**User guide on IPUM DHS calendar data (with examples)**

# Calendar data: an overview

For many surveys, DHS interviewers collect "calendar data” or “reproductive calendar data” from women of childbearing age. These calendar data are longitudinal data covering at least the preceding 5 years (although the length of the calendar varies from woman to woman, depending on the month of interview). The data provide a month-by-month history of reproductive events that occurred to a woman during her calendar period. For every month of the calendar period, the calendar data document whether a woman gave birth, was pregnant, terminated a pregnancy, or used a (specified) contraceptive method, and, if she stopped using a contraceptive method, the reason for stopping. While these events represent the core of the calendar data, some samples supply additional types of calendar data, including: marital status, source of contraception, moves and types of communities, breastfeeding, postpartum amenorrhea, and specific type of pregnancy termination (abortion, miscarriage, or stillbirth). The calendar data are collected and stored as part of the Individual Recode (IR) file.

The image below, taken from the final report for the India 2015 survey, reproduces the form the interviewer fills in while collecting the calendar data, with the country-specific codes used to the left. Here, the three columns collect data on three topics: births, pregnancies, and contraceptive use; antenatal ultrasound; and reason for discontinuing contraceptive use.

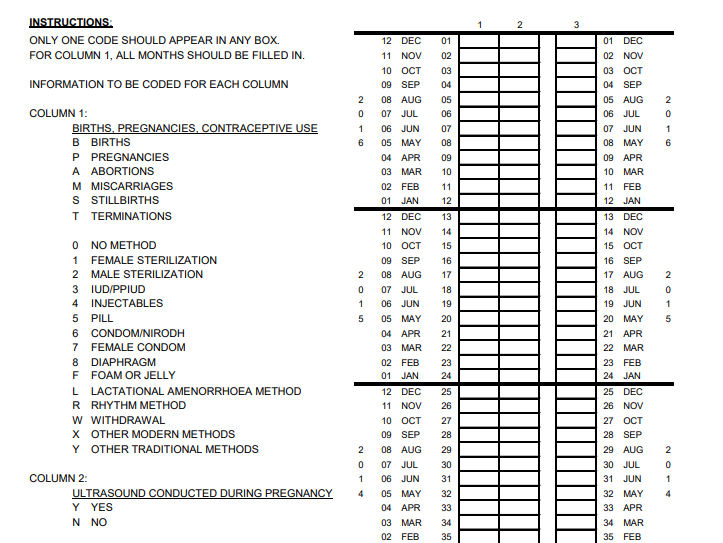
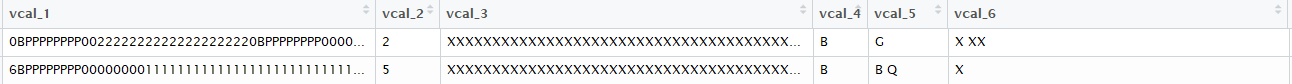


Fig. 1. Calendar data questionnaire (from India 2015 DHS)

For information on the question sequence the interviewer uses to collect the calendar data and how the responses are recorded during the interview, consult modules 1 and 2 of The DHS Program's <link id="83">"Contraceptive Calendar Tutorial"</link>.

In the original DHS data, the calendar variables are stored as 80-character strings. In principle, each character represents a month of the calendar period, so up to 80 months of reproductive history can be stored in the calendar. In practice, the duration of calendar data varies across women, but a total of 80 characters are stored for all women, with months without meaningful calendar data filled with blank spaces. In the IR file, calendar variables are named as follows: VCAL\_1, VCAL\_2, … VCAL\_9. To check how many calendar variables are in a given DHS sample, one can look at variable v019a or at a sample-specific Final Report.

When you read the original DHS calendar data from the IR file into a statistical package, the calendar variables will resemble the image below:



Here, every row represents a unique woman from the IR file, and the characters in vcal\_1 denote various events that occurred to that woman, with only one reproductive event (birth, pregnancy, pregnancy termination, or contraceptive use) recorded per month. Note that the positions of characters in all variables relate to a chronological month in a woman's reproductive history, starting from the most recent month. As such, the very first character in these 80-character strings represents the most recent month of the calendar for the woman.

Interviewers use the country-specific codes documented in a sample's final report when recording responses for a given survey. The DHS Program recodes those responses into a standard coding system when the calendar data are added to a public use IR file. In some cases, however, a response falls outside the standard coding system's categories. If so, the meaning of that response is made clear in the final report documentation.

