

## *Growing Up in Australia: The Longitudinal Study of Australian Children*

# **Imputations to solve missing data problems in Wave 2.5**

A number of the variables in the Wave 2.5 data files have higher levels of missingness than is usual for LSAC self-complete questionnaires. This attachment details the imputations made, and also those considered and rejected, in order to limit the amount of missing data. Using answers to other questions could impute some of these ‘missings’, and this was done wherever possible.

Examination of the questionnaires revealed the following main reasons for the high levels of missing data (note that many of these are not exclusive to Wave 2.5, but appear to be exacerbated by other problems with the Wave 2.5 questionnaire):

1. Formatting issues. On pages where questions were in two columns at the top of the page, but then in only one column at the bottom, some respondents missed the second column at the top of the page. This affected the following questions:

Cohort	Question no.	Description
B	2	TV/ computer in other rooms
B	3	Electronic games system
K	2 (a)	TV/ computer in other rooms
K	3 (a)	Electronic games system, mobile, iPod
K	22	Required to look for work
K	23	Partner working
K	24	Income

(a) Number of missings was exacerbated by the instruction that appears under Q1 that said “If you do not have a computer at home, go to Q6”.

2. Instructions to skip questions. There were a number of questions where the lead in instructions requested that only people with particular characteristics (eg, people who are currently working) complete the following question. It appears that some respondents may have skipped reading the preamble and made their own decisions about whether the question was relevant to them. Where this inconsistency led to people answering the question who should not have, their data was removed. However missing data could not be replaced in most cases. This affected the following questions:

Cohort	Question no.	Description
B	23	Main reasons not in paid work
B	24	Plans about paid work

Cohort	Question no.	Description
B	31	Effect of government benefits on attitudes to work
B	32	Attitudes to work
K	4	Computer use at home
K	5	Internet use at home
K	26	Effect of government benefits on attitudes to work
K	27	Effect of work on school involvement
K	28	Attitudes to work

3. 'None of the above' questions. A number of questions provided a 'none of the above' option if none of the other response categories applied. Experience with other self-complete forms has shown that it is not uncommon for people to omit ticking 'none of the above'. In general, it could be assumed that many of the responses to questions that had no response categories ticked are in fact 'none of the above' - however some people may skip questions for reasons that are not readily apparent. This affected the following questions:

Cohort	Question no.	Description
B	2	TV/ computer in other rooms
B	11	Childcare
B	18	Life events
B	24	Plans for paid work
B	25	Government benefits
B	28	Current study etc
B	35	Child support arrangement services
B	43	Help from other parent
K	1	TV/ computer in child's bedroom
K	2	TV/ computer in other rooms
K	3	Electronic games system, mobile, iPod
K	11	Childcare
K	16	Life events

Cohort	Question no.	Description
K	21	Government pensions
K	31	Child support arrangement services
K	39	Help from other parent

4. ‘Yes/ No’ questions. As for ‘none of the above’, it seems that some of the missing data for these questions could be explained by respondents for whom the ‘no’ response was relevant, omitting to tick the ‘no’ option. In addition, if the question included a ‘go to’ instruction, then sometimes respondents forgot to tick whichever option they selected. The main questions affected are:

Cohort	Question no.	Description
B	16	Is study child the youngest
	19	Do you currently have a paid job
	33	Does the study child have a PLE
K	14	Is study child the youngest
	17	Do you currently have a paid job
	20	Are you currently looking for work
	22	Are you required to look for work
	29	Does the study child have a PLE

5. Questions where ‘0’ is a valid response. These are often left blank as respondents feel they don’t apply. This affected Q14 for the B-cohort (number of changes to childcare arrangements) particularly since those currently without arrangements had been instructed to skip the previous two questions.
6. There were a number of cases in the B cohort (7) and the K cohort (29) that had roughly 90% of missing data. These cases have been excluded from the raw data files and the final files.

### ***Rationale for imputations***

#### **Presence of media devices in the home and amount of time spent using these devices (Q2 to 4 B-cohort)**

An attempt was made to impute whether a child had access to the facilities listed in these questions by whether they had reported using them at Q4. Unlike many of the other imputations mentioned in this attachment, respondents were expected to answer the subsequent question regardless of response to the previous ones, so some meaningful checks of the concordance between responses was possible.

However, most of the children (30 out of 41) who don't watch any TV still have TV in the home, so it couldn't be assumed if children don't watch TV, they don't have access to one.

In Q1 and Q2 on the devices in their home, 46 of the 52 children were reported as watching some television in the home, even though these questions indicated they did not have a TV. While this indicates a misinterpretation of at least one of the questions, it would appear that presence of a television in the home couldn't be reliably inferred from a response indicating that the child watches television in the home.

So therefore we can neither confirm nor refute the presence of a television in the home from the response to Q4a.

Likewise, 268 out of the 384 children that don't use a computer still indicated that they have one in the home and 30 of the 51 children without a computer still use one. Again the correspondence between the items isn't reliable enough to impute a response on whether there is a computer in the home.

