The Longitudinal Study of Australian Children:

an Australian Government initiative

Children's time use in the Longitudinal Study of Australian Children: Data quality and analytical issues in the 4-year cohort

Technical Paper No. 4

Jennifer Baxter
July 2007

Australian Government

Australian Institute of Family Studies

Growing Up in Australia: the Longitudinal Study of Australian Children.

In 2004, over 10,000 children and families around Australia agreed to take part in *Growing Up in Australia*, the longitudinal study of Australian children (LSAC). This study is designed to identify policy opportunities for children and their families and for early intervention and prevention strategies.

This longitudinal study involves two representative cohorts of children and their families – approximately 5,000 infants aged 0-1 years (B or infant cohort) and 5,000 children aged 4-5 years (K or child cohort) when the families agreed to take part in 2004. It is following the development of these children until 2010 and possibly beyond.

Growing Up in Australia was initiated and funded as part of the Australian Government's *Stronger Families and Communities Strategy* by the Australian Government Department of Families, Community Services and Indigenous Affairs. The study is being undertaken in partnership with the Australian Institute of Family Studies, with advice being provided by a consortium of leading researchers at research institutions and universities throughout Australia. In Wave 1, the data collection was undertaken by I-view, in conjunction with Colmar Brunton Social Research. From Wave 2, the data collection is being undertaken for the Institute by the Australian Bureau of Statistics.

Table of Contents

About the author

Dr Jennifer Baxter is a Research Fellow at the Australian Institute of Family Studies. Jennifer's interests lie in the analyses of the ways in which work and family interact. Since commencing at AIFS, she has used quantitative techniques to explore mothers' return to work and work-family spillover for parents of young children. She has also undertaken extensive analysis of children's time use and parental time with children, making use of the Time Use Data from *Growing Up in Australia: the Longitudinal Study of Australian Children* (LSAC).

1 Introduction

The first wave (2004) of the Longitudinal Study of Australian Children (LSAC) included children's time use diaries for two cohorts of young children, one aged around one year old and the other around four to five years old. These data have proved immensely valuable, especially given the uniqueness of these data, and the large sample of children from which these data were collected. There have, however, been some difficulties in analysing these data, with a significant issue being the amount of missing data. The purpose of this paper is to analyse this missing data for children in the 4-5 year cohort. This is done in order to better understand the relationships between missing activity data, children's participation in child care, preschool or school and parental employment. As other issues emerged in undertaking these analyses, this paper also provides an overview, and highlights some of the strengths and weaknesses of these TUD data.

The Time Use Diary (TUD) collects details of the activities of the study children in LSAC over two 24-hour periods, one a specified weekday and one a specified weekend day. While data were collected of children in both the infant and 4-5 year cohorts, this analysis is restricted to the elder cohort. This is because more significant problems with missing activity data were anticipated for this cohort, given that these children were more likely than the infants to spend a portion of their day in non-parental care, when missing activity were most likely to occur.

The paper is structured as follows. First, in Section 2, is an overview of the diary and the resulting data. Following that, Section 3 examines data quality issues, including missing data. Section 4 then presents an option for imputing missing activity data, and Section 5 presents analyses of the TUD data by time of day. The analysis of the average time in activities is presented in Section 6, with a discussion of how important it is to consider the treatment of missing data when doing these types of analyses. This is followed in Section 7 by analyses of these mean time data to identify factors associated with more or less time spent in different activity groups. A discussion of issues and a list of recommendations for analysts concludes this paper.

The LSAC technical paper series provides detailed insights to the survey design and data. Papers available to date cover the sample design of LSAC (Soloff, Lawrence, & Johnstone, 2005), the outcome index (Sanson et al., 2005), and weighting and non-response (Soloff, Lawrence, Misson, & Johnstone, 2006).

2 Data

2.1 Overview of the Time Use Data

After the LSAC personal interview, the respondents were left with some self-complete forms, including the Time Use Diaries. The interviewer worked through an example of how to complete the diary with the respondent, and the respondent was advised of the dates for which they should complete the diary. These dates were selected by the interviewer to ensure a random allocation of weekdays and a random allocation of weekend days. The respondent was asked if they could not complete the diary on their allocated date to wait another week before completing it, such that the completion day was on the same day of the week as was the date selected for them. The diaries were collected by the interviewer at an agreed-upon later date, or mailed back in some cases (a mail-back envelope was provided).

The diaries divided the 24-hour day into 96 15-minute intervals. The day started at 4am on the diary day and finished at 4am the following day. Parents were asked to mark the times in which their child was involved in any of 26 pre-coded activities. For this analysis, the activities were grouped into broader categories, as shown below:

Table 1 Classification of activities

Activity category	Coded activity
Sleeping/resting	Sleeping/ napping, awake in bed, do nothing/ bored/ restless.
Personal care	Bathe/ dress/ hair care/ health care, eating/ drinking/ being fed, held/cuddled, crying/ upset, being reprimanded/ corrected, destroy things/ create mess.
Leisure	**Exercise**: Walk/ ride bike/ other exercise. **Achievement**: Colour/ look at book/ educational game, use computer, read a story, talked to/sung to, taught to do chores or read. **Other leisure**: Watching television, movies, listening to tapes, radio, music, other play/ other activities; visiting people/ special event/ party, organised lessons/ activities.
Travel/ taken places	Taken places with adult, taken out in a pusher or bicycle seat, travel in a car or on public transport.
Unknown	Unsure what child was doing (where not coded to another activity), no activity coded.

While other published work on children's time use was consulted in developing this classification (for example, Yeung, Sandberg, Davis-Kean, & Hofferth, 2001), close correspondence was not possible, given differences in the data collection. There is a fair degree of subjectivity around how these data are assigned to broad groups (for example, should 'destroy things/create mess' be coded as 'leisure' rather than 'personal care'?), and for

some activities, the detail of what is captured in the underlying data is not clear, in particular in relation to 'other play/other activities', 'organised lessons/activities' and 'taken places with adults'.

The diary data were analysed by assuming the activity recorded in each time period lasted for the full 15 minutes. This may result in an overestimation of time spent in specific activities, when those activities take less than 15 minutes. For example, cleaning teeth before bed may be recorded in one 15-minute block, and therefore allocated 15 minutes, while in effect the actual duration was much shorter. Children could be coded to a number of activities concurrently, which meant that the sum of time spent in different activities could be greater than 24 hours. Unlike some adult time use surveys, the diary did not differentiate between the main activity being undertaken (primary activity) and any activities being undertaken concurrently (secondary activities).

Information was collected on where the child was in each time period. Possible categories were own home (indoors), other person's home (indoors), day care centre/playgroup, other indoors and other outdoors. Preschool and school were not specifically identified in the response categories, so it is not clear whether children in these structured early education environments would have been recorded as being in 'day care centre/playgroup' by the parents, or whether they would have been coded to elsewhere. A likely place is 'organised lessons/activities' in the activity coding.

Pre-coded context information was collected on who the child was with, defined as in the same room, or if outside, nearby to the child. Clearly this is not a complete measure of who was responsible for or caring for the child, as parents or others can do this from another room in the house. Further, it does not reflect the intensity of caring, as it does not distinguish between those who are interacting with the child, and those nearby but involved in some other activity. Parents could mark whether their child was with their mother (including step-mother), father (including step-father), with other adult relatives, with other adults, with siblings or other children, or alone.[1]

---

[1] Measures of parental time spent with children using time use data are described in Budig and Folbre (2004) and Craig (2006). Estimates of time spent with children based on primary child care, primary plus secondary child care and time with children are given in Craig (2006).

For each time period, information was also collected on whether a pet (excluding fish) was with the child, and whether the activity was paid for. These data have not been analysed.

2.2 Data source

The dataset used in this analysis was created from the first release of the 4-year cohort of the LSAC Wave 1 TUD data, and is described as Version 1. Since undertaking this analysis, Version 2 of the TUD has been released. As discussed further in Section 3.1, Version 1 had less extensive data edits applied than did the later version (LSAC Project Operations Team, 2006).

As with the entire LSAC data collection, the diary collection was spread over a number of months of 2004. Most diary dates were between March and August 2004, and some more through September to November. This collection, therefore, does not cover the summer period, when children's activities may be somewhat different. (For some analyses, it may be worth considering differences in activities by month or season.)

The dataset contains diary weights that adjust the sample weights according to the distribution of days of the week. Because the diary days were not distributed equally over the seven-day week (especially with regard to weekday versus weekend), the diary weights should be used in analyses of these data. Further, analyses should take into account the survey design of LSAC. To simplify the analytical process, this paper uses unweighted data, keeping weekday and weekend analyses separate. Future work with the data will modify these analyses by taking into account weights and survey design.

2.3 Characteristics of the sample

The TUD sample comprised 3,813 weekday diaries and 3,636 weekend diaries. The weekday diaries were fairly equally distributed over Monday through to Friday. The weekend diaries had a slight bias towards Saturdays (56 per cent were Saturdays).

Most diaries were completed by the child's mother (91 per cent), with seven per cent completed by the child's father. The remaining two per cent were completed by other family or carers.

About 76 per cent of diaries were completed on the diary day. Some filled out the diary more than three times a day (26 per cent of respondents), others two or three times a day (28 per cent) and others when the child went to bed (22 per cent). Another 11 per cent filled the diary

out the next morning and 12 per cent filled it in at a later time. Some respondents reported that they filled the diary in for a day considerably later than the date they were supposed to (55 records were completed more than 30 days after).

Respondents were asked to describe the diary day, by selecting one or more of a number of pre-set options (see Table 2). The most commonly selected option was that this was an ordinary day (78.2 per cent of weekday diaries and 67.7 per cent of weekend days). On weekends a considerable proportion reported that the diary day was a holiday or family celebration (12.5 per cent). All respondents were included in the analyses in this paper, even those who said this was not an ordinary day.

Table 2 Description of the diary day

	Weekend (%)	Weekday (%)
An ordinary day	67.7	78.2
A holiday or a family celebration	12.5	4.4
A school holiday for brother/sister	4.1	6.2
A parent took time off work	1.7	4.2
Our family dealt with a crisis	0.8	0.5
We had guests staying with us	6.2	3.2
A family member was away	3.8	2.4
I was ill	1.5	1.7
This child was ill	3.2	3.7
This child was a great deal more stressed than normal	0.7	0.6
An unusual day for another reason	14.1	12.0

Note: Respondents could select more than one of these categories.

To analyse the relationships between time use and other parental, family or child characteristics, the TUD data were matched to other Wave 1 data. Children were aged between 4 years 3 months and 5 years 7 months at interview, with an average age of 4.8 years. The sample was equally divided between girls (49 per cent) and boys (51 per cent). Only nine per cent of the children were an only child, 50 per cent had younger siblings (of which some also had older siblings) and 41 per cent had older but no younger siblings. Most children (86 per cent) were from couple-parent families. Of couple families, almost all the fathers (95 per cent) were employed, and only six per cent were employed part-time (that is, working less than 35 hours a week). Of couples, 42 per cent of mothers were not employed, 44 per cent worked part-time hours and 14 per cent worked full-time hours. Of the single parent families, most were headed by a mother (93 per cent).