The calendar data available from IPUMS DHS are recoded using the standard coding system for DHS Phase VII as a starting point. Those standard Phase VII codes make considerable use of letter characters, however, indicating births, pregnancies, and pregnancy terminations by B, P, and T, respectively, for example. To make the calendar data easier to manipulate by a statistical package or programming language, IPUMS DHS recodes alphabetic characters with numeric codes, assigning a code of "100" to replace the "P" for pregnancy, for example. IPUMS DHS staff also consulted final reports to determine the meaning of country-specific responses outside the standard coding system and gave numeric codes to those country-specific responses.

# Transformed calendar data by IPUMS-DHS

To analyze the original DHS calendar data, a researcher needs to manipulate or transform the data strings. The DHS Contraceptive Calendar Tutorial describes three ways of handling the string calendar data: 1) string parsing of the data; 2) restructuring the data into a file of single months; and 3) creating event files.

IPUMS DHS uses the second approach of restructuring the data into units representing woman months of exposure. These woman-months are available as a unit of analysis in IPUMS DHS, with the characteristics of the associated woman (e.g., her level of education) available along with the specific information collected and summarized for the calendar data. This conversion of the calendar data simplifies analysis of these valuable but underused longitudinal data.

In Figure 2 below, the left image shows the beginning of some original calendar data strings. Each string relates to a different woman, as shown by the woman's i.d. variable in the far left column. Each cell within the string holds data on the reproductive event (birth, pregnancy, pregnancy termination, or contraceptive use) that occurred within sequential months, starting with the most recent (last) month. For the woman with i.d. 182, for example, the first cell (with a value of "1") represents October 2004; the second cell (with a value of "1") represents September 2004; the sixth cell (with a value of "0") represents May 2004; and the cells directly following (a single B followed by eight P's) represent April 2004 through August 2003. According to DHS Phase VII standard codes, the woman used the Pill for the final two months, preceded by four months of using no method, preceded by a birth and 8 months of pregnancy.

The image on the right shows the transformed calendar data created by IPUMS DHS staff, for the same woman with i.d. 182. Here, each month of exposure (starting with the latest) represents a row. The second column documents the date of the month in century months. The third column shows the same events in each month as described above. However, string characters of "B" for birth and "P" for pregnancy are recoded into numeric values of "100" and "200" respectively, for easier manipulation.

