5 to 14) were reported significantly less consistently. Further, weekend trips were reported significantly less consistently than weekday trips. This maybe explained by the more complex travel patterns of females, non-workers, children and weekend days.

4 Conclusions

One-day travel diaries are often the only source of data available for investigating travel behaviour. It is therefore important to identify their deficiencies, especially if simple ways can be proposed to improve their usefulness in travel behaviour understanding.

Joint household travel is an important component of understanding household travel behaviour and this paper sheds light on the accuracy of using matching criteria to identify joint travel in activity travel surveys. A method to detect and correct data inconsistencies is presented and demonstrated to improve the performance of flexibly defined matching criteria in identifying joint household travel. The analysis of the paper also offers practical suggestions for understanding the reliability of traditional trip diaries since inconsistencies arise not only in departure and arrival times but also in trip origin and destination. Moreover, as complex travel patterns suffer from higher incidence of inconsistency, allocating relatively more resources for their validation could improve the accuracy of activity-travel surveys.

References

[1] Bhat, C. R. & Pendyala, R. M., "Modeling intra-household interactions and group decision-making", Transportation 32, 443-448, 2005.

[2] Timmermans, H. J. P. & Zhang, J., "Modeling household activity travel behavior: examples of state of the art modeling approaches and research agenda", Transportation Research Part B: Methodological 43, 187-190, 2009.

[3] Kang, H. & Scott, D., "Impact of different criteria for identifying intra-household interactions: a case study of household time allocation", Transportation, 38, 81-99, 2011.

[4] Shaz, K. & Corpuz, G., "Serving passengers - are you being served?" The 4th Annual PATREC Research Forum, 1-15, 2008.