The measure of parental employment was the usual weekly hours worked. The actual parental employment arrangements for the diary day were not known, so it was not possible to determine if, or at what times the parents worked on the diary day. Instead, parental employment details were sourced from other Wave 1 data, and therefore referred to the work arrangements at the time of these other collections.[2] Parents who worked full-time hours were fairly likely to have worked on the diary weekday but those classified as part-time were likely to include a mix of those who worked on this day and those who did not.[3]

2.4 Response bias

In the 4-5 year cohort, 77.6 per cent of respondents completed at least one diary, using those who responded to the Wave 1 interview as a base (Table 3). The majority completed one weekday and one weekend diary.[4] Overall, 3,582 respondents completed two diaries and 285 completed one diary.

Table 3 Diaries completed for LSAC Wave 1 respondents, 4-5 year cohort

	Number	Per cent
No diaries completed	1,116	22.4
At least one diary	**3,867**	**77.6**
1 weekend and 1 weekday diary	3,485	69.9
2 weekend diaries	20	0.4
2 weekday diaries	77	1.6
1 weekend diary only	111	2.2
1 weekday diary only	174	3.5
Total number of respondents	**4,983**	**100.0**

Using as a response rate the proportion completing at least one diary, the response rate was higher in couple families (79 per cent), compared to single parent families (67 per cent) and in families with more highly educated mothers (84 per cent bachelor degree or higher, down to

---

[2] Some details were collected in the Wave 1 interview and others in the self-complete component. In 89 per cent of cases the diary date was in the same month as the initial interview, and in 10 per cent of cases it was in the following month.

[3] According to the Australian Bureau of Statistics Working Arrangements Survey, at November 2003, of women employed full-time, 76 per worked Monday to Friday while of women employed part-time, 21 per cent worked Monday to Friday, although another 39 per cent usually worked weekdays only.

[4] While respondents were supposed to complete one weekday and one weekend diary, some children had two weekday diaries and some two weekend diaries, instead of having one of each. Problems with the recorded date and day of interview may have contributed to this. Interviewers were to fill in both date of diary completion and the corresponding day of the week prior to giving the TUD to the respondent. On analysing the data it was discovered that the day of diary completion did not always correspond with the date. See Appendix 1.

70 per cent those with incomplete secondary education only). Families in which the mother worked full-time had a lower response rate than others (72 per cent), but those working part-time had a higher response rate (82 per cent) than those who did not work (76 per cent).

The response bias can be considered by analysing the likelihood of responding through multivariate analysis, with controls for various family characteristics (Table 4). As well as confirming the effects described above, the logistic regression showed that response rates were lower when the primary carer was younger and when the study child was older. The number of children in the family and the sex of the study child were not significantly related to response rates.

Table 4 Time Use Diary respondents, response rates and logistic regression results

	Response Rate (per cent completed at least one diary)	Logistic regression (completed at least one diary=1) Odds Ratio
Primary carer's education		
Incomplete secondary only (reference)	70.0	1.00
Incomplete secondary and diploma/certificate	73.2	1.12
Complete secondary only	76.7	1.30*
Complete secondary and diploma/certificate	81.6	1.69***
Bachelor degree or higher	84.2	1.87***
Primary carer's employment		
Not employed (reference)	75.9	1.00
Part-time hours (<35 hours)	81.5	1.13
Full-time hours (>=35 hours)	72.1	0.65***
Number of children in the family		
1 child (reference)	74.8	1.00
2 children	80.6	1.12
3 + children	74.8	0.82
Family type		
Couple (reference)	79.4	1.00
Single parent	66.7	0.62***
Sex of child		
Boy (reference)	78.0	1.00
Girl	77.2	0.94
Age of study child (months), centred at average of 57 months	-	0.94**
Age of primary carer (years), centred at average of 35 years	-	1.03***
Constant	-	3.07**
Sample size (number of respondents)		**4,925**
Adjusted R-square		**0.04**

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.
Note: Reference categories are those omitted in the logistic regression.
            The odds ratio of the constant term indicates, for those in all reference categories and
            average age, the probability of responding, calculated as 3.07/(1+3.07), equal to 75 per cent.
*** p <0.001, ** p <0.01, * p<0.05

3 Data issues

3.1 Initial data edits

There were some data quality issues with the first release of time use data used in this analysis. These related both to the occurrence of missing data (when the activity, 'who with' or 'where' data were not provided for a time period) and of 'false positives' (where over-sensitive scanning of the diaries recorded some positive entries in places where a null entry should have been recorded). While work on this paper was underway, the LSAC Project Operations Team was doing further edits of the data to address these problems, with input from Jude Brown and Michael Bittman of University of New England. The work done in preparation of this paper fed into that process. Since then, a second release of these data has become available, and analysts should use these data and note the edits that have been applied to improve data quality, as discussed in the User Guide (LSAC Project Operations Team, 2006). As these edits had not been completed before this work was done, data edits had to be applied to this dataset to address these problems. Most were the same as, or similar to, those used in the creation of Version 2 of the TUD data. The edits applied here and the effects on the amount of missing data are shown in Appendix 2.

3.2 Missing data

Despite these additional data edits, the existence of missing data remained a significant problem. Missing data occurred in the TUD in two ways. Firstly, it occurred when the activity, 'who with' or 'where' details were not completed for a period. Secondly, for activity data, it occurred when the activity was recorded as 'not sure what child was doing'. An important question addressed by this paper, is to what extent the missing data was related to children being away from the parents, for example, due to parental employment or due to the child's involvement in early education. For this time, while parents should usually have been able to code where the child was or who they were with, they may not have felt they could say exactly what the child was doing. With regard to these activity data, an interviewer instruction advised interviewers to tell respondents, if required, that it was alright for them to code this time as 'not sure what child was doing' although suggested that respondents could seek information from the carer, to enable them to fill in the day's details. The extent to which these details were collected is discussed in Section 3.3.

Table 5 shows the distribution of missing data, for activity, 'where' and 'who with' across the 24-hour day after the edits discussed in Appendix 2.

Table 5 Missing data distribution after data cleaning, all diaries

		Missing activity data	Missing 'where' data	Missing 'who with' data
Minutes of missing data in a day	Mean	113	232	119
	Std. Dev.	215	355	292
As per cent of total time (1,440 minutes)	*Per cent*	*7.8*	*16.1*	*8.3*
Diaries with data all missing data	Number	19	135	97
Diaries with no missing data	Number	4,041	2,360	5,257
As per cent of all diaries (7,449 diaries)	*Per cent*	*54.2*	*31.7*	*70.6*

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

An approach used by some analysts, when using these time use data, is to exclude records with more than 90 minutes of missing data (for example, Baxter, Gray, Alexander, Strazdins, & Bittman, 2007). For analyses of activity data, this means excluding just over 2,000 diaries, of which two-thirds are weekday diaries. In the following sections, I show that there are some factors associated with the presence of missing data, which means that by excluding a significant number of diaries based on this 90-minute 'rule', the resulting sample is biased, and there are other approaches that reduce the need to remove so many diaries.

To help understand what this missing data relates to, Figure 1 shows the proportion missing by time of day and weekend/weekday.

Figure 1 Proportion of diaries with missing activity data by time of day
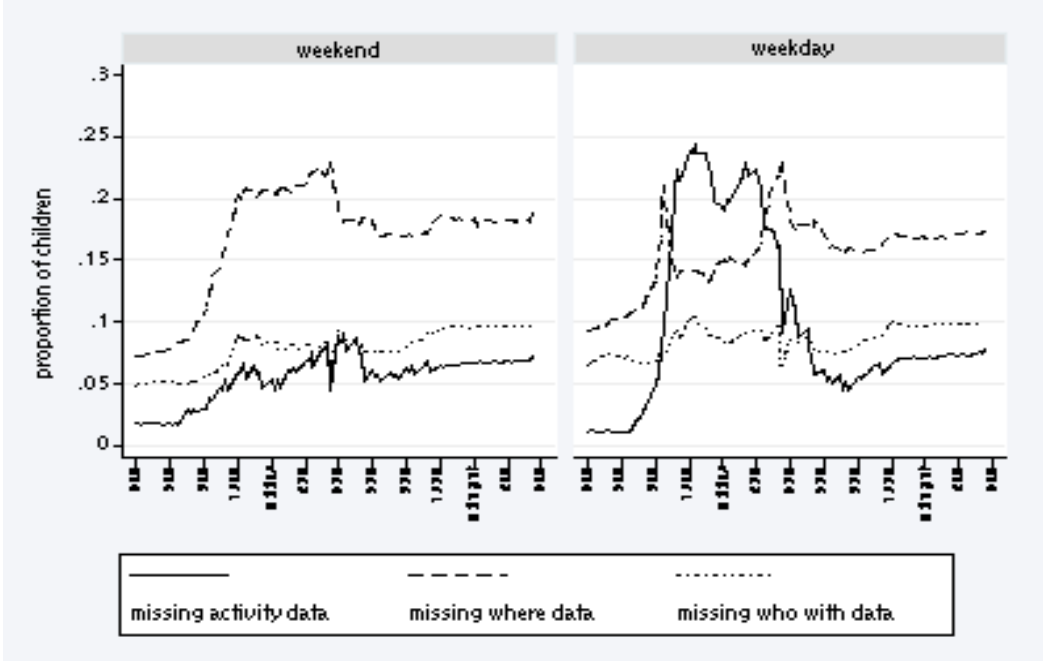
Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

Clearly, missing activity data were most likely on weekdays in the middle of the day, when children were most likely to be away from their parent/s. However, across weekdays and weekends there were missing data at other times as well.

In addition to this peak in missing activity data in the middle of the day, the likelihood of activity, 'who with' or 'where' data being missing was considerably greater at the end of the day than it was at the beginning of the day. For example, less than ten per cent of 'where' data were missing in the early morning of the diary day, but around 20 per cent was missing by the end of the diary period (4am of the next day). Similar differences occurred for the other data. It is likely that this was related to respondent fatigue.

Missing 'where' data were spread over the day, although there were peaks on weekdays coinciding with the beginning and the end of the work-day, and on weekends, there was more missing data between about 10am and 4pm. On a weekday, these times aligned with travel time (see Figure 5) – this makes sense, as parents may not have felt that they could exactly state where the child was. The reason for the large amount of missing 'where' data on the weekend is less clear, but may be related to the child being with someone other than the respondent, for example, with the other parent, at this time (see Table 9 for how missing activity data were related to the child being with the other parent).

3.3 Non-parental care

As shown above, the time when activity data were most likely to be missing aligned with the time of day when parents were most likely to be in employment, and/or children were likely to be at preschool, school or child care.

Figure 2 shows the extent to which missing data were related to children being in day care or in other types of care. This shows the total percentage missing (as was given in Figure 1) and this missing proportion classified according to whether the child was also recorded as being in day care/playgroup ('missing and at day care'), not in day care but with other (non-parental) adults ('missing and with others') or other ('other missing'). On the weekend, being in non-parental care explained very little of the missing data. On the weekday, however, a large amount of the daytime missing data corresponded to time children were in day care. Also, there was some missing data associated with other adult care. If we take these missing data away, the distribution of missing data looks much like that of the weekend, and of the missing 'who' and 'where' data, as shown by the "other missing" line in this chart.