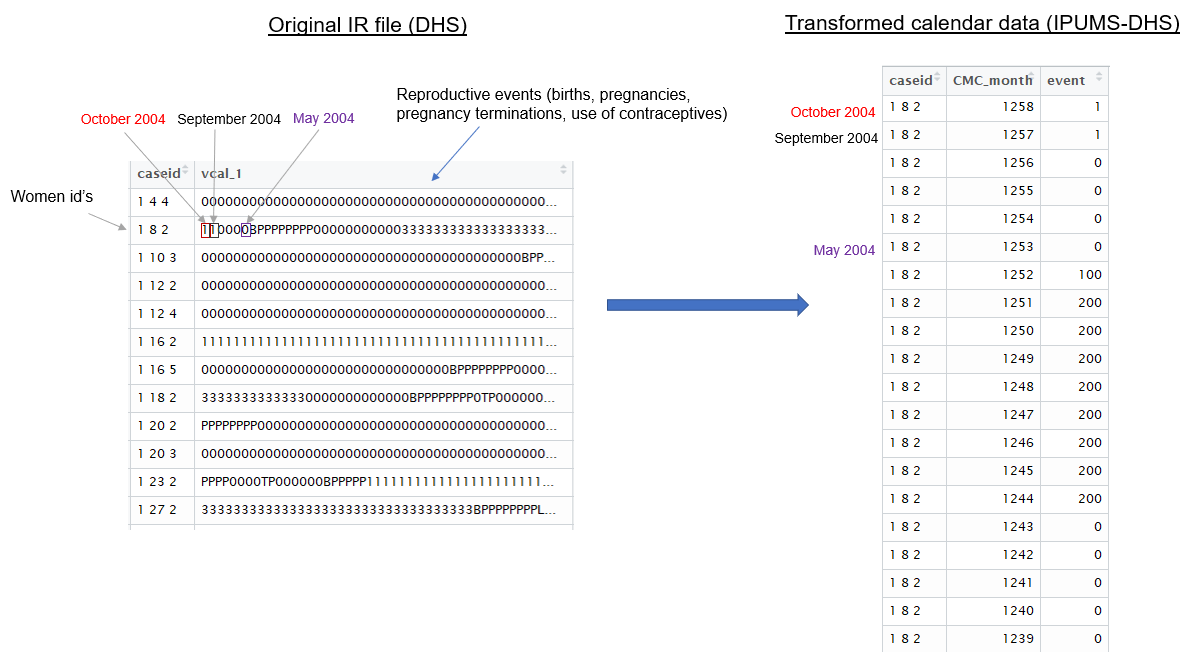


Fig. 2. Schematic representation of the transformation of the original DHS calendar data into women-months by IPUMS-DHS

Breaking up one string per woman into woman-months makes it easier and more intuitive to follow the history of the woman's reproductive health. To take advantage of the richness of the calendar data, after transforming the original strings into woman-months, IPUMS DHS staff also created a number of summary variables based on the transformed calendar data. The next section describes these variables, while the following section provides data examples for three women with calendar data.

## Variables created based on vcal\_1 (births, pregnancies, terminations, and contraceptive methods)

CMC\_month - Century Month Code (CMC) for *every month* of the woman's calendar. The first CMC for all women in the sample is constant and is the century month code of January of the first year of the calendar. CMC\_month depends on the duration of the calendar for every woman, since women in the sample have reproductive calendars of different lengths (depending on their month of interview). Note that CMC\_month is listed for every woman in descending order, so it starts with the most recent month.

Event - This variable contains a single reproductive event that happened to a woman in every CMC\_month, as contained in the original variable vcal\_1. Event is based on breaking up the vcal\_1 string and attributing an event to the corresponding CMC\_month (see Figures 1 and 2 above). The Event variable can contain only one code per month and is created by recoding alpha-numeric codes from vcal\_1. The variable has the following codes: 100 – birth, 200 – pregnancy, 300 – pregnancy termination, and numbers 0-99 representing the use of specific contraceptive methods. (Absence of any of these events is represented by a 0.) Codes for contraceptive methods are based on standard codes from DHS Recode Phase VII, plus some additional numeric codes based on country-specific responses not recognized in the standard codes. For the full list of codes for Event, see the Codes and Frequencies section of the variable description for Event.

Preg - Binary variable indicating whether a woman was pregnant during that CMC\_month (1-yes, 0-no). Note that a woman was also pregnant during the month that she gave birth, so eight months of pregnancy followed by a live birth represent a 9-month pregnancy.

Birth - Binary variable indicating whether a woman delivered a live birth during that CMC\_month (1-yes, 0-no).

Term - Binary variable indicating whether a woman terminated a pregnancy/delivered a non-live birth during that CMC\_month (1-yes, 0-no).

Contr - Binary variable indicating whether a woman used any contraceptive method during that CMC\_month (1-yes, 0-no).

Total\_preg - Total number of months a woman was pregnant during the entire length of her reproductive health calendar.

Count\_birth - Total number of live births a woman had during the entire length of her reproductive health calendar.

Count\_term - Total number of terminations (abortions and miscarriages) and non-live births a woman had during the entire length of her reproductive health calendar.

Contr\_duration - Total number of months a woman used any contraceptive method during the length of her reproductive calendar.

Switch - Binary variable indicating whether in any given month a woman was using a different contraceptive method compared to the previous month (1-yes, 0-no). This variable indicates cases in which a woman 1) switched to a different contraceptive method or 2) started using a contraceptive method prior to not using any method in the preceding month. Comparison to previous months when a woman was pregnant, gave birth, or had a pregnancy termination does not make sense, so the month immediately after a pregnancy, birth, and/or pregnancy termination are filled with code 99 (NA) for Switch.

Switch\_new - Binary variable indicating whether in a given month a woman was using a different contraceptive method compared to the method used the month before (1-yes, 0-no). This variable describes instances of switching from one method to another, if a woman reported using a method in the previous month.

Duration\_1 … duration\_N - Duration of use of specific contraceptive methods. The number after the underscore refers to the method code as recorded in the variable Event. For example, duration\_2 refers to the total number of months a woman used an IUD during her reproductive calendar.

