Reproductive Variables

Age Cohorts	1973-78 (Young),
Surveys	All (1 to 6)
Derived Variables:	Birth (including still births), term, misc (Surveys 1 ,2, 3, 4)
Definition	Number of births, terminations, miscarriages
Statistical Form	Discrete variable
Index Number:	
Prepared by	David Fitzgerald
Endorsed	

Background

The surveys for the 1973 – 78 cohort (previously known as the Young cohort) have questions asking about the reproductive histories of the ALSWH women. This document explains the ALSWH data relating to number of births, terminations, and miscarriages. These data have been re-coded by longitudinally cleaning. That is, they were changed when there were inconsistencies across surveys. This is not generally done to ALSWH data.

The reproductive questions were not always consistent across the surveys of this cohort. The first two surveys only asked about the number of certain reproductive events or counts. From Survey 3 onwards the number of reproductive counts and also the dates of each birth were requested. Surveys 4 and 5 (and possibly subsequent surveys) also asked about complications within each birth. This information is contained in the CHILD data sets and is not discussed here.

In Surveys 1 to 4 there are derived reproductive variables: births, terminations, and miscarriages. These were derived from survey responses that had been longitudinally cleaned. This longitudinal cleaning has not been continued after Survey 4 and nor were these derived variables calculated. However, it would be a simple matter to derive them from the survey data, although it would not be a simple matter to longitudinally clean the reproductive variables.

There are other ALSWH reproductive variables that are not mentioned in this document. The purpose of the document is not to list all the reproductive variables but to explain the longitudinal cleaning of the survey questions that lead to the births, terminations, and miscarriage counts.

Reproductive Counts

Each 1973-78 survey has a question asking the number of reproductive events.

Table 1 Reproductive Counts

Survey	How many times have you had each of the following:	Survey Items	Responses:
1	Pregnant, Miscarriage, Termination, Given birth	Y1q22a – y1q22d	0 to 4 or more
2	Live birth, Live premature birth, Stillbirth, Miscarriage, Termination	Y2q35a – y2q35e	1 to 5 or more
3	Live birth, Live premature birth, Stillbirth, Miscarriage, Termination	Y3q35a – y3q35e	0 to 5 or more
4	Live birth, Live premature birth, Stillbirth, Miscarriage, Termination for medical reasons, Termination for other reasons, Ectopic pregnancy	Y4q33a – y4q33g	0 to 5 or more
5	Live birth, Stillbirth, Miscarriage, Termination for medical reasons, Termination for other reasons, Ectopic pregnancy	Y5q35a – y5q35f	0 to 5 or more

The questions have changed with each survey. More importantly, all the Survey 2 questions did not have a zero option. This was a mistake and it lead to many missing values, because it was unclear whether a missing value to any question was indeed missing or a zero response.

At the time when the fourth survey data was available the reproductive data in all Surveys 1 to 4 were edited to make them consistent. This editing meant that the number of births, terminations and miscarriages was not allowed to decrease.

Here is a list of all the longitudinally cleaned variables for each survey

Table 2 Longitudinally Cleaned Reproductive Variables

Survey	How many times have you had each of the following:
1	Y1birth, y1term, y1misc
2	Y2birth, y2term, y2misc , y2q35a, y2q35b, y2q35c
3	Y3birth, y3term, y3misc , y3q35a, y3q35b, y3q35c
4	Y4birth, y4term, y4misc , y4q33a, y4q33b, y4q33c, y4q33e, y4q33f, y4q33g (not y4q33d)
5	None

Note that the some, but not all, of the reproductive count questionnaire items were longitudinally cleaned. Also note that the variable *birth* (y1birth - y4birth) includes full term, premature and still births.

The child's dates of birth were also used to longitudinally clean the reproductive counts in the second stage of re-coding.

Table 3 Child Dates of Birth

	If you have ever given birth to a child, please write the date of each birth in the box. (If you had twins, please write the date twice.)