Figure 2 Missing activity data by context and location
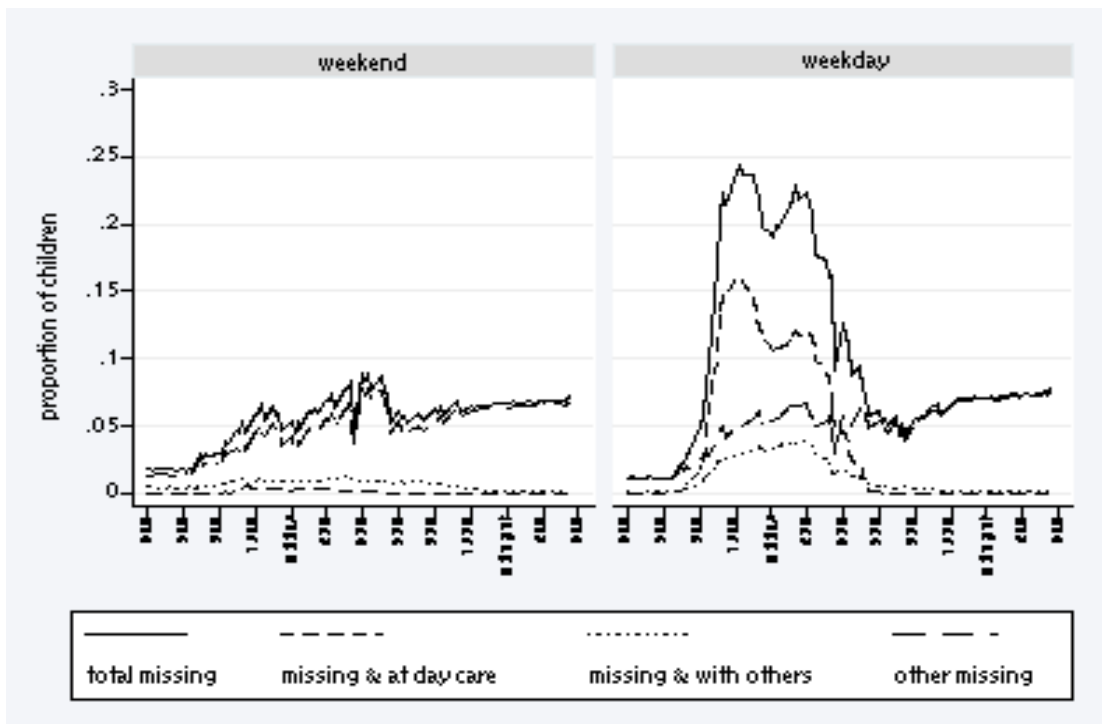
Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

There is evidence, then, that a child's involvement in day care, and to a lesser extent other non-parental care, was associated with less detail being provided for the times that children were in care.

When children were in non-parental care (that is, in day care or with other adults), they did not always have missing activity data for these times. Of 3,736 children, 2,482 (66 per cent) were in non-parental care for some time on the weekday (including 41 per cent in day care/playgroup and 25 per cent in other adult care only – see Table 6). For about half of these children, the respondent provided some activity details for that time, including seven per cent who provided activity details for the whole time the child was in non-parental care.

To what extent did the reporting of activities differ for those who used some non-parental care? Figure 3 shows, for weekdays, the proportion of children reported to be in sleep, leisure, personal care and travel/taken places by whether or not non-parental care was used at all on that day. This includes only non-missing data, that is, if an observation had activity information missing for one or more time periods, that observation was excluded from the analysis for that/those time periods that are missing.

Figure 3 Activities by care types, weekdays
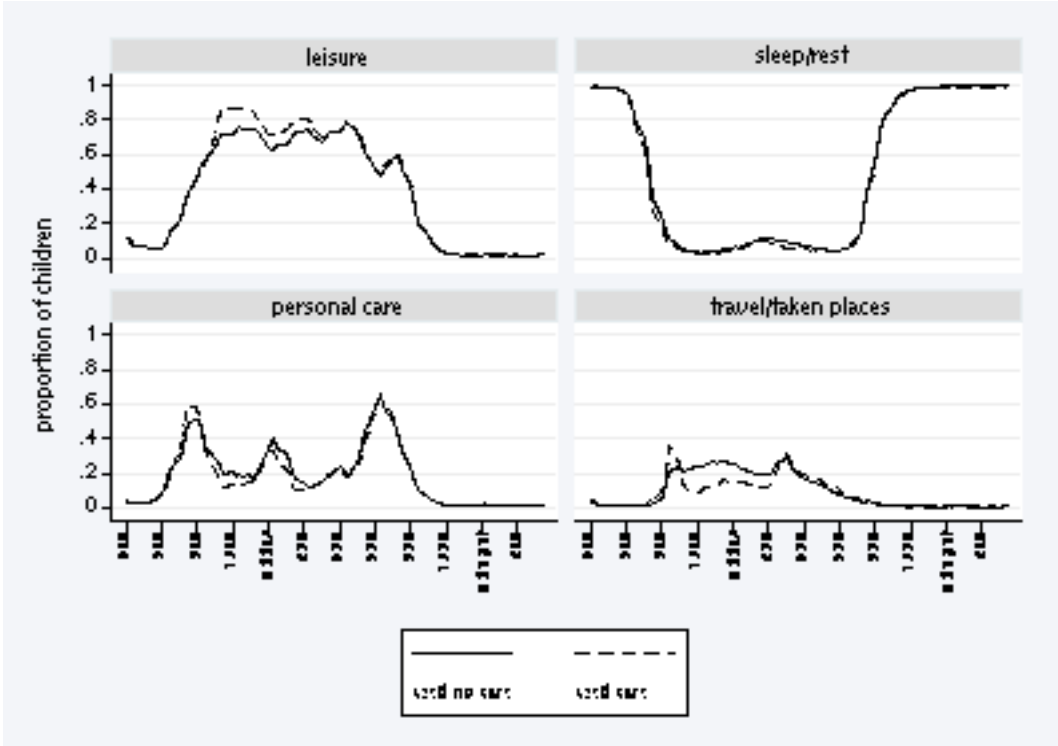
Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

Those who used non-parental care, who reported their child's activities, were slightly more likely to report that the child was doing 'leisure' activities during the day. This possibly relates to those children not in non-parental care having been more likely to be 'travel/taken places' during the day. For those in non-parental care, the 'travel/taken places' was most likely at the beginning and end of day, while for other children 'travel/taken places' was more uniform over the day, no doubt associated with a greater likelihood of being 'taken places' with adults throughout the day. There were small differences in patterns of sleep and personal care.

3.4 Parental employment and non-parental care

The issues related to parental employment and recording of child's activities are no different to those discussed above on non-parental care. In fact, the non-parental care issue is the more significant one because, for children of this age, many were in non-parental care (preschool or school) even when there was a parent not in paid employment (Table 6). Amongst families with a not-employed mother, 60.1 per cent reported the child to be in non-parental care at some time (21.2 per cent day care, 47.2 per cent other adult care). When the mother worked part-time the figure was 68.6 per cent (including 34.6 per cent day care) and when the mother

worked full-time the figure was 79.3 per cent (including 45.9 per cent day care). The time in non-parental care also differed according to parental employment.

Table 6 Non-parental care and parental employment on weekdays

	Primary carer usual weekly hours in paid employment			
	Not employed	<35 hours	35 hours or more	Total
	Percentage in day care/other adult care on 24-hour day			
Percentage of children in non-parental care for some time on the diary day	60.1	68.6	79.3	66.4
in day care	*21.2*	*34.6*	*45.9*	*30.4*
in other adult care	*47.2*	*53.6*	*65.6*	*52.5*
	Minutes in a 24-hour day			
Mean time child was in non-parental care on the diary day	165	228	324	214
in day care	*92*	*132*	*184*	*122*
in other adult care	*73*	*96*	*140*	*92*

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

It is interesting to note the relatively high percentage in 'other adult' care for this age group. For 4-5 year olds, few children were in informal care only, with most attending preschool, school or formal child care (Baxter, Gray, Alexander, Strazdins, & Bittman, 2007). It is unlikely, then that, this 'other adult' care refers solely to informal care. The category 'day care/ playgroup' was to capture the whereabouts of children when in formal day care (or playgroup), but parents may not have felt this was the appropriate place to code their child's location when they were in school or preschool. This 'other adult' care may then include children in preschool or school. One way to check for this is to tabulate the child care use, as recorded in the personal interview, by the 'where' data on the diary day. We would not expect an exact match, since children did not necessarily attend care or preschool on the diary day, but some associations may be apparent. Table 7 shows these data.

Of those who reported attendance at a child care centre, 59.4 per cent had some time in 'day care' on the diary day (Table 7). These children were the most likely to have time recorded in day care. The next most likely were those who used other formal child care or who used preschool. Children who attended school, and who would have been the most likely to be attending every day, were less likely to have time recorded as 'day care', but they were more likely to have time recorded as 'other inside'. It appears that the 'day care' data may not have captured children in early education settings. (Also, it should be noted that 'playgroup' is quite a separate activity, where parents almost always remain in attendance – this should possibly not be included in this category alongside 'day care' as it is now.)

Table 7 Child care use compared to TUD 'where' data, weekday 9am to 6pm

Child care use	At home	Other person's home	Daycare/ playgroup	Other inside	Other out- side	Missing	Total
							Sample size
	Per cent spent some time in this location during 9am to 6pm						
School	86.6	20.5	32.5	42.6	49.7	61.9	551
Preschool	90.6	24.3	41.8	35.2	52.2	63.1	2,285
Child care centre	86.6	20.3	59.4	23.6	43.0	59.6	1,131
Other formal	88.0	27.7	44.0	36.3	49.0	58.3	300
No formal	92.8	18.4	5.6	24.0	50.4	65.6	125
	Minutes in this location (out of possible 540 minutes)						
School	159	25	98	107	55	110	551
Preschool	223	32	104	49	57	91	2,285
Child care centre	179	28	207	28	43	70	1,131
Other formal	174	55	121	70	49	86	300
No formal	293	36	10	25	75	117	125

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.
Note: Child care use was collected at the Wave 1 interview. Children could be in multiple types of care.

The activity 'organised activities/lessons' was likely to have captured some of children's attendance at school or preschool, as parents may have felt this most closely described early education. This was explored further, in order to understand what this activity captures, and also to help understand where or how children's attendance at school or preschool might be identified from these data.

Figure 4 shows the proportion of children reported to be doing 'organised activities/lessons' and the proportion in day care by time of day. Children are further classified according to their child care arrangements from the Wave 1 interview (grouped to create a summary variable from the categories in Table 7). Overall, organised activities were often reported as occurring throughout the day, in a similar pattern to children's attendance at day care. In fact, children were often reported as being in day care and doing organised activities at the same time. Children in only school or preschool were more likely to be reported doing organised activities over the day than they were to be reported to be in day care, although, as Table 7 shows, some of these children were recorded as spending some time in day care. Amongst those reported to be using formal child care (and not preschool/school), more than 20 per cent of children were recorded as doing organised activities at any time during typical school hours. Those children in formal childcare as well as school/preschool had patterns somewhere between the formal child care only category and the school/preschool only category.