Seq – Sequential (in chronological order) number for the month of the calendar interview for every woman. For example, the earliest month of the calendar will have the value of seq=1. In some cases, you may wish to limit months of exposure to the five years before the year of interview, so all women have the same amount of time "at risk" of an event, despite being interviewed in different months during the survey year. In such cases, limit the months analyzed to those with seq less than or equal to 60 (the five years preceding the interview year.)

cml\_preg – Cumulative duration of pregnancy in months. The final month of a woman's pregnancy – whether it resulted in a live birth or termination – is counted toward the cumulative duration. For example, a woman who was pregnant for 8 months and gave birth in month 9 will have a cumulative duration of pregnancy equal to 9 months.

preg\_flag – Binary variable indicating whether the cumulative duration of a pregnancy was greater than 9 months (1-yes, 0-no).

prfirst – Binary variable denoting whether a woman was pregnant in the first (earliest) month of her calendar (1-yes, 0-no). If a pregnancy had begun before the start of a woman's calendar, it is not possible to determine accurately the total duration of such a pregnancy. Variable prfirst can be used to easily identify such cases.

trunc – Binary variable indicating whether a pregnancy is truncated at the end of the calendar (1-yes, 0-no). If a woman is pregnant during the last month of her calendar, we cannot find out the outcome of that pregnancy and include it in the data. This truncation can lead to discrepancies between the number of months a woman has been pregnant (total\_preg) relative to the number of pregnancy outcomes (births/terminations) captured in the calendar.

Id – Simple numeric id for every woman.

Caseid\_CMC – Unique woman-month variable based on concatenating a woman's unique caseid with the CMC code for every month of her calendar.

## Variables created based on vcal\_2 (reasons for discontinuation)

While vcal\_1 stores information about various reproductive events and the use of contraceptive methods, vcal\_2 stores the reasons why women stopped using a contraceptive method.

Reason - Reason for discontinuation of a contraceptive method. The variable is recorded for the last CMC\_month of using a contraceptive method. Codes for that variable use the DHS standard codes for Phase VII as a starting point. However, some standard alphabetic codes (e.g., "F" for "fatalistic") are recoded as numeric codes (e.g., "11" for "fatalistic") in IPUMS DHS to simplify manipulation. IPUMS DHS also adds numeric codes for some country-specific responses (e.g.,"IUD fell out") not recognized in the standard coding system. For the full list of codes for Reason, see the Codes and Frequencies tab of the variable description for Reason.

Disc\_event - Binary variable indicating whether a woman stopped using a contraceptive method during that CMC\_month (1-yes, 0-no).

Disc\_total - Total number of times a woman stopped using contraceptive methods throughout the entire length of her reproductive health calendar.

## Variables created based on vcal\_3 (marital status)

Married – Binary variable based on vcal\_3, refers to the CMC\_month when a woman reported being married/in a union (1-yes, 0-no).

## Variables created based on vcal\_4 (residence/moves and types of communities)

This variable is available for some older samples, such as Zimbabwe 1994 and Jordan 1997.

Move – Categorical variable reporting whether a woman moved communities or the type of community where she resided in a given month. Codes for these variables are country- and sample-specific and can be found in the respective DHS Final Reports. For example, Jordan 1997 has the following codes for the variable Move: (0-change of community, 1-Amman, 2-another city, 3-countryside/village, 4-outside Jordan). Zimbabwe 1994 has the following codes: (0-change of community, 1-main town, 2-other urban area, 3-rural area).

The content of the Move codes in calendar column 4 is shown from the 1994 Zimbabwe sample's final report in Figure 3, below.

IPUMS DHS has placed the country-specific codes for Move into a harmonized variable. For the exact meaning of some responses (e.g., "Amman" as "main city" for Jordan 1997), consult the country-specific comparability text in the Move variable description.

