Calculating the Age of Children More Precisely

The calculation of children's age has recently changed in DHS surveys. Get all of your questions answered about what changed, why the change happened, and how this might affect your analysis.

What has changed in the questionnaire?

For most of DHS history, the DHS woman's questionnaire collected month and year of birth of the respondent and her age, month and year of marriage or age at marriage, and month and year of birth of each of her children as well as the age of living children. As part of the biomarker data collection, day of birth was also collected for children age 0-5 years and was used in the calculation of the anthropometric indices. Beginning with the DHS-7 questionnaires, (surveys with fieldwork in about 2015 and later), the woman's questionnaire collects the **day of birth** for all children listed in the birth history (all children of the respondents to the woman's questionnaire) in addition to their month and year of birth.

Note that while DHS had previously collected day of birth for children weighed and measured, these data were not used in conjunction with the birth history data as they were not available for a) children who had died, b) children of the respondent who did not live in the household, c) older children, and d) children for whom height and weight was not collected if it was only collected in a subsample of the households. Additionally, it may also have been reported by a different respondent than the mother of the child.

Why was day of birth added for children in DHS-7?

Adding day of birth permits calculating the age of children more accurately. Calculating age in months using just month and year of birth and month and year of interview meant that age in months could be off by one month in approximately half of all cases. For example, a child born February 2017 was considered as 3 months old in May 2017, however, if the birth took place on February 25, 2017, and the interview was May 3, 2017, then the child is actually only two completed months old. Thus, if the day of birth is greater than the day of interview (roughly half of all cases) then the age would be over-estimated by one month. Similarly, a birth in May 2016 would be considered as 12 months (1 year) old in May 2017, but if the birth is on May 26, 2016, and the interview on May 3, 2017, then the child is 11 months (0 years) old. For most analyses, this difference has very little effect, but for a few, it can matter. Note that for anthropometry a different procedure is used and age in days is calculated as part of the creation of the anthropometric indices, but this is restricted to living children under the age of 5 that were reported as living in the household.

Why wasn't day of birth added before?

Historically DHS has not collected the day of birth of all children as the quality of reporting of dates of births and ages was simply not good enough to use more accurate calculations of age. The exception to this was the collection of day of birth for living children under the age of 5 who live in the household in order to be used in the anthropometric calculations. For this narrow group, the quality of data was acceptable but was not previously considered good enough for older children or for children who had died. The quality of date and age reporting has improved over time and by the start of DHS-7, The DHS

Program felt that the quality was sufficient to introduce the collection of day of birth and capture sufficiently reliable data.[1]

How was age calculated previously?

Previously, DHS calculated age for children by subtracting the month and year of birth from the month and year of interview to give age in months.

For analysts, the age of the child in months was calculated as follows:

age = V008 – B3

where V008 is the century month code[2] (CMC) of the date of interview, and B3 is the CMC for date of birth of the child. The CMC date of birth (B3) is calculated using a combination of the reported month and year of birth and reported age for living children as well as an imputation process for incompletely or inconsistently reported information.

How is age calculated now?

In DHS-7, The DHS Program introduced the calculation of age taking into account the day of birth and the day of interview. To do this, DHS introduced a new concept – the century day code (CDC). The century day code is analogous to the century month code and gives the number of days since the beginning of 1900. The century day code for a date of birth is computed in a similar manner to the century month code, using the reported day, month and year of birth, and the reported age in years for living children, as well as an imputation process for incompletely or inconsistently reported information. Note that for simplicity in the calculation of the century day code the assumption is that 1900 is a leap year, but this is also how Excel works too. This doesn't affect any calculations of age as all calculations are using a consistent base.

Age in years is then calculated by subtracting the century day code for the date of birth from the century day code for the date of interview, and dividing the result by 365.25 (allowing for a leap year every 4 years) and taking the integer part of the result to get completed years. For age in months, divide the difference between the date of interview CDC and the date of birth CDC by 30.4375 (365.25 divided by 12 months), and then take the integer part to get completed months.

What has been added to the datasets, and what's changed?

Several variables related to the century day codes have been added, as well as several new age-related variables:

V008A: Century day code (CDC) for date of individual interview, similar to the century month code variable V008.

---

[1] In fact a very small number of surveys prior to DHS7 had also included day of birth for children, and a similarly small number of recent surveys also included day of birth for the respondent, however, day of birth has not been taken into account in the calculation of age for those earlier surveys or for the age of the respondent.

[2] Century month code (CMC) is the number of months since the beginning of 1900, calculated as follows:
CMC = (Year – 1900) * 12 + Month. Thus January 1900 = CMC 1, January 2000 = CMC 1201, May 2017 = CMC 1409

B17: Imputed day of birth for children of the respondent (similar to the imputed month and year of birth in B1 and B2).