It appears that those interested in identifying children in early education or childcare on the diary day should consider using a combination of information, including the 'day care/play group' item, but making use of other data from the 'organised activities/lesson' activity and perhaps from the other 'where' data. This classification might need to be reviewed for future TUD collections, as it makes it questionable how useful the measure of 'day care' is. Users will need to be aware of the limitations of this item.

Figure 4 Participation in organised activities and attendance at day care, by child care arrangements, weekdays
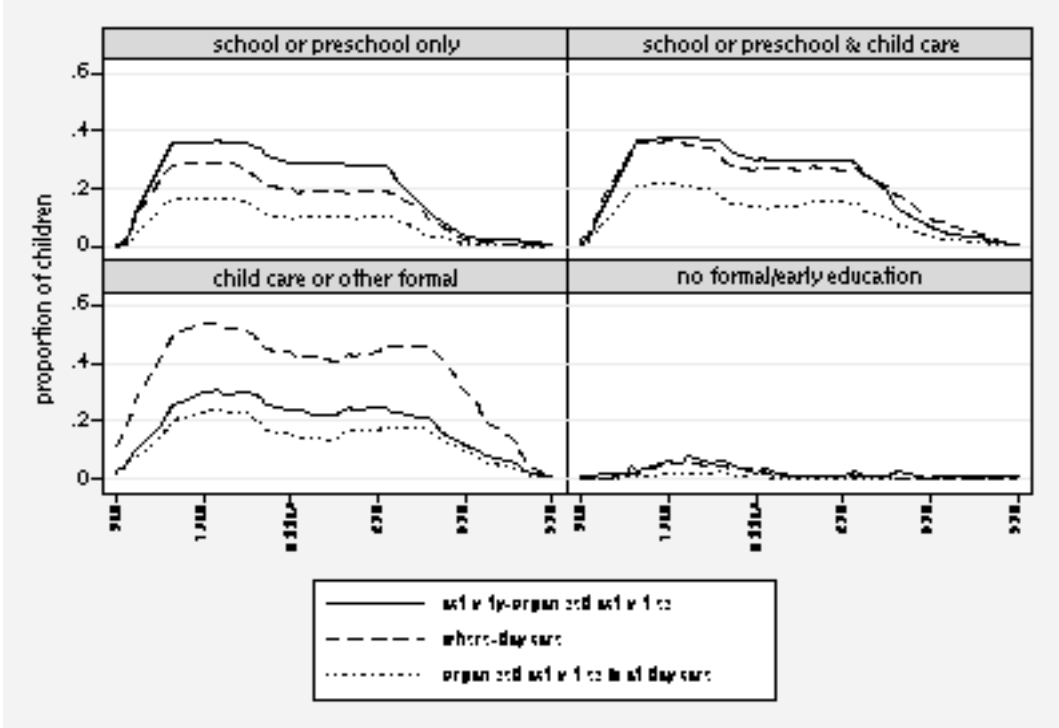
Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

What is the relationship between parental employment, non-parental care and missing data? Children who spent some time in non-parental care were more likely to have some missing data, and to have more missing data, on average, than those who did not spend time in non-parental care (Table 8). Amongst children that spent some time in non-parental care, there was more missing data when the primary carer worked full-time, but the differences between those who did and who did not use non-parental care were greater than the differences by primary carer's work hours.

For children of this age, any analysis of TUD needs to consider the effect of children being in non-parental care for some of the day.

Table 8 Parental employment, non-parental care and missing data on weekdays

Primary carer's employment	Percentage of children with some missing activity data (%)			Mean time child activity data were missing (minutes)		
	No non-parental care	Used non-parental care	Total	No non-parental care	Used non-parental care	Total
Not employed	42.8	56.8	51.2	95	163	136
Part-time	39.2	55.9	50.6	67	172	139
Full-time	45.4	60.8	57.6	161	193	187
Total	**41.6**	**57.1**	**51.9**	**90**	**172**	**144**

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

3.5 Single parents

The analyses so far have considered the times children were in non-parental care to be when parents were less likely to know what the child was doing. It is, however, quite plausible that the respondent (usually the mother) did not know about the child's activities when the child was with the other parent. In particular, this may have been the case when the parents were not living together. Table 9 very clearly shows that this was the case. Especially in single parent families, but also in couple families, the child's activity data were more likely to be missing when the child was not with the respondent, but was with the other parent. While little can be done about this, users of these data should be aware of this missing data if attempting to compare activities in couple and single parent families. Note however, that when the child was with the respondent, there was very little missing data, regardless of family type and weekend/weekday.

Table 9 Family type and missing activity data

	Couple		Single parent	
	Weekend	Weekday	Weekend	Weekday
	Missing activity data as per cent of all activity data (%)			
Child was with respondent	2.4	3.2	2.8	3.8
Child was not with respondent	*3.3*	*4.7*	*28.0*	*18.2*
Child was with other parent	7.7	14.4	15.9	21.3
Child was with others but not with other parent	2.4	3.2	2.8	3.8
Total	**4.9**	**9.5**	**10.2**	**13.9**
	Per cent of time child with respondent (%)			
Total	**48.2**	**40.5**	**48.2**	**42.2**

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.
Note: Excludes families in which the TUD respondent was neither the mother nor father.

Another consideration for analysing children applies to couple families in which the child's biological parent lives elsewhere. In the TUD, the 'who with' data included 'father/step-

father'. This did not allow for the distinction to be made between the step-father and the biological father. Care might need to be taken, in these families, when analysing children's time with father/step-father, perhaps using the 'where is' data and/or other Wave 1 data on non-resident-parent contact to help determine whether the time referred to the father or the step-father. This also applies to mother/step-mother time, if there were some children residing with the biological father, but spending time in the day with the non-resident biological mother.

3.6 Other characteristics associated with missing data

More missing data were associated with a child's participation in non-parental care, longer carer working hours and being in a single-parent family. Were other characteristics associated with a greater likelihood of having missing activity data? We can answer this question by looking at the probability of a diary having missing data according to some of the family/ child characteristics, using multivariate analyses.

Table 10 confirms that diaries from single parent families were more likely to have missing data, as were those with a primary carer working full-time or the child in day care some time. Further, the likelihood of the activity data being missing declined with the age of the primary carer. Missing activity data was also less likely if the primary carer had a complete secondary education only or a bachelor degree or higher. Diaries for girls and older children were also somewhat more likely to have missing data. Some of these effects, while significant, were very small.

Table 10 Probability of diary having some missing data, logistic regression results

	Odds ratio		
	Missing activity data	**Missing who data**	**Missing where data**
2 children	0.92	1.00	1.13
3 or more children	0.96	0.91	1.16
Single parent family	1.37***	1.37***	1.14
Incomplete secondary plus diploma/ certificate only	0.86*	0.80**	0.94
Complete secondary education only	0.71***	0.57***	0.81*
Complete secondary and diploma/ certificate	0.85*	0.70***	0.87
Bachelor degree or higher	0.60***	0.45***	0.70***
Age of primary carer	0.98**	1.00	1.01
Primary carer works part-time	1.00	0.84**	0.98
Primary carer works full-time	1.15	1.01	0.99
Age of child (months)	1.03**	1.05***	1.03***
Child is a girl	1.13*	1.03	1.04
Tuesday	1.07	1.16	1.16
Wednesday	1.04	1.16	1.17
Thursday	1.03	1.03	1.08
Friday	0.91	1.07	1.11
Saturday	0.60***	0.86	0.92
Sunday	0.61***	0.80*	0.91
Is an ordinary day	1.11	0.93	0.93
Some time in day care	1.74***	0.94	1.51*
Constant	0.45	0.05***	0.26*
Sample size	**7,348**	**7,348**	**7,348**
Adjusted R-square	**0.03**	**0.03**	**0.01**

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.
Note: Reference categories are single child family, couple family, education (of primary carer) is incomplete secondary only, primary carer is not working, child is a boy, diary day is a Monday, it is not an ordinary day child and spends no time in non-parental care on the diary day.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

4 Imputation for missing data

The existence of missing data is clearly an issue, particularly so for those children in non-parental care. For analyses by time of day, those with missing information for a time period can be excluded from calculations just for that time period, by setting the values to missing rather than zero just for that time or those times that are affected. While this means that certain types of children, particularly those in day care, are excluded from that analysis, it is perhaps reasonable to assume that their activity patterns were somewhat similar to those whose activity details were provided, as was shown in Figure 3. This may be less robust for

those activities that differ for those in day care, for example, travel/taken places was likely to have a different distribution for those in day care.

There are various techniques available to researchers when analysing missing data, and this paper does not attempt to review different methodologies. Those interested in analysing these data should consider whether there are more suitable methods for their application, (for example, Acock, 2005).

This method uses data at the person-time-period level to impute the likelihood of undertaking each activity when there was missing activity in a time-period. Multivariate analyses (logistic regression) were run on the non-missing data for each activity group to estimate the likelihood that the child did that activity at that time. This approach is demonstrated with these data, using models that included parameters to capture the time period (in hours rather than 15-minute blocks) as well as other predictors that are known to be related to child activities (age and sex of child, family structure, education level of parents) (Bianchi & Robinson, 1997; Hofferth & Sandberg, 2001; Timmer, Eccles, & O'Brien, 1985). Further, an indicator of whether the child was in 'day care/playgroup' in that time period was also included. Models were estimated separately for weekend/weekdays. The coefficients for the activity groups, excluding the time-coefficients, are shown in Table 11. Standard errors were adjusted to take into account the clustering of observations per person.

As this process was used to impute for missing data, these coefficients have not been interpreted. However, it is worth noting that there is a 'day care' effect. Children in day care were less likely to be sleeping, in leisure and travel/taken places, but more likely to be in personal care. This may be related to differences in the activity details provided for children in day care as opposed to those that are not, and should not be interpreted as having a substantive meaning beyond this.

These models were then used, for each person-time observation, to predict the likelihood that a child undertook this activity at this time. The predicted likelihood was used in place of the missing value wherever a missing value occurred. These predicted values were not converted to zero/one indicators of whether or not the child was predicted to have done this activity, as doing so resulted in considerable overstatement of activities. Retaining the predicted values as probabilities can be thought of as the predicted proportions of that 15-minute period spent on each activity.

Table 11 Probability of a child undertaking sleep, leisure, personal care, or travel/taken places at time *t*, selected logistic regression odds ratios

	Weekends				Weekdays			
	Sleep/ rest	Leisure	Personal care	Travel /taken places	Sleep/ rest	Leisure	Personal care	Travel /taken places
Age of child (months)	0.98***	1.00	1.01*	0.99	0.97***	0.99**	1.01	0.98*
Girl	1.14***	1.00	0.93**	0.94	1.05	0.98	0.94**	0.99
Single parent	1.08	1.15***	0.97	0.91	1.15**	1.15**	0.93	1.03
Complete sec. education only	1.07	0.93	0.97	1.24**	0.97	0.92*	1.07	1.11
Diploma/ certificate	1.06	0.97	0.99	1.11	0.97	0.98	1.03	1.09
Bachelor degree or higher	1.04	0.93*	1.08*	1.22**	0.94	0.92*	1.17***	1.08
In day care	0.32	0.48**	2.33**	1.24	0.59***	0.41***	2.56***	0.36***
Adjusted R-square	0.65	0.19	0.32	0.17	0.68	0.22	0.38	0.18

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.
Note: Probability for each activity is determined for each 15-minute time period over a 24-hour day (t=1 to 96). Model also included time parameters, one for each hour of the 24-hour day. Education is education of primary carer, with incomplete secondary education as reference category. Diploma/ certificate include those with and without complete secondary education.
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 12 shows the proportion of total time spent on each activity group, comparing the original data (missing data time periods excluded) with the post-imputation data. It was hoped that the data would be similar, and this was the case. There was some difference in the proportion of weekday time spent in sleep/rest (imputed data average is lower) while the weekday proportion of time spent in leisure was slightly higher for the imputed values data. The differences, however, were small equating to around half an hour difference in the imputed and original data in an average weekday 24-hours.