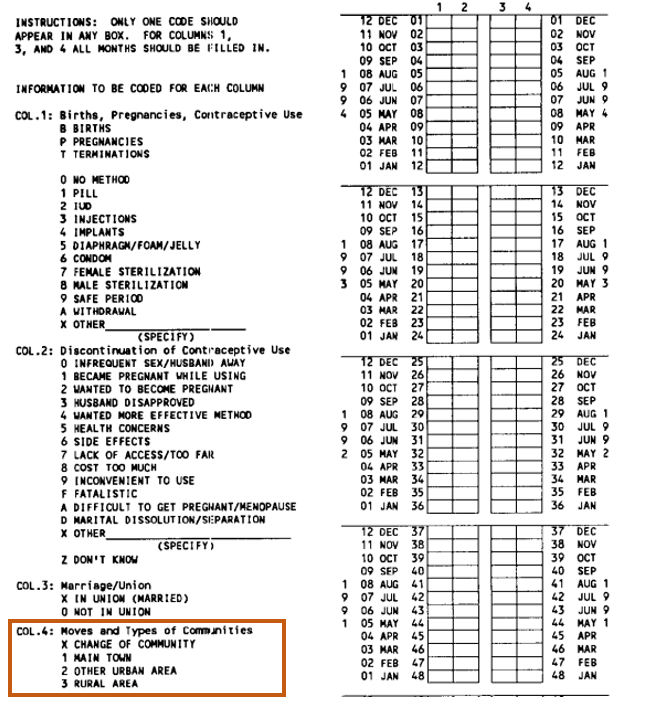


Fig. 3. An example of vcal\_4 from Zimbabwe 1994 DHS

## Variables created based on vcal\_5 (source of contraception)

contr\_source – A categorical variable reporting the source of contraception. The code is reported for the first month of the use of a method. Codes for this variable are country-specific and can be found in the respective DHS Final Reports. Examples for Egypt 2000 and Tanzania 2000 are shown directly below, in Figure 4.

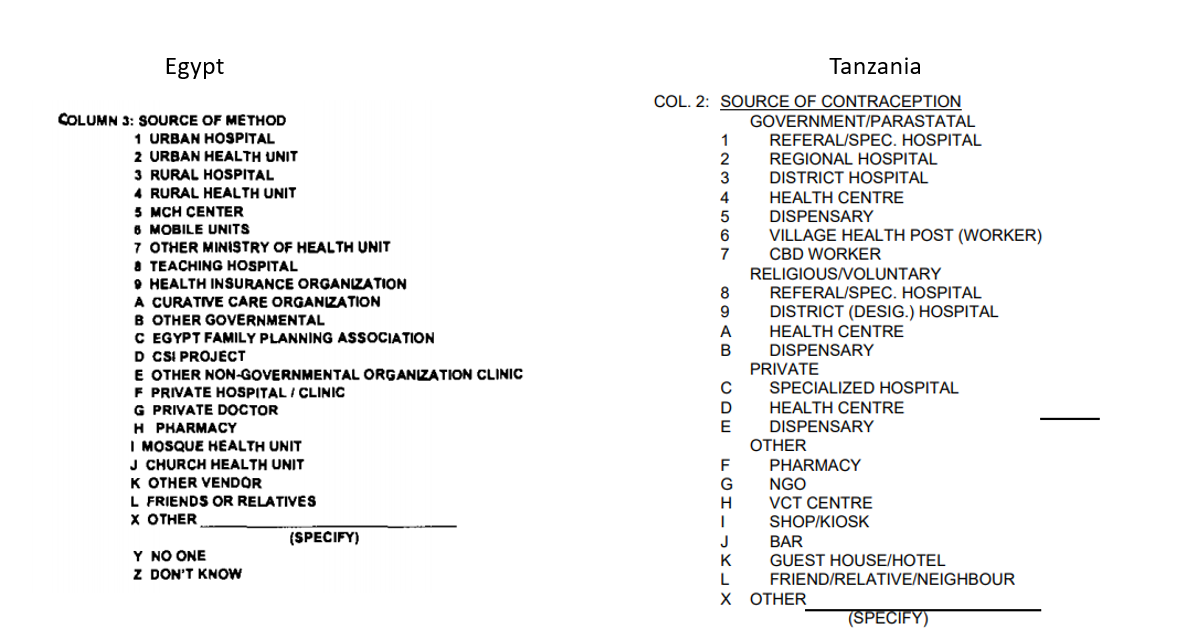


Fig. 4. Codes from calendar variable vcal\_5 from the Final Reports for Egypt 2000 DHS and Tanzania 2000 DHS

IPUMS DHS uses this country-specific information to create general and country-specific variables for the source of contraception reported in calendar data. The approach is similar to that followed for the "Known source for FP for non-users, general" and "Known source for FP for non-users, country-specific" variable groups in IPUMS DHS.