Survey	*Variable*
1	None
2	None
3	Y3Childbd1 – Y3ChildBd8
4	Y4Childbd1 – Y4ChildBd8
5	Y5Childbd1 – Y5ChildBd9

All these dates have been converted to be the 15th of the month of birth. This was to de-identify the respondent. They are kept in a date format, for example, the SAS date format.

Recoding Reproductive Counts

The reproductive counts values are available in the survey data under the question number. For example, in Survey 3 the number of live full-term births is the variable y3q35a. For most ALSWH data the survey question values were exactly what the respondents entered, however, these reproductive counts have been longitudinally cleaned. A complete explanation of the first stage of the cleaning/ recoding is in the ALSWH Technical Report 29[1] . A brief summary is given here.

First Stage : Recoding of Reproductive Variables

There were two motivations for recoding the reproductive variables. One was that the Survey 2 questions did not have the zero option and so the zero values could not be distinguished from the missing and they had to be imputed somehow. The second reason was that some reproductive pathways did not make sense. There were some counts that decreased, which is not possible. For example, a woman may have reported one miscarriage at Survey 2 and then zero miscarriages at Survey 3.

The rule for recoding was that a count could not decrease. Therefore, once a value was established it must either remain that value or increase. If a subsequent survey value was lower than the earlier one it was replaced with the earlier one. This was only done to variables that had unambiguous pathways. An example of an ambiguous pathway is from Survey 1 to Survey 2 where the question on number of births divided into full term births and premature births. Therefore, knowing there was one birth at Survey 1 did not allow any re-coding at Survey 2, since it was not known whether that birth was full term or premature. However, if there was one stillbirth at Survey 2 and at Survey 3 the number of stillbirths was either missing or zero, then that value at Survey 3 would be changed to one.

After these recodes had been done at the question level, the summary variables births (full-term, premature, and still birth), terminations, and miscarriages *(birth*, *term*, and *misc)* were derived by adding up the number of counts.

Table 2 above has the list of all the longitudinally cleaned variables for each survey.

Second Stage: Using the Dates of Births in Longitudinal Cleaning

Only the *birth* variable was re-coded in the second stage. The preceding sectioned explained the first stage of how the derived variables, including birth, were calculated. The birth variable was further edited to use the child date of birth information. In surveys 3 and 4 the dates of birth for all children were reported. If the number of dates of births was more than the births count variable (y1birth to y4birth) then it replaced the lower birth count value at the relevant survey. The date of birth was used to determine at which survey the birth had taken place. For example, if there were two births counts recorded for Survey 3, but the respondent had also recorded three different singleton dates of birth in the child date of birth question and they all occurred before survey 3 had been returned, then the number of births variable, y3birth, would be increased to three. This method would not decrease the number of births. If fewer dates were recorded than the count number of births then the count number would remain the same. In essence, where the number of births did not agree then the larger of the two was used.

Longitudinally Cleaning Discontinued

For Survey 5 and later surveys the ALSWH Data Manager decided that this longitudinal cleaning should not be continued. The reason not to continue was primarily because there was no need. The number of each birth event is available on the questionnaire data. The variables y5birth, y5term, and y5misc could easily be created from the survey data, although they would not be exactly equivalent to the variables on the surveys 1 to 4, since these had been longitudinally cleaned. A user could somewhat recreate the cleaning by raising any value that is lower than the previously value. It should be noted that this recoding will not necessarily be correct, because any mistakes from early surveys would be carried forward. Another reason for not continuing this longitudinal cleaning is that the methodology may not be sound. For example, if a woman mistakenly over-reported the number of events and in each subsequent survey correctly recorded the true lower amount, the higher incorrect value would always replace the corrected value. Furthermore, the initial problem of correcting the 'zero or missing' problem at survey 2 had been completed.

References

1. Australian Longitudinal Study on Women's Health Technical Report 29 , 2007, pp 103 – 109. Available at http://www.alswh.org.au/Reports/Technical/Report29ALSWH.pdf [Accessed at 14 July 2011]