Table 12 Proportion of 24-hour day in each activity group, comparison of original and after-imputation data

	Original data (missing excluded)		Post-imputation data (with imputed values)	
	Weekend	Weekday	Weekend	Weekday
Sleep/rest	0.498	0.516	0.496	0.493
Personal care	0.174	0.177	0.173	0.175
Leisure	0.388	0.371	0.390	0.394
Travel/ taken places	0.088	0.080	0.089	0.082

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

If viewed by time of day over the 24-hour day and weekday/weekend using tempograms, the original and the post-imputation distributions were virtually indistinguishable (therefore not shown – the distribution for original data is shown in the next section).

Imputations were not attempted for the missing 'who with' or the missing 'where' data.

5 Activities and other data by time of day

In this section, the data were used to show children's activities over the day. The 'where' and 'who with' data are also shown. At each 15-minute period, the proportions were calculated using data for which activity, 'where' or 'who with' details were known. That is, missing data were treated by excluding data at the time-period level, rather than excluding entire diary records.

Analysed across the day, the activity patterns are in accordance with expectations (Figure 5):

- Sleep occurred primarily between around 9pm and 7am, with a small proportion of children sleeping in the early afternoon.

- Personal care was observed around meal-times, with a greater peak in the evening, no doubt associated with bathing and other personal care activities.

- Leisure encapsulated much of children's day-time activities.

- On weekdays, travel/taken places was most likely at the beginning of the day then at around 3pm. On weekends, time in this activity was most likely in the morning-late morning, and then tapered off to the evening.

Figure 5 Activity by time of day
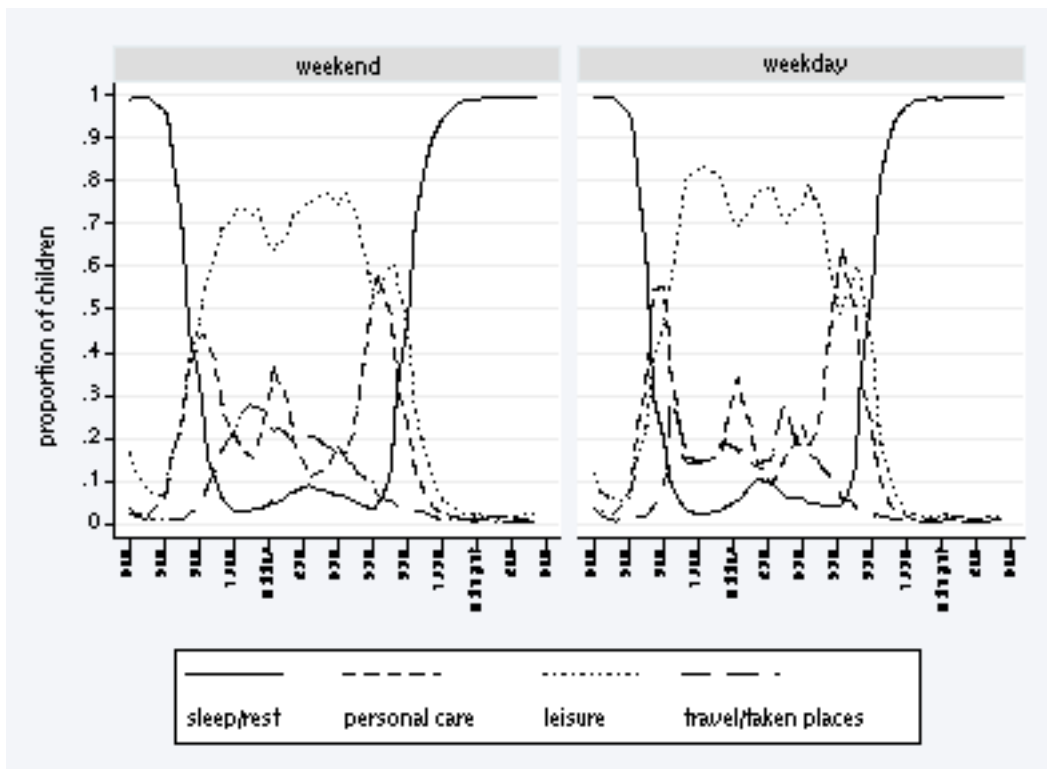
Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

The 'who with data' in Figure 6 illustrate:

- Children spent much of the day with other children – even overnight between 40 and 50 per cent of children were with other children.

- Mothers were more likely than fathers to be with the child, on weekdays as well as weekends.

- The effect of father's employment is seen clearly on weekdays, when fathers were considerably less likely to be with the child during the day.

Figure 7 shows the following patterns of where the child was:

- Most children were at home at night on weekdays, and a small percentage were away from home (usually other home indoors) on the weekend.

- Just over half the children were at home indoors during the day on the weekends, with just under 40 per cent outside during the day on the weekend. After school-time was when children were most likely to be outside on weekdays.

- Children were more likely to be in a day care centre on weekdays, especially in the morning.

Figure 6 Who child was with by time of day
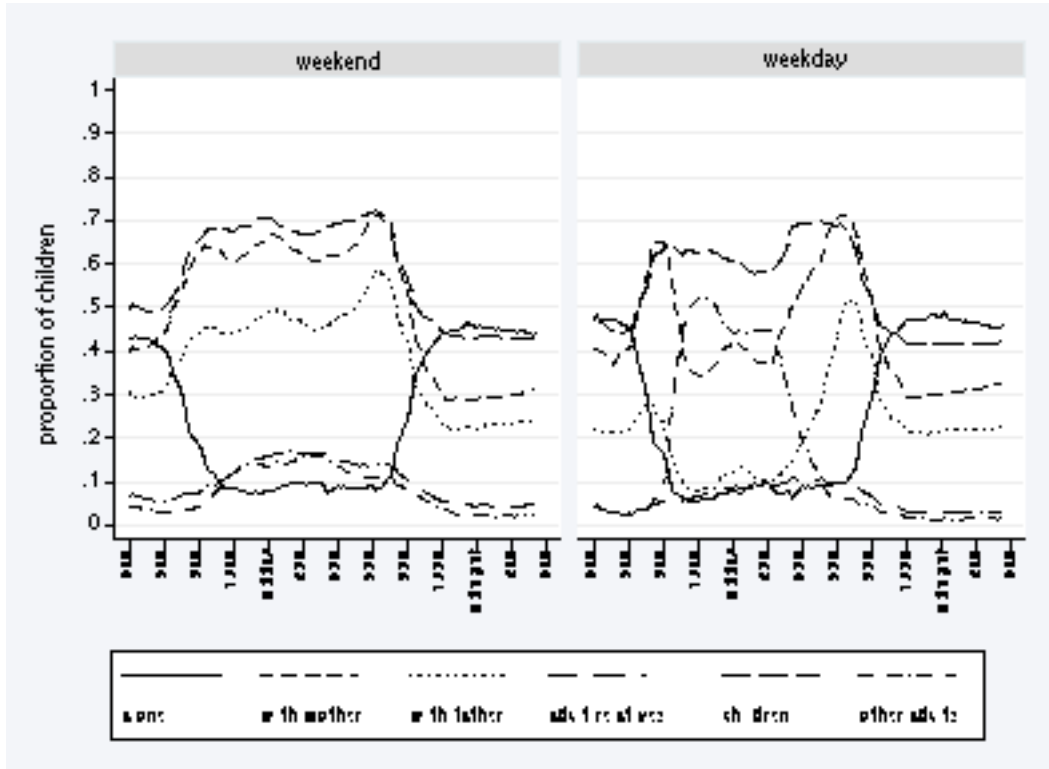
Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

Figure 7 Where child was by time of day
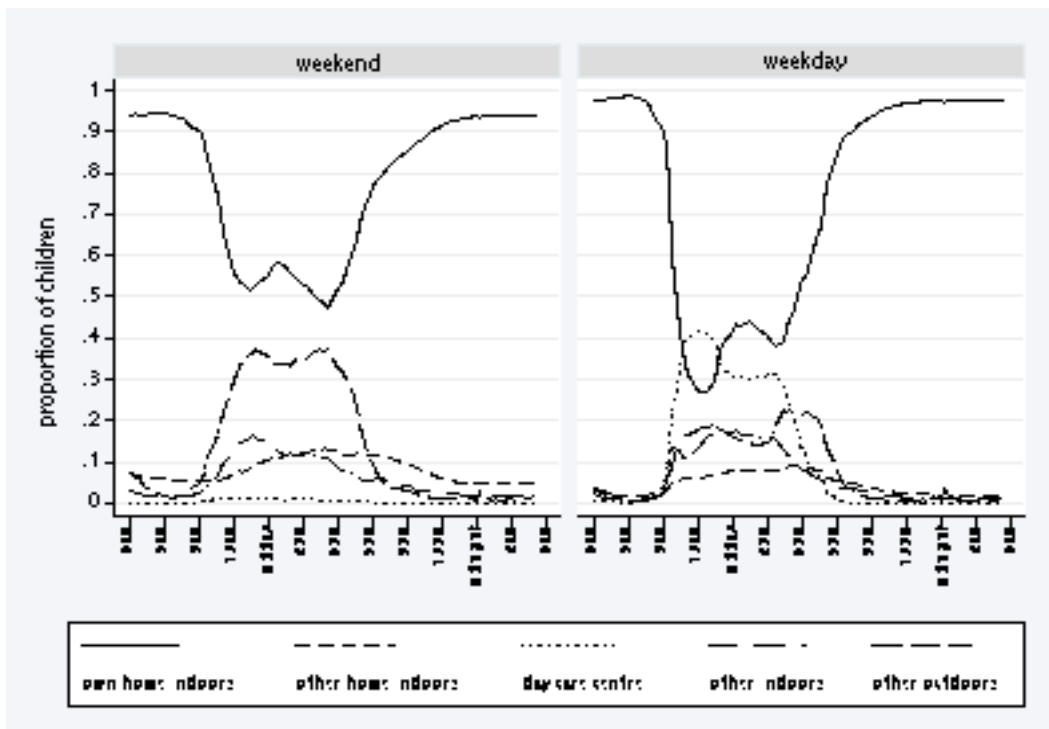

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

Data in this time-period format allow a clear picture to be built up over a child's day. Such charts are visually appealing, and can be produced to compare families or children with different characteristics. Despite these possible uses, to assess differences in distributions across different groups in a more systematic way, other techniques are required. The logistic regression parameters used to impute missing data are useful for this purpose as they show which types of children/families are more or less likely to do each activity at any point in time. However, a more typical way of comparing groups is to compare the mean times spent over a day in a type of activity. This approach is discussed in the next section.