## Country-specific variables

Beyond variable vcal\_5, calendar variables vcal\_6 – vcal\_9 are country-specific. Only a few samples have country-specific calendar variables, and most countries have variables vcal\_1 and vcal\_2 only. To find out what calendar variables following the variable vcal\_5 represented, researchers had to consult the DHS Final Reports. Fortunately, IPUMS DHS staff have done that work for you, creating additional woman-month variables that are available for just a few samples. Here are a few examples:

Samples for India 2005 and 2015 have a variable about antenatal ultrasound checks:

ultrasound – A binary variable denoting whether a woman had an ultrasound during her pregnancy (1-yes, 0-no). This variable is recorded in the last month of a woman's pregnancy (i.e., the month of birth).

Some Egypt samples have the following variables:

ppam – A binary variable denoting whether a woman experienced postpartum amenorrhea in that month (1-yes, 0-no).

breastf - A binary variable denoting whether a woman reported breastfeeding her baby in that month (1-yes, 0-no, 2-never breastfed that child).

Finally, a number of samples have a calendar variable describing the type of pregnancy termination from vcal\_1. The position of that variable in the calendar recode varies across samples, and the codes are sometimes non-standard. This information on the type of pregnancy termination is put into the following consistently coded and consistently named variables in IPUMS DHS:

abort - A binary variable denoting whether a woman had an abortion in that month (1-yes, 0-no).

misc - A binary variable denoting whether a woman had a miscarriage in that month (1-yes, 0-no).

sbirth - A binary variable indicating whether a woman had a stillbirth in that month (1-yes, 0-no).

count\_abort - The total number of abortions a woman had during her calendar period.

count\_misc - The total number of miscarriages a woman had during her calendar period.

count\_sbirth - The total number of stillbirths a woman had during her calendar period.

Important note: Keep in mind that codes for contraceptive methods in the Final Reports differ from the codes in vcal\_1, in the DHS IR files, because the codes for vcal\_1 have been standardized by The DHS Program. These differences between contraceptive codes in final reports and the standard contraceptive codes in vcal\_1 are illustrated by two examples in Figure 5, below. **The standard contraceptive codes in vcal\_1 undergo further changes in the woman-month data from IPUMS DHS, which changes letter codes into numeric codes to simplify analysis.**

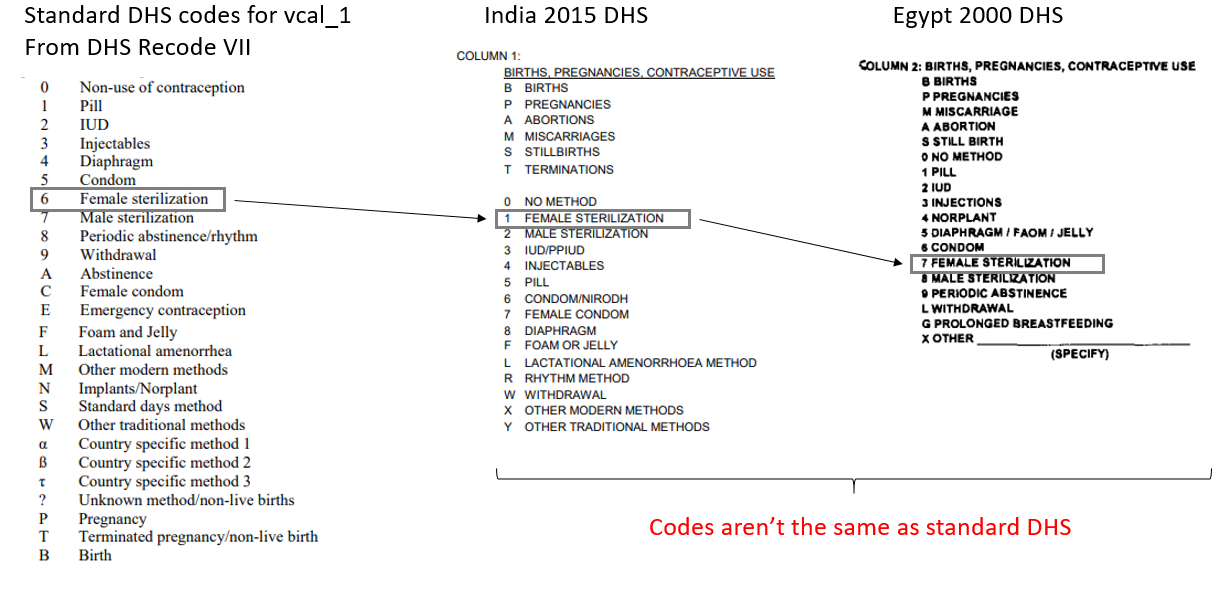


Fig. 5. Comparison of codes from country-specific DHS Final Reports and standard DHS codes