B18: Century day code (CDC) for date of birth of children of the respondent.

B19: Age of child, or months since birth for children who have died, in completed months, computed as:

B19 = int((V008A - B18) / 30.4375)

HV008A: Century day code (CDC) for date of household interview

HV807A: Century day code (CDC) for date of biomarker data collection

HML16A: Age of child in months for children. Used for reporting malaria testing for children

HC1A: Age of child in days for children included in the biomarker questionnaire

HC20: Century day code (CDC) for date of biomarker data collection for children. If date of measurement is included for individual children the variable may be different than HV807A, for example when retaking child's measurements

HC32A: Century day code (CDC) for date of birth of children included in the biomarker questionnaire

HW1A: Age of child in days for children included for anthropometry in the biomarker questionnaire

Note that equivalent variables to those above are actually created in the raw data files at the time of the imputation process, and it is these raw data variables that are used to create the above set of variables in the recode files.

Changes have also been made to the method of calculating several existing variables:

B8: Age of child in years – now calculated based on B19, instead of V008 - B3

B11, B12: Previous and succeeding birth interval. Previous birth interval was previously calculated as B3(i) - B3(i+1), but is now calculated as int((B18(i) - B18(i+1)) / 30.4375), and a similar change is made for the succeeding birth interval

V208: Births in the five years preceding the survey

V209: Births in the 12 months preceding the survey

V222: Interval between last birth and date of interview in months

V238: Births in the three years preceding the survey

V337: Months of use of current contraceptive method

HC1: Age of child in months for children for whom anthropometric measures were taken, computed as follows:

HC1 = int((HV807A - HC32A) / 30.4375)

HW1: Age of child in months for children of respondents for whom anthropometric measures were taken, as follows:

HW1 is set equal to HC1. Take note that HC1 and HW1 are calculated with reference to the date of biomarker data collection, and this can occasionally differ from the date of individual interview, and so, on rare occasions, HW1 and B19 may differ slightly.

HML16: Age of child in months for children included in the malaria bed net roster

Additionally, anywhere that a restriction based on the age of the child or the number of months since a birth, e.g. selecting all children born in the last five years, the condition has been changed to refer to B19 instead of V008 - B3.

How does this affect analysis?

In surveys that introduced the day of birth of the child, changes have been made in the analysis of the data in two main ways:

1) The restrictions on the denominator for tables now all use the age variables based on the calculation to the day, rather than to the month as was previously done. In most cases, this means changing selections such as

$$\text{if } (V008 - B3 < 60)$$
$$\text{to}$$
$$\text{if } (B19 < 60)$$

If the original restriction is applied, the analysis will exclude a small number of cases that are included with the newer calculation. On average this will add approximately half a month's worth of additional births or children into the analysis.

2) All background age group variables used in analysis are now based on the revised ages. Previously, on average, because the calculation method only considered month and year and not day of birth, the age group of 0 months would have roughly half the number of cases of age group 1 month or other older single month age groups. With the new method, age group 0 months will have a roughly similar number of cases as other single month age groups.

These changes affect virtually all tables related to children, particularly to children under the age of five.

Fertility rate and childhood mortality rate tables are not currently changed as these tables exclude the month of interview from calculations and effectively use complete months in the calculations.

More precise calculation results in a shift in age

The below diagrams show the age of the child calculated using the old and new methods, given a particular month of interview and month of birth, giving examples here for interviews in January to June of 2017, and births in December 2015 to June 2017. For any birth taking place on a day in the month on or before the day of interview there is no change in the calculation, but for any birth taking place on a day in the month after the day of interview the age of the child is now calculated as 1 less than previously. For example a child born in late April 2017 and included in an interview in early June 2017 (equivalent to a point in the bottom right corner of box "2" in the first row below, marked with a red star) was calculated as 2 months using the old method, but looking at the equivalent position in the second example, this child is calculated as age 1 month in the new calculation method.