For respondents with non-missing data for both the device ownership question (Q3) and the amount of time spent playing computer games (Q4) only 25 of the 362 cases where the child doesn't have an electronic games system do they play one. Also, in 385 of the 435 cases where the child does have an e-games system they spend some time playing them. Since this data follows the basic correspondence that would be expected, the presence or absence of a console has been imputed by whether the child plays with one when this information is missing. This has added 141 'no' responses and 107 'yes' responses. It has also been imputed that if the child doesn't have access to an e-games system at home, the time spent playing with one at home will be nil, altering 1,541 responses from missing.

### **Devices in the home and possession of personal devices (K-cohort, Q2 and 3)**

No checks are possible on television use, electronic games systems, or i-pods. For mobile phones, there is no implication in the 'use of mobile phones' items at Q6 that the child has to use their own mobile phone to do these things, so ownership of a mobile phone can't be imputed.

For computers and internet, cases where the child has a computer in their room are actually a little less likely to have one somewhere else in the home (87% v 93%), so it can't be assumed if they have one they'll have the other. Generally, if the respondent has given good answers to Q4 and 5 they do have a computer in the home, but it's not necessarily the case that because the respondents has answered these questions incompletely that they don't have a computer. Therefore it's impossible to impute accurately.

### **Presence or absence of childcare (B-cohort Q11, K-cohort Q11)**

If the respondent indicated that the child did spend time at childcare, the child was assumed to have an 'other' type of childcare for the K-cohort, however for the B-cohort they were imputed as having childcare but the type of childcare was set to missing since no 'other' option was available. This affected 3 cases in the B-cohort and 2 cases for the K-cohort. If the respondent reported 0 for number of days or hours per week of childcare it has been imputed that the child had no childcare. This affected 1 case for the B-cohort and 2 cases for the K-cohort.

### **Is the study child the youngest child in the home (B-cohort Q16, K-cohort Q14)**

The study child was aged either 3-4 years for the B-cohort or 7-8 years for the K-cohort at the time of the Wave 2.5 questionnaire. Therefore if the respondent indicated that the age of their youngest child corresponds with this, it has been imputed that this is the study child, effecting 6 cases for the B-cohort and 14 cases in the K-cohort. If the age given was younger than this it has been imputed that this wasn't the study child, effecting 14 cases for the B-cohort and 4 cases for the K-cohort.

### **Does the respondent have a paid job (B-cohort Q19, K-cohort Q17)**

If the respondent indicated they did work some hours then they were imputed to have a paid job, effecting 2 cases for the B-cohort and 13 for the K-cohort that were previously missing. If they said they generally work 0 hours then they were imputed to have no job, effecting 1 case for the B-cohort. If B-cohort respondents were missing data for work hours and their desired number of work hours, but had a response for why they were not currently working or their future work plans, they were imputed as being out of work, affecting 4 cases for the B-cohort. This question was not asked for the K-cohort so no similar imputation was possible. Some of the remaining missing cases had data for the desired work hours question, however those with or without a job could logically answer this question, so this provided little indication of the true response to whether they were working. The attitude items for those in work (Q32 for the B-cohort, Q27 and Q28 for the K-cohort) could also be answered by some non-workers on the basis of previous work experience, so imputation based on responses to these was deemed unreliable.

### **Whether government benefits are received (B-cohort Q25, K-cohort Q21)**

For the K-cohort, if the respondent indicated that they are required to do an activity test, it could be imputed that a benefit is received, however none of the missing cases met this criteria. The only other possibility for imputing this question would be to look at the effect of government benefits on work plans items (B-cohort Q31, K-cohort Q26), however there is no way of knowing if the respondents that didn't answer these questions were getting family tax benefits. Also, the skip is not very well highlighted in the formatting, so there can be little confidence that those who answered the question understood whom it was for.

### **Whether the study child has a PLE (B-cohort Q33, K-cohort Q29)**

A number of the missing cases have been classified on a case-by-case basis based on responses to the follow up questions on child support. The criteria for these classifications involved the amount of missing data, the amount of data that might indicate the presence of a PLE (e.g. having a child support arrangement vs not having one), as well as whether a PLE was present at Wave 2. This created 1 extra 'yes' response and 1 extra 'no' response for the B-cohort and 4 extra 'no' cases for the K-cohort. After this process there were 16 cases that were missing all subsequent information for the B-cohort and 9 for the K-cohort. Most children do not have a PLE, however some of these cases could be from people who gave up on the questionnaire. It can be reasonably assume that if they answered 50% of items in the most recent question required of them that they haven't given up on the survey and therefore are just cases without a PLE. This added an extra 12 'no' cases for the B-cohort and 3 for the K-cohort.

### **Respondent information (B-cohort Q47)**

Initially there were 100 records that were missing the respondent information (“who completed this form?”) in the B cohort file and 85 records in the K cohort file. ABS was able to correct 88 B cohort records and 69 K cohort records by matching the names of the people that completed the Wave 2.5 form to the names of people who participated in Wave 2. The location of this question may have been a factor in why there are missings, because the question is located at the bottom on the back of the form.