# Transformed calendar data example: Egypt 2000

Below is an excerpt from some calendar data as transformed by IPUMS-DHS, along with an interpretation of variables that IPUMS DHS staff have created. This material provides concrete examples of the relationship between reproductive events and the woman-month variables.

Keep in mind that you can select calendar variables directly to add to your customized data file through the IPUMS extract system, if you log in and use woman-months as the unit of analysis.

The examples below are based on the Egypt 2000 DHS. Egyptian DHS samples have more calendar variables than most other DHS samples. Overall, the Egypt 2000 DHS has the following calendar variables:

Vcal\_1 Pregnancies & contraception

Vcal\_2 Reason for discontinuation

Vcal\_3 Marriage & unions

Vcal\_5 Source of method

Vcal\_6 Postpartum amenorrhea (months)

Vcal\_7 Breastfeeding (months)

Vcal\_8 Original Pregnancies & contraception (includes type of termination). This variable is the same as vcal\_1, except that it describes a specific type of termination. So, if an event is coded as T (termination of pregnancy) in vcal\_1, in vcal\_8 it will represent either A (abortion), M (miscarriage), or S (stillbirth). In addition, the country-specific codes for contraceptive methods in vcal\_8 are not the same as the standard contraceptive codes in vcal\_1 (see note immediately preceding this example).

For these examples we will look at four women's reproductive health histories, derived from the calendar data in the Egypt 2000 DHS. Open *eg2000\_lr\_example.xlsx* to follow the examples below. Note that the calendar data are stored in reverse chronological order; more recent events appear first. This is how the data are originally stored in the DHS IR file. If you prefer earlier events to appear first, you can sort the data by the CMC\_month variable. The reverse order essentially means that we start reading a woman's history from the bottom up.

First, let's look at the woman with id=7 (in column C). This woman starts her reproductive calendar in Century Month Code (CMC) 1141 (in column B), and she reports being pregnant in that month (in column D). Since she reported a pregnancy in the first month of her calendar (which means we don't have enough information to ascertain how many months her pregnancy had lasted before that first month), variable **pfirst** (in column AB)has code 1 to flag that the woman was pregnant in the first month of her calendar. That pregnancy resulted in a termination in CMC 1146, as indicated by code **300** in the variable **event** (in column F). Because the Egypt 2000 DHS has data on the kinds of pregnancy termination, we can figure out how exactly the pregnancy terminated by looking at variable **event2** (in column AO) (or the original DHS variable **vcal\_8.dhs**). According to these variables, the woman had a miscarriage in CMC 1146. (also indicated by a "1" for the variable **misc** (in column AQ) in that month).

Then, in CMC 1151, the woman became pregnant and had a live birth in month 1159. In CMC 1160 she experienced postpartum amenorrhea, as indicated by variable **ppam** (in column AK). Starting in CMC 1160, the woman also breastfed for the next 11 months after giving birth, as indicated by variable **breastf** (in column AM). The woman became pregnant for a third time in CMC 1172 and had a live birth in CMC 1180 after 9 months of pregnancy, as indicated by the variable **cml\_prg** (in column AX). The woman experienced postpartum amenorrhea in CMC 1181, as indicated by variable **ppam**. After this third pregnancy, the woman reported never breastfeeding this child in CMC 1181 (variable **breastf**).

The woman got pregnant for a fourth time during her calendar period in CMC 1194 and had a live birth in CMC 1202 after 9 months of pregnancy. In CMC 1203 she reported postpartum amenorrhea and reported that she was breastfeeding.