6 Mean time in activities

The missing data cause particular difficulties in attempting to calculate mean times in activities from the TUD. Ideally, we would like to be able to sum up the times children spend in different activities and have the whole day accounted for. With missing data, the only way this can occur is to treat 'missing' as a category. This means, however, that as the amount of missing data increases, the more some other category is lessened. For some (especially those in day care) this can have a significant impact on the time spent in leisure activities, and possibly other activities that occur throughout the day. Clearly, this has implications for calculating mean values, especially if the missing data are more prevalent in some family types than others. Any differences in the means may be just an artefact of more or less missing data.

Analysis of mean times in activities is done at the person-level, rather than the person-time level, so it is not feasible to drop partial records, as was possible in the previous section in dealing with missing data. Entire person-records need to be removed, or missing data replaced with imputed data, as was demonstrated previously. Dropping entire records with any missing data throws away a large amount of data, and the retained sample is biased towards children who are not in non-parental care. There may be other biases, as mentioned earlier. In other analyses of these TUD data, means have been analysed by excluding those with more than 90 minutes of missing data. Using this approach, there are cases in the retained data with up to 90 minutes of missing data, which reduces by a small amount the durations of some activity groups. To investigate how this method compares to others, analyses are included which adopt this approach. These data are referred to as original-less-missing.

Another option is to use all data and to incorporate the imputed data for those time periods in which the activity data are missing. These data are referred to as the post-imputation data.

Note that this is not entirely appropriate, as some diaries are included for which a significant proportion (more than 50 per cent) of the data are imputed. In future analyses of these data, those with too much missing or imputed data should be excluded.

Table 13 shows the mean time in the broad activity groups using the original data (that is the data without special treatment of records with missing data), original-less-missing data, and the post-imputation data. In the original data, the times for which data are missing were not allocated to any of the activity groups. The original data produced the smallest estimates of mean times in each category, but affecting the larger figures more – sleep and leisure times. The exclusion of records with more than 90 minutes of missing activity data made a considerable difference to all estimates.

Table 13 Broad activity distributions, original, original-less-missing and post-imputation data

Variable	Original data (Sample size=7,449)		Original-less- missing (Sample size =5,409)		Post-imputation data (Sample size =7352)	
	Mean	*Std. Dev.*	Mean	*Std. Dev.*	Mean	*Std. Dev.*
Weekend						
Sleep/rest	691	*151*	729	*90*	714	*90*
Personal care	237	*112*	242	*108*	250	*110*
Leisure	529	*178*	563	*157*	561	*152*
Travel/ taken places	121	*122*	127	*124*	128	*119*
Weekday						
Sleep/rest	683	*144*	723	*81*	709	*79*
Personal care	231	*110*	248	*111*	252	*103*
Leisure	480	*198*	577	*141*	567	*130*
Travel/ taken places	104	*102*	115	*110*	118	*100*

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.
Note: In original data, missing activity periods are not assigned to a category, but all records are included. In original-less-missing, those with more than 90 minutes missing data are excluded. For those with 90 minutes or less missing data, the missing data are not assigned to a category. In the post-imputation data, all records are included, except that these data exclude one diary per person when two diaries were completed for the same day of the week. The missing data are replaced with the imputed data.

The estimates using the post-imputation data were more similar to the original-less-missing data than they were to those using the original data. Compared to the original data, the post-imputation data produced higher estimates in all categories. This was as expected, given that in the original data missing time periods were simply not attributed to an activity, reducing the total time in those activities that were most likely to be affected by the missing data. The largest difference between original and imputed data was for weekday leisure time, which was the activity that took up the most time during the day, when data were most likely to be

missing – through imputation much of this missing weekday time had a high probability of being leisure time, and therefore boosted this estimate.

Table 14 shows these same data as percentages of the 24-hour day. For weekends, the distribution across activity groups was fairly similar, regardless of the data source, although the percentages in the listed activity groups were lower, of course because some of the time was attributed to an 'unknown' category. The lack of very large differences between data sources was because the missing data, for weekends, was more fairly distributed across the day (Figure 1), relative to that of the weekday missing data. It was not surprising to find that the percentages were quite different for the original weekday data, with a smaller percentage in the leisure category, as was clear in the mean minutes shown above.

Table 14 Broad weekday activity distributions as percentages, original, original-less-missing and post-imputation data

Weekday	Weekend			Weekday		
	Original	Original-less-missing	Post-imputation	Original	Original-less-missing	Post-imputation
			Percentage of day			
Sleep/rest	48.0	50.6	49.6	47.4	50.2	49.2
Personal care	16.5	16.8	17.4	16.0	17.2	17.5
Leisure	36.7	39.1	39.0	33.3	40.1	39.4
Travel/ taken places	8.4	8.8	8.9	7.2	8.0	8.2

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.
Note: Unweighted data, sample size=7,352

As much of the missing data was related to the child's presence in non-parental care during the week, we expected larger differences between original and imputed data for those who had used non-parental care. Table 15 shows that, again for 'leisure', the original data for those who used non-parental care produced a considerably smaller estimate. For the other activity groups, the differences between original and imputed data were only slightly greater for those who used non-parental care.

The similarity of results using the original-less-missing and the post-imputation data suggest that either approach could be used to analyse the TUD data to address the missing-data issue.

Using the post-imputation data, Table 16 shows the mean time in each of the activities as listed in the TUD, as well as the overall categories. Also, this table shows the proportion of children reported to have spent some time in each activity (or activity group) on the diary day. This is another important way of summarising these data.

Table 15 Broad weekday activity distributions, original, original-less-missing and post-imputation data

	Used non-parental care			Did not use non-parental care		
Weekday	Original	Original-less-missing	Post-imputation	Original	Original-less-missing	Post-imputation
	Mean minutes					
Sleep/rest	662	716	704	682	734	721
Personal care	219	238	244	251	262	266
Leisure	469	601	580	502	542	540
Travel/ taken places	92	100	108	126	137	136
	Percentage of day					
Sleep/rest	46.0	49.7	48.9	47.4	51.0	50.1
Personal care	15.2	16.5	16.9	17.4	18.2	18.5
Leisure	32.6	41.7	40.3	34.9	37.6	37.5
Travel/ taken places	6.4	6.9	7.5	8.8	9.5	9.4

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.
Note: Unweighted data, sample size=7,352.

These data show the following, for example:

- At the broadest level, there was considerable similarity across weekdays and weekends for time spent in sleep, personal care and leisure, and even within these categories weekends and weekdays were similar for some activities. For example, almost 4 hours was spent on personal care activities on either weekday or weekend, and the largest activity within this time was eating, drinking or being fed.

- The total amount of time in leisure was the same for weekdays and weekends. Within this broad category, there were considerable weekday/weekend difference in how this time was distributed across specific activities. Of the different types of play identified, watching television or movies took up the most time, although much activity was recorded as 'other play or other activities' (likely to include creative play, playing with toys). Children spent less time doing activities defined as exercise on weekdays, and on weekdays and weekends spent less time in 'exercise' than watching television. Children spent just under two hours a day in activities defined as 'achievement-oriented', spending more time on these activities during the week.

- Participation in organised lessons/activities was more likely on a weekday, but 'visiting people/special event/party' was more likely on a weekend.

- Most of the 'travel/taken places' involved 'taken places with adults' (note that this is not necessarily travel, ie, it includes being at the supermarket) and travel in car. The former was more likely on weekends while the latter was slightly more likely on a weekday.

- The 'per cent doing' data showed that around 60 per cent of children spent some time in the day being held or cuddled, while only around one in four spent some time crying or upset and just a little more than this spent some time being reprimanded or disciplined.

- On a weekday, most children did an achievement-oriented activity. Exercise was less common, although still done by more than half the children, and more likely on the weekend (72 per cent on the weekend and 64 per cent on a weekday did some exercise). About one in four children walked for fun or travel and about one in four (more on the weekend) rode a bike/tricycle. Around nine in ten children watched television: one in four used a computer and about the same number listened to the radio/ CDs.

Table 16 Children's activities over a 24-hour day

	Weekend			Weekday		
	Mean minutes	**Per cent of total time**	**Per cent doing activity**	**Mean minutes**	**Per cent of total time**	**Per cent doing activity**
Sleep	**714**	**49.6**	**100**	**709**	**49.3**	**100**
Sleeping/ napping	674	46.8	100	669	46.5	100
Awake in bed	39	2.7	62	41	2.9	65
Nothing/ bored/ restless	9	0.6	18	8	0.6	18
Personal care	**250**	**17.3**	**100**	**252**	**17.5**	**100**
Eating/ drinking/ being fed	140	9.7	99	140	9.7	100
Bathe/ dress/ hair/ health	62	4.3	94	68	4.7	97
Held/cuddled	45	3.1	59	46	3.2	63
Crying/ upset	11	0.8	27	11	0.8	27
Being reprimanded/ corrected	13	0.9	30	14	1.0	32
Destroy things/ create mess	8	0.6	16	9	0.6	17
Leisure	**561**	**39.0**	**100**	**567**	**39.4**	**100**
Achievement	*118*	*8.2*	*83*	*132*	*9.2*	*89*
Read a story, talked/sung to	63	4.4	56	75	5.2	61
Colour/ look at book/ educational game	45	3.1	65	52	3.6	77
Taught to do chores or read	20	1.4	35	22	1.5	38
Exercise	*104*	*7.2*	*74*	*77*	*5.4*	*67*
Walk for fun or travel	16	1.1	22	16	1.1	25
Ride bike	25	1.7	33	18	1.3	27
Other exercise	70	4.9	56	49	3.4	49
Other 'leisure'						
Watching television, movie	134	9.3	90	125	8.7	90
Listening to radio, music	22	1.5	27	21	1.5	28
Use computer	21	1.4	29	17	1.2	25
Visiting people/ special event/ party	103	7.1	50	37	2.6	28
Other play/ other activities	159	11.0	79	115	8.0	72
Organised lessons/ activities	15	1.0	17	144	10.0	56
Travel/taken places	**128**	**8.9**	**82**	**118**	**8.2**	**88**
Taken places with adult	67	4.7	52	51	3.6	46
Travel in a car	65	4.5	71	70	4.9	80
Travel on public transport	6	0.4	10	7	0.5	13
Taken in a pusher/bike seat	5	0.3	10	5	0.3	11
Total	**1,440**	**100.0**		**1,440**	**100.0**	

Source: LSAC TUD 2004 Version 1, 4-5 year cohort (post-imputation).
Note: The mean values include imputed data for missing time periods.