Old age calculation method example:
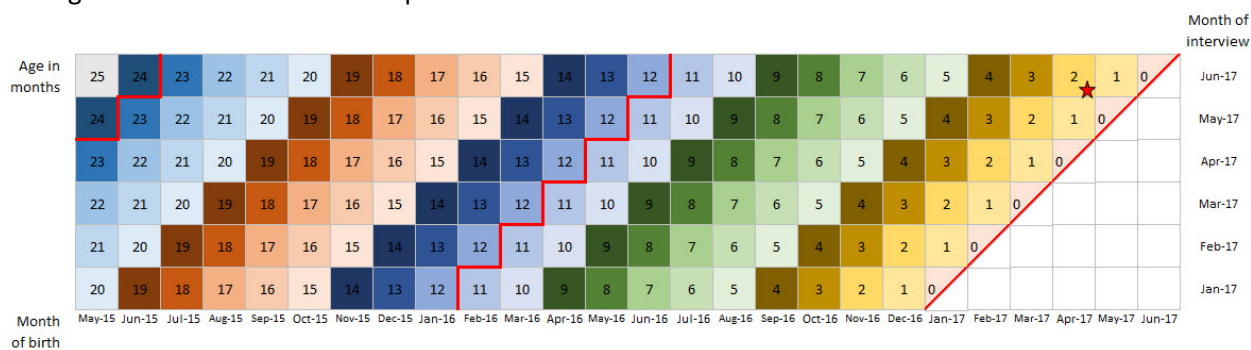
New age calculation method example:
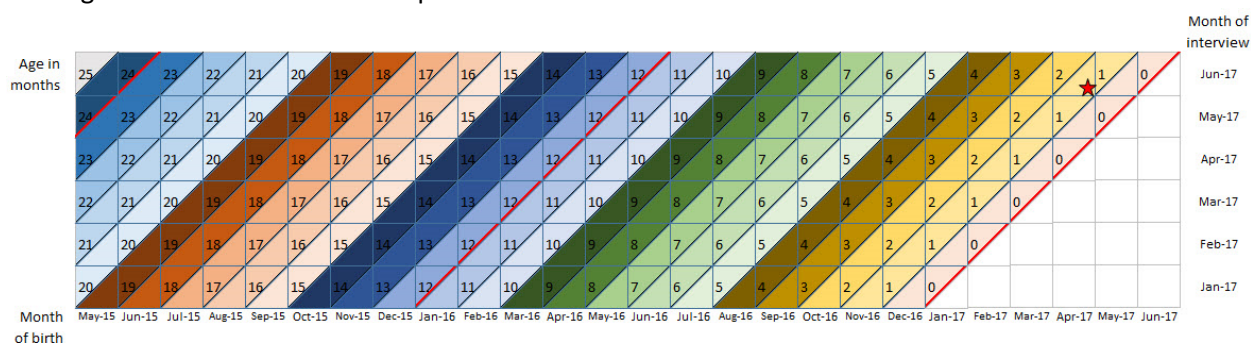
This shift in age in month affects roughly half of all children, but only has an effect on age in years for roughly 1/24 of children – those previously classified as 12 months old, but now classified as 11 months old, and similarly around ages 24 months, 36 months, etc.

Notes for Stata users:

a) The variable names used above are generic forms, not specific to any software. When referring to any variables mentioned above, use lower case letters in the variable names.

b) To understand the calculation of b18 and b19, you can think of them as being:

```
gen b18=mdy(b1,b17,b2)+21916
* mdy uses Jan.1, 1960 as its base – adding 21916 adjusts to Jan.1, 1900.
gen b19=int((v008a-b18)/30.4375)
* 30.4375 days in a month on average = (365.25/12)
```

c) You can generalize code to work with datasets that do not include b19 and later datasets that include the new age calculation by using code such as:

```
capture confirm variable b19
if _rc { // b19 does not exist, so create equivalent for old calculation method
 gen b19 = v008 - b3
 label variable b19 "Age of child in months or months since birth"
}
recode b19 …, gen(agegrps)
…
tab agegrps xxx [iw=wgt] if b19 < 60
```

This approach will use b19 it exists in the dataset, but otherwise, create its equivalent using the old age calculation method, and allow the production of analyses that are consistent with the tabulations in the DHS reports.

In general, avoid using references to b3 (CMC date of birth) in most cases (except for the above) and instead refer to b19 (age of child in months or months since birth) instead.

Notes for SPSS users:

d) To understand the calculation of B18 and B19, you can think of them as being:

```
compute V008A = yrmoda(V007,V006,V016) - (yrmoda(1900,1,1)-1) + 1.
compute B18 = yrmoda(B2,B1,B17) - (yrmoda(1900,1,1)-1) + 1.
compute B19 = trunc((V008A-B18)/30.4375).
```

[The century day codes (CDC) algorithm assumes for simplicity that 1900 is a leap year (as does Excel), so it is necessary to take that into consideration, thus the +1. This does not affect our method of calculating ages as the base is the same in all cases.]

e) You can generalize code to work with datasets that do not include B19 and later datasets that include the new age calculation by using code such as:

```
* check recode type is earlier than DHS7.
if (char.index("123456789", char.substr(V000,3,1)) < 7) B19 = V008-B3.
variable label B19 "Age of child in months or months since birth".
recode B19 … into agegrps.
…
compute filter_$=(B19 < 60).
filter by filter_$.
crosstab tables=agegrps by xxx.
```