

2009–2010 Survey of HIV Status and Fertility Preferences in Sub-Saharan Africa (Community-Based Survey of Zambian Women)

Data Users' Guide

Data collected by the Guttmacher Institute

Public-use dataset prepared by the Guttmacher Center for
Population Research Innovation and Dissemination

April 2016

Contents

Introduction; about the data	1
Contact person	2
Available data formats	2
Weights and sample design variables	2
Other key variables	3
Missing values	3
Programming for recodes.....	4
Bibliography of articles based on the dataset	28

Introduction; about the data

These data come from one of eight surveys conducted in Zambia and Nigeria exploring how HIV status relates to attitudinal and behavioral measures regarding HIV services and fertility preferences. 1,441 Zambian women aged 18–49 took part in this community-based survey in 2009 and 2010. Information collected includes age, urban vs. rural location, education, religion, marital status, fertility desires, HIV status, pregnancy intention, current and previous pregnancies, attitudes toward and use of family planning, current and past sexual activity, past abortions, attitudes toward and knowledge about HIV, and attitudes toward and use of HIV services, among other variables. The final data file contains 858 variables.

Note that while this dataset only covers the community-based survey of Zambian women, there were also seven other surveys within this project on HIV and fertility desires. These surveys focused on men or women from Nigeria or Zambia, who were interviewed either at a health facility or in the community. All possible combinations of these three “variables” of survey type were covered, yielding eight total surveys ($2 \times 2 \times 2 = 8$). We plan to make the other seven data files publicly available over time. Researchers interested in those datasets should consult Guttmacher’s or DSDR’s data archive to determine if they are currently available.

We have generated a number of recodes to facilitate use of the data. Programming for these recodes can be found below in the “Programming for recodes” section. We have provided a codebook in a separate document with information on each variable in the dataset. The English version of the questionnaire that was used in the field in Nigeria is available as well. There are only very minor differences between this questionnaire and the one that was used in Zambia.

Contact person

For inquiries about this dataset, please contact Suzette Audam through the Guttmacher Center for Population Research Innovation and Dissemination at popcenter@guttmacher.org or via phone at (212) 248-1111.

Available data formats

The data are available in the following formats:

- A **Stata** data file (.dta) is available for Stata users.
- An **SPSS** data file (.sav) is available for SPSS users.

Weights and sample design variables

All analyses should be weighted. The main weight variable for community surveys is `wgt`.

In order to obtain accurate standard errors, analyses should also account for the fact that the sample was stratified by using the `stratum` variable (`strata1`). This variable was created in Stata using the following code (note that `rov` represents the state in which the survey took place):

```
gen stra1=rov==1 & urb_rur==1
gen stra2=rov==1 & urb_rur==2
gen stra3=rov==2 & urb_rur==1
gen stra4=rov==2 & urb_rur==2
gen stra5=rov==3 & urb_rur==1
gen stra6=rov==3 & urb_rur==2
```

`gen strata1 = stra1 + 2*stra2 + 3*stra3 + 4*stra4 + 5*stra5 + 6*stra6`*Applying weights in analyses:*

- Stata users: The complex-sample settings (`wgt` and `strata1`) have already been set and saved in the data file using the `svyset` command. For reference, the command to set these settings was:

```
svyset clnum [weight=wgt], strata(strata1)
```

You can use the dialog boxes to generate your analysis commands (click Statistics > Survey data analysis), or by simply adding the `svy:` prefix before your analysis commands. For example:

```
svy: tab urbrural
```