Woman 7's calendar period spans a total of 63 months (indicated by variables **v019** or **seq** (in column E)), from CMC 1141 through CMC 1203. During that time, the woman was pregnant four times, for a total of 33 months of pregnancy (**total\_preg,** in column K). The first pregnancy resulted in a miscarriage in CMC 1146 (variables **misc** (in column AQ); **count\_misc** (in column AT) for the total number of miscarriages). The other three pregnancies resulted in live births (**count\_birth** in column L). The woman breastfed the first and the third births and did not breastfeed after the second birth. She also experienced postpartum amenorrhea for a month after each of her live births (**ppam**).

Woman 7 was married throughout the entire calendar period (variable married (in column AG) and did not report using any contraception (**contr\_duration** = 0 in column N).

As an example of tracking contraceptive use as well as other reproductive events using the calendar data, consider the woman with id=1130. That woman's calendar period started in CMC 1141 and ended in CMC 1203. Woman 1130 first became pregnant in CMC 1147 but reported a termination in CMC 1149. Variable **event2** (in column F) indicates that the pregnancy resulted in a miscarriage in CMC 1149 (see also variable **misc** in columnAQ). After that, the woman reported becoming pregnant again in CMC 1162, a pregnancy that lasted 9 months (variable **cml\_prg** in columnAX) and resulted in a stillbirth in CMC 1170, as indicated by the variable **sbirth** (in column AR). Then, in CMC 1186, the woman became pregnant for a third time and had a live birth in CMC 1194 after 9 months of pregnancy (**cml\_prg** in columnAX). After that birth, the woman reported postpartum amenorrhea (**ppam** in columnAK) for one month in CMC 1195 and reported breastfeeding for at least nine months after giving birth (**breastf** in columnAM), after which her calendar period ended.

In CMC 1197, woman 1130 reported using contraception, namely an IUD (variable **event** in column F), which she used for a total of seven months (variable **duration\_2** in column R). According to the country-specific codes for the source of contraception, the woman obtained her IUD from an urban health unit (code 2 in the variable **contr\_source** in columnAI).

Woman 1130 reported being married during her entire calendar period (**married** in column AG). Overall, she had three pregnancies that lasted a total of 21 months (**total\_preg** in columnK), had one live birth (**count\_birth** in columnL), two terminations (**count\_term** in columnM), which we know were one miscarriage (**count\_misc** in columnA) and one stillbirth (**count\_sbirth** in columnAU). The woman used contraception for a total of 7 months (**contr\_duration** in column N).

As a third example, let's look at woman id=3, who demonstrates how switching contraceptive methods is documented. Woman 3's calendar period started in CMC 1141, and she was pregnant in that month. In CMC 1144, the woman started using an IUD (code 2 in variable **event** in columnF), which she obtained from a private doctor (code 16 in variable **contr\_source** in columnAI). The woman used an IUD for 26 months, through CMC 1169. The woman stopped using the IUD in CMC 1169 because of health concerns, as indicated by the variable **reason** (in column AC). Variable **disc\_event** (in column AD) is coded "1" to indicate that the woman discontinued a contraceptive method in that month. In CMC 1170 the woman began using an injectable (code 3 in variable **event** in columnF), which she also obtained from a private doctor (variable **contr\_source** in columnAI). Because in CMC 1170 the woman was using a different method than in CMC 1169, variable **switch** (in column O) has code "1", indicating a switch in the method of contraception. The woman used injectables for a total of 3 months, before stopping their use in CMC 1172 because she wanted to become pregnant (code 2 in variable **reason** in column AC). This discontinuation is also reported in CMC 1172 for the variable **disc\_event** (in column AD). Three months after that, in CMC 1176, the woman became pregnant and had a live birth in CMC 1184. Two months after giving birth, the woman started using an IUD again, which she procured through a private doctor. The woman used the IUD until the end of her calendar period in CMC 1203. As shown by the variable **duration\_2** (in column T), the woman used an IUD for a total of 44 months, and the injectable for a total of 3 months, as shown in **duration\_3** (in column W). Overall, the woman used contraception for 47 months during her calendar period (variable **contr\_duration** in columnN), and had two episodes of discontinuation of contraceptive methods (**disc\_total** in columnAE).

Finally, consider woman id=28, who was pregnant during the last month of her calendar. The variable **trunc** (in column AZ) has code "1" for that woman, to indicate we do not have data on how her pregnancy resolved.

None of the women in these four examples had a pregnancy lasting more than 9 months, as indicated by "0" in variable **preg\_flag** (in column AY).