7 Correlates of mean time in activities

With these data, multivariate analyses can be used to find associations between child and family characteristics and time spent in activities. The choice of which type of multivariate technique is, however, not straightforward. The analysis of time diary data is often done using Tobit estimation. The Tobit model takes into account those cases reporting no time spent in an activity, by adjusting coefficients to analyse the length of time as well as the likelihood of there being non-zero data. There is, however, some disagreement about the appropriateness of this method for analyses of time use data. Some time diary analysts recommend Ordinary Least Squares (OLS) over Tobit as a more appropriate method (Brown & Dunn, 2006; Gershuny & Egerton, 2006; Stewart, 2006). The choice of which of these estimations to use only makes a difference for those activities that are not universal, because they occur less frequently over a time period, or because a proportion of the population never or rarely undertakes these activities. The purpose of this paper is not to argue for the benefits of one method over the other, nor to comprehensively address the differences between the two methods. To analyse the duration data, OLS have been used initially, and then Tobit has been included in the comparison of different missing data treatment options.

For this analysis OLS estimation was first conducted using the post-imputation data (Table 17). These multivariate analyses controlled child characteristics (sex and age of child) as well as parental characteristics (hours worked, education level, country of birth/English language proficiency). Some of the results evident in these data are:

- Children in single parent families spent around 20 minutes longer in personal care compared to children in couple parent families. When the primary carer worked, children spent less time in personal care.

- Girls spent longer sleeping (just on weekends) and less time in leisure.

- Older children spent less time sleeping, less time in personal care (on weekdays) and more time in leisure.

- Children with non-Australian-born primary carers spent more time in personal care, especially those with poor English language proficiency. They spent less time in leisure but on weekends were more likely to travel and/or be taken places.

Table 17 OLS coefficients, mean daily time on selected activities (minutes)

	Sleep	Personal care	Leisure	Travel/ taken places
	Weekends			
Single parent	5.33	21.19***	-1.67	-9.16
2 children	-2.58	-3.97	1.36	-6.76
3+children	-6.52	-9.27	7.45	-14.66*
Primary carer part-time	-8.01*	-16.04***	0.02	7.98
Primary carer full-time	-3.31	-11.91*	-12.87	15.75*
Age of primary carer (centred)	1.15***	0.62	-1.26*	-1.40***
Incomplete secondary and diploma/certificate	6.62	-11.46	-4.92	20.12**
Complete secondary education only	6.00	0.67	-4.03	1.94
Complete secondary and diploma/certificate	6.85	-6.02	2.80	8.51
Bachelor degree or higher	7.83	-8.26	17.12*	9.93
Age of child (months-centred)	-2.19***	-0.65	2.24*	-0.79
Study child is girl	13.66***	-0.13	-15.24**	-5.62
Not Australian-born, good English	2.47	5.71	-17.31**	13.05**
Not Australian-born, poor English	8.89	39.60**	-27.72	0.76
Constant	708.55***	265.08***	566.09***	124.29***
Sample size	**3,582**	**3,582**	**3,582**	**3,582**
R-squared	**0.02**	**0.02**	**0.01**	**0.02**
	Weekdays			
Single parent	11.03**	19.47***	-15.20*	5.87
2 children	7.40	-4.28	-26.08***	14.08*
3+children	5.56	-1.00	-30.02***	13.64*
Primary carer part-time	-4.71	-13.61***	-6.55	1.21
Primary carer full-time	-13.08**	-12.33*	-1.19	-6.46
Age of primary carer (centred)	0.92***	0.42	-1.50***	-1.44***
Incomplete secondary and diploma/certificate	3.52	-0.74	-3.46	2.62
Complete secondary education only	-3.15	-13.20*	12.43	9.15
Complete secondary and diploma/certificate	-3.63	-7.45	13.95	8.50
Bachelor degree or higher	-1.30	-11.24*	26.70***	2.32
Age of child (months-centred)	-2.21***	-1.96**	0.93	-1.73**
Study child is girl	4.90	-3.68	-11.51**	-1.63
Not Australian-born, good English	-0.47	2.65	1.25	-6.43
Not Australian-born, poor English	11.19	29.43*	-16.66	-4.59
Constant	705.31***	267.04***	589.36***	103.73***
Sample size	**3,643**	**3,643**	**3,643**	**3,643**
R-squared	**0.02**	**0.02**	**0.02**	**0.01**

Source: LSAC TUD 2004 Version 1, 4-5 year cohort (post-imputation).

Note: Education is education of primary carer, with incomplete secondary education as reference category. Country of birth and English proficiency refer to primary carer, with reference category Australian-born. Other reference categories are couple parent, single-child family, primary carer not employed, study child is a boy. Age of primary carer and age of child are centred (35 years and 57 months respectively).

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

- The education level of the primary carer had the strongest relationships with time spent in personal care and leisure. For example, on weekdays, where the primary carer had a Bachelor degree or higher, children spent less time in personal care and more time in leisure compared to children with a primary carer who had an incomplete secondary education.

For all activity groups and the more detailed categories, alternate model specifications were tested to compare the results across different ways of treating missing data. Specifically, OLS using the original data, the original-less-missing data and the post-imputation data were compared. Also, different methodologies were tested. In addition to the three OLS estimations, a Tobit estimation was also tested using the post-imputation data. Selective results are shown in Table 18 ('leisure') and in Table 19 (a subset of this leisure category, 'exercise').

The original data OLS results were often quite different to the other models, this is not surprising, given that a number of the family/child variables within the model, also explained the presence of missing data (see Table 10) – the results therefore explained differences in the variable of interest as well as the extent of missing data. The removal of records with a large amount of missing data, as in the original-less-missing sample, resulted in estimates that more closely resembled those in the post-imputation model. While the coefficients were often of different size, they were almost all in the same direction and the levels of significance were usually similar.

The OLS and Tobit results on the grouped activities were practically the same, in terms of size of coefficients, direction of effects, and levels of significance. This is because all or almost all children did these activities. When similar models were estimated for activities which were done by a smaller percentage of children, there were more differences between the Tobit and OLS estimates.

Table 18 Total weekday daily time spent on leisure, comparison of models

	Original OLS	Original OLS, missing excluded	Post-imputation OLS	Post-imputation Tobit
Single parent	-38.81***	-10.59	-15.20*	-15.20*
2 children	-10.75	-39.88***	-26.08***	-26.09***
3+children	-10.11	-39.56***	-30.02***	-30.02***
Primary carer part-time	-27.29***	-9.39	-6.55	-6.57
Primary carer full-time	-46.98***	8.52	-1.19	-1.19
Age of primary carer	-2.42***	-1.59**	-1.50***	-1.50***
Incomplete secondary and diploma/certificate	1.98	-10.43	-3.46	-3.50
Complete secondary education only	31.70**	7.14	12.43	12.44
Complete secondary and diploma/ certificate	27.93**	19.26	13.95	13.96
Bachelor degree or higher	55.47***	15.94	26.70***	26.69***
Age of child (months)	-0.76	0.93	0.93	0.94
Girl	-17.31**	-10.84	-11.51**	-11.53**
Not Australian-born, good English	-15.24	2.24	1.25	1.26
Not Australian-born, poor English	-53.27*	-40.15	-16.66	-16.65
Constant	498.17***	615.19***	589.36***	589.37***
Sigma				128.83***
Sample size	**3,643**	**2,307**	**3,643**	**3,643**
R-squared	**0.03**	**0.02**	**0.02**	

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.
Note: Education is education of primary carer, with incomplete secondary education as reference category. Country of birth and English proficiency refer to primary carer, with reference category Australian-born. Other reference categories are couple parent, single-child family, primary carer not employed, study child is a boy. Age of primary carer and age of child are centred (35 years and 57 months respectively).

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 19 Total weekday daily time spent on exercise, comparison of models

	Original OLS	Original OLS, missing excluded	Post-imputation OLS	Post-imputation Tobit
Single parent	4.85	13.53*	9.72*	11.77*
2 children	8.48	5.73	7.69	7.14
3+children	12.93**	10.03	11.76*	12.91*
Primary carer part-time	-3.78	1.94	-1.97	-2.39
Primary carer full-time	-8.87*	-1.46	-4.08	-5.01
Age of primary carer	-0.42	-0.31	-0.30	-0.29
Incomplete secondary and diploma/certificate	-2.47	-5.30	-3.75	-4.15
Complete secondary education only	2.60	-1.80	-0.06	-1.10
Complete secondary and diploma/certificate	1.42	-0.58	-0.69	-1.11
Bachelor degree or higher	2.38	-5.37	-1.67	-3.01
Age of child (months)	-0.86	-0.45	-0.63	-0.59
Girl	-8.93**	-10.16**	-9.51***	-9.19**
Not Australian-born, good English	-2.88	0.57	-1.07	-0.32
Not Australian-born, poor English	-8.03	-32.75	-4.09	-2.11
Constant	64.58***	79.21***	75.13***	68.56***
Sigma				128.83***
Sample size	**3,643**	**2,307**	**3,643**	**3,643**
R-squared	**0.01**	**0.01**	**0.01**	

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

Note: Education is education of primary carer, with incomplete secondary education as reference category. Country of birth and English proficiency refer to primary carer, with reference category Australian-born. Other reference categories are couple parent, single-child family, primary carer not employed, study child is a boy. Age of primary carer and age of child are centred (35 years and 57 months respectively).

*** p <0.001, ** p <0.01, * p<0.05

8 Discussion

The main purpose of this paper was to analyse the LSAC TUD missing data. Clearly, missing activity data was related to times when children were in non-parental care. This did not just reflect parental employment, since many children attended some form of early education or care for some time during the day, irrespective of their parents' employment arrangements.

These analyses showed that there were significant amounts of missing data. These missing data can affect the results of analyses, especially of weekday activities and so it is important that researchers adopt appropriate techniques when analysing these data. In this paper, the type of analysis determined the approach used. For analysis by time of day, the missing data could be excluded at the person-time level, so that data could still be retained at times when

activity data were non-missing. These non-missing data, across all children, were used to predict the likelihood of undertaking each activity when there were missing data for a time period. This method is one way the TUD data can be explored.

For analyses of time spent in activities, the treatment of missing data was particularly important, as durations in activities at the person-level were smaller when there was more missing data for that person. Some activities were affected more than others. In these data, the leisure estimates were smaller when records with missing data were included (for weekdays in particular), since most missing data were during the daytime when leisure activities were the most likely to occur. This analysis showed that two approaches produced fairly similar results: excluding those with more than 90 minutes of missing data and including all diaries, but replacing missing data with imputed data. Either of these methods could be used to address the problem of missing data.

There is no perfect way of taking into account the missing data associated with children's time spent in non-parental care. This analysis attempted to correct for this missing data by imputing activity data using the non-missing data, but there was still a 'day care' effect in the models of time spent in activities. Without much more detailed analysis of the recording of activities in day care time versus non-day-care time, it is recommended that these data not be used to make comparisons between children in day care and children not in day care.

In addition to addressing the missing data issue, this paper highlights some of the ways these data can be used to analyse children's activity patterns and time use. It also highlights some of the data issues that should be considered when analysing these data. Note that the multivariate analyses in this paper were meant to be illustrative of what is possible with the TUD. To simplify the analytical process, these analyses were conducted using unweighted data, without adjustments for sample selection effects. Future work should allow for these adjustments.

As a final note, it is important to note that there were biases in the TUD data, with lower educated families, single parent families, families with full-time employed primary carers being under-represented in the TUD relative to the LSAC sample.

This paper does not explore all data quality issues for the TUD, and only applies to the 4-5 year cohort. This analysis has not yet been repeated for the infant cohort, although it is expected that missing data will be less of a problem for the infants, as they are less likely to

be in non-parental care. Those wishing to analyse these data should also refer to the User Guide (LSAC Project Operations Team, 2006) for more information.

9 Recommendations

I offer the following suggestions for analysts:

- Carefully consider the treatment of missing data. Consider removing those with too much missing data, or imputing for those missing time periods (or a combination of these). 'Too much' is, of course, a measure to be decided upon by individual researchers. A limit of 90 minutes has been used with these data previously (Baxter, Gray, Alexander, Strazdins, & Bittman, 2007).

- Unless planning to undertake detailed and careful analysis of the activity and 'where' coding, do not use these data to compare activities of children in day care with those not in day care. The inadequate coding of where the child is when not with parents, combined with difficulties in dealing with missing data make such a comparison problematic.

- Consider the use of tempograms as a means of exploring and presenting these data.

- Take note of the bias in the TUD data, compared to the original sample.

- When day of week is important to this analysis, consider excluding records for which the day of the week information are less certain, specifically those for which the date of completion is much later than the requested date of interview. When compiling measures across the week, use diary weights (not used in this analysis, since actual day of the week was not included in the analysis) to take into account the day of week distributions.

- Consider including a variable to capture month or season of collection if activities are likely to vary across the time of year.

- If concerned about where children are during the day on weekdays, consider using 'organised activities/lessons' and possibly the 'who with' data to supplement the 'where' data. This is important as the 'where' data do not appear to have adequately captured preschool/school attendance.

- Be careful when analysing single versus couple families, given the greater extent of missing data in single parent families, and the greater likelihood of missing data when the child is with the other parent. Keep in mind that time with father could be time with non-resident father or time with resident step-father, and similarly time with mother/step-mother could be confused. Other information on where the child is may help in these analyses.

References

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family, 67*(4), 1012-1028.

Baxter, J., Gray, M., Alexander, M., Strazdins, L. & Bittman, M. (2007). *Mothers and fathers with young children: Paid employment, caring and wellbeing* (Social Policy Research Paper No. 30). Canberra: Department of Families, Community Services and Indigenous Affairs.

Bianchi, S. M. & Robinson, J. (1997). What did you do today? Children's use of time, family composition, and the acquisition of social capital. *Journal of Marriage and Family, 59*(2), 332.

Brown, J. & Dunn, P. (2006, 16-18 August). *Too little time? A fresh look at right skewed data and how Poisson-gamma distributions may provide a viable solution.* Paper presented at the International Association of Time Use Research (IATUR) - XXVIII Annual Conference, Copenhagen.

Budig, M. & Folbre, N. (2004). Activity, proximity, or responsibility? Measuring parental childcare time. In N. Folbre & M. Bittman (Eds.), *Family time: the social organization of care* (pp. 51-68). London: Routledge.

Craig, L. (2006). Does father care mean fathers share?: A comparison of how mothers and fathers in intact families spend time with children. *Gender and Society, 20*(2), 259-281.

Gershuny, J. & Egerton, M. (2006, 16-18 August). *Evidence on participation and participants' time from day- and week-long diaries: implications for modelling time use.* Paper presented at the International Association of Time Use Research (IATUR) - XXVIII Annual Conference, Copenhagen.

Hofferth, S. L. & Sandberg, J. F. (2001). How American children spend their time. *Journal of Marriage and Family, 63*(2), 295.

LSAC Project Operations Team. (2006). *The Longitudinal Study of Australian Children Data Users Guide Version 1.7* Melbourne: Australian Institute of Family Studies.

Sanson, A., Misson, S., Wake, M., Zubrick, S., Silburn, S., Rothman, S. & Dickenson, J. (2005). *Summarising children's wellbeing: The LSAC Outcome Index* (Technical paper No. 2). Melbourne: Australian Institute of Family Studies.

Soloff, C., Lawrence, D. & Johnstone, R. (2005). *Sample Design* (LSAC Technical Paper No. 1). Melbourne: Australian Institute of Family Studies.

Soloff, C., Lawrence, D., Misson, S. & Johnstone, R. (2006). *Wave 1 weighting and non-response* (LSAC Technical Paper No. 3). Melbourne: Australian Institute of Family Studies.

Stewart, J. (2006, 16-18 August). *Tobit or not tobit?* Paper presented at the International Association of Time Use Research (IATUR) - XXVIII Annual Conference, Copenhagen.

Timmer, S., Eccles, J. & O'Brien, K. (1985). How children use time. In F. T. Juster & F. P. Stafford (Eds.), *Time, goods, and well-being* (pp. 353-382). Ann Arbor, Michigan: Survey Research Center, Institute for Social Research University of Michigan.

Yeung, W. J., Sandberg, J. F., Davis-Kean, P. E. & Hofferth, S. L. (2001). Children's time with fathers in intact families. *Journal of Marriage and Family, 63*(1), 136.

Appendix 1 Day and date of interview data problems

The TUD included a 'please complete diary on [date]' line, and a day of the week indicator, which needed to be completed by the interviewer. I refer to the date as requested-date and the day as requested-day. For 13 per cent of the records, the requested-day did not align with the requested-date, that is, either the requested-day or the requested-date were incorrectly entered by the interviewer, and for such cases we do not know if the respondent paid more attention to the requested-date or the requested-day when filling in the diary. In coding these data to weekday/weekend, seven per cent of records were in different categories depending on whether the date or the day of the week was used.

Some of the diary data were used to assist in making a decision about whether to use the requested-day or requested-date to derive weekday/weekend. The data used were the information on the child's attendance at daycare centres/playgroup (which for now I refer to as "daycare") as it this was more likely to occur on weekdays. Some 26 per cent of children spent some time in daycare on the diary day. If weekday/weekend is derived from requested-day, 46 per cent of children were in day care on a weekday. Of those derived as weekday using these data but weekend using requested-date, the proportion is still quite high (43 per cent). As it does not seem realistic that such a high proportion would be using day care on the weekend, it is more likely that the diary day is actually a weekday. That is, it appears that the requested-day may have been followed more-so than the requested-date.

Respondents were also asked for the date they completed the form (called completed-date) so it is also possible to derive day of the week from this item. For just under two-thirds of respondents the completed-date was the same as requested-date and for another 29 per cent the completed-date was one of the two days following the requested-date. It is not clear whether this date refers to the date *for which* the diary was completed or the date *on which* the diary was completed.[5] There is some discrepancy between this information and other information collected on when the form was completed (more than 3 times a day, 2 to 3 times during the day, once-when the child went to bed, once-the next morning and, at a later time). Some said they filled in the diary on the diary day, but then the date of completion was one day after the requested diary date. Does that mean the diary information refers to one day later than requested, or that the respondent finished filling out the form the following day? Or does

---

[5]   This question wording was improved in Wave 2.

it mean that the day and date information were not aligned and the respondent chose to fill out the diary for the day specified rather than the date? Clearly, some respondents would have had some difficulty identifying which day they were supposed to complete the diary when the diary date and the given weekday did not agree. Given the uncertainty about what this date of completion referred to, it was decided it should not be used to derive weekday/weekend.

In this analysis, I use the requested-day to derive the weekend/weekday distinction.

Appendix 2 Missing data and data cleaning

A time period was said to contain missing activity data in one of two situations: if the child's activity was coded as 'not sure what child was doing' (where no other activity details were provided for that time period), and if no activity details were provided for that time period. In the same way, missing 'where' data and missing 'who with' data occurred when no 'where' or 'who with' data were entered for a time period. These missing data could be aggregated across the diary day to calculate the mean number of minutes of missing data, as shown in Table 20.

Table 20 Missing data distribution prior to data cleaning

		Missing activity data	Missing 'where' data	Missing 'who with'
Minutes of missing data in a day	Mean	161	232	276
	Std. Dev.	243	355	436
As per cent of total time (1,440 mins)	*Per cent*	11.2	16.1	19.2
Diaries with data all missing data	Number	20	135	418
Diaries with no missing data	Number	2,574	2,360	3,465
As per cent of all diaries (7,449 diaries)	*Per cent*	34.6	31.7	46.5

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.

Table 20 shows that 11.2 per cent of activity data were missing, meaning 88.8 per cent of time periods had complete activity data. The 'where' and 'who with' data were more likely to be missing, with 135 diaries with no 'where' data and 418 cases with no 'who with'. There were only 20 cases with no activity data.

To reduce the amount of missing activity data and to correct some other data problems the following edits were applied:

- Activity data were often missing for the night-time, although it was recorded that the child was at home. For the time periods before 7.30am and after 9pm, if activity data were missing and the child was recorded as being at home, it was assumed that the child was sleeping. These times were selected because around half the children were awake by 7.15am on weekdays (7.30am on weekends) and most were asleep by 9pm.

- Television viewing was often recorded for the 4am to morning time periods. While it is possible that the television was on at this time, it is unlikely that the child was actually watching television then, especially if they were also recorded as being asleep. This was probably related to the false positives arising from scanning. When children were recorded as being sleeping as well as watching television, the television data were reset to zero.

- Where activity data (or 'who' or 'where' data) existed for one time period but not the next, the activity for the non-missing time period was copied to the missing time period. If data were missing for consecutive time periods, only the first of these was replaced using this method. It was largely used to correct for single periods of missing activity data. The LSAC Project Operations team used a similar, although more sophisticated, approach.

The following edits were applied to the 'who with' data:

- When the 'who with' data were missing, but the child was in day care centre/playgroup, the 'who with' data were changed to say that the child was with other adults and with other children.

- If the 'who with' data were missing but the child was in 'other person's home' (indoors), the child was recorded as being with other adults. Depending on whose home this was, this may undercount the presence of another parent, other relatives or children.

- If the child was at home, but the 'who with' data were missing, it was assumed the child was alone. This largely affected the night-time coding.

The following edits were applied to the 'where' data:

- Data were edited when children were coded, in the one time period, to both in their own home (indoors) and at 'day care centre/playgroup'. In this situation the 'day care centre/playgroup' item was reset to zero. This corrected some obviously incorrect data where children were appearing as at day care in the middle of the night. (This problem was probably related to scanning errors.)

Table 21 (also repeated in Table 5) shows the distribution of missing data, for activity, 'where' and 'who with' across the 24-hour day after these edits. The extent of missing data is reduced for the activity and 'who with' data, compared to prior to these edits. The number of diaries with no missing activity data increased from 2,574 to 4,041 as a result of these edits, and the percentage of missing time fell from 11.2 per cent to 7.8 per cent.

Table 21 Missing data distribution after data cleaning, all diaries

		Missing activity data	Missing 'where' data	Missing 'who with'
Minutes of missing data in a day	Mean	113	232	119
	Std. Dev.	215	355	292
As per cent of total time (1,440 mins)	*Per cent*	*7.8*	*16.1*	*8.3*
Diaries with data all missing data	Number	19	135	97
Diaries with no missing data	Number	4,041	2,360	5,257
As per cent of all diaries (7,449 diaries)	*Per cent*	*54.2*	*1.7*	*70.6*

Source: LSAC TUD 2004 Version 1, 4-5 year cohort.