- SPSS users: To obtain accurate standard errors, the Complex Samples module is required, and you must also use the “HIV and Fertility Complex Sample Plan Community Women.csaplan” file, which specifies the weight and stratum variables. It’s easiest to use the dialog boxes to generate your desired tables, etc. (click Analyze > Complex Samples > ...). However, here is some sample syntax:

```
CSTABULATE  
/PLAN FILE= 'HIV and Fertility Complex Sample Plan Community women.csaplan'  
/TABLES VARIABLES=urbrural  
/CELLS POPSIZE TABLEPCT  
/STATISTICS SE CV CIN(95) COUNT DEFF DEFFSQRT  
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

See the extended methodology section below for more information about how the universe of clinics was stratified. Labels identifying strata were suppressed in order to deidentify the data.

Other key variables

The unique questionnaire identifier is `fqn`.

The household identifier is `hnum`.

Key demographic variables include `f104` (age at last birthday), `urbrural` (urban vs. rural location), `f111` (highest level of education completed), `f119` (religion), `f503` (current marital status), `f410a-ot` (last method used), and `f801` (pregnancy intendedness).

Missing values

In Stata, numeric variables have extended missing values (i.e. `.a`, `.b`, `c.`, etc.) with value labels and string variables have been assigned a value of `""`.

In SPSS, missing values are assigned a value of 9999, and are marked as missing in their value labels. For string variables, missing values are assigned a value of “Missing.”

Values of “N/A” were only declared missing if their value labels contained the word “missing” (i.e. “NA/Missing”). No values of “Don’t Know” were declared missing. “N/A” values are typically assigned numerical values of 8, 88, 888, or 8888, while “Don’t Know” values were generally assigned values of 98, 998, or 9998.

Programming for recodes

Some variables have been recoded in order to facilitate analysis. Below is a subset of the SPSS syntax that was used to create recoded variables.

Most of the original, non-recoded variables begin with an “f” followed by three numbers. Recoded variables can be found at the bottom of the codebook document, starting with the variable `filcomp`. The names of the variables that were used to create the recoded variables are included in brackets in the variable labels of those variables.

Some variables have been removed from the dataset because the information that they contain is included in other variables. These variables are marked as “removed” in the variable labels of any recoded variables that were created from them. Other variables were temporary and so were never included in the permanent data file. These have been marked as “temp” in the variable labels of recoded variables that rely on them.

When multiple original variables are used to create a single recoded variable (for instance, when variables for each contraceptive method type, `f405a-f405ot`, are coded into a single variable for contraceptive use, `cumeth`), values of “N/A” and missing values in original variables may be treated as valid “No” responses. Continuing with the example of contraceptive use, one respondent may have a value of “Not using” for `cumeth` because she has valid “No” responses recorded for each `f405` variable, but another respondent could also have a “Not using” value for `cumeth` and missing values for each `f405` variable.

Please note that some original variables in the dataset contain invalid values and discrepancies, as they were not used in previously published analyses. **Whenever available, recoded variables should be used instead of original variables.** There are two cases when it may be necessary to use original variables: 1) when variables were not related to analyses, and so were never cleaned and recoded in separate variables, and 2) when recoded variables differ from their corresponding original variables in the information that they contain and are not ideal for answering a specific research question. When using original variables for either of these reasons, it is the responsibility of the user to cross-check and clean those variables.

Recoding Syntax

```
*****.
*Age
*****.
*create new age variable based on age calculation using CMCs.
*note: cases with calculate age of 17+ coded as 18.

*if difference between calculated and reported age is less than or equal 2
and month and year provided for DOB, use calculated.
compute agec=$sysmis.
if difagecalc le 2 and missing(f103moimp) agec=dif103_104yrs.
```

```
*if difference between calculated and reported age is less than or equal 2
and month was imputed for DOB, use reported.
if difagecalc le 2 and f103moimp=6 agec=f104.
```

```
*difference between calculated and reported is greater than 2, use
calculated.
if difagecalc gt 2 agec=dif103_104yrs.
```

```
*if only calculated age available, set age to calculated.
if sysmis(agec) and dif103_104yrs gt 0 agec=dif103_104yrs.
```

```
*if only reported age available(ie interview or birthdate missing), use
reported.
if sysmis(agec) and ((f103yr=9999 or f103yr=9998) or missing(year)) and (f104
lt 90) agec=f104.
```

```
*code those 17+ as 18 years old.
recode agec (17=18).
*EXCLUDE AGE 16+ AND YOUNGER.
IF AGE LT 17 COMPLETED=-1.
IF AGE LT 17 filcomp=0.
```

```
*Community zambia women - 1 case 56 years old ineligible.
if fqn=135 and survey='ZCW' completed=-1.
```

```
*****
```

```
*Interview date CMC.
```

```
*****
```

```
Compute intvcmc=(year-1900)* 12 + month.
recode intvcmc (sysmis=9999).
missing values intvcmc (9999).
value labels intvcmc 9999 'missing'.
```

```
*****.
```

```
*Century month for respondent's date of birth--from facility recode.
```

```
*****.
```

```
*community only has f103mo and f103yr.
do if f103mo lt 13 and f103yr lt 9000.
compute dobcmc=(f103yr-1900)*12 + f103mo.
end if.
```

```
*impute month=6 for missing month.
if missing(f103mo) or f103mo=98 or value(f103mo)=99 f103moimp=6.
```

```
*Impute year for missing year.
*year of birth= interview year -age.
if missing(f103yr) or f103yr=9998 or f103yr=9999 f103yrimp=year-f104.
```

```
*dob cmc for cases missing month only.
do if f103yr lt 9000 and (f103mo=98 or value(f103mo)=99) .
```

```
compute dobcmcflg=1.
compute dobcmc=(f103yr-1900)*12 + f103moimp.
end if.
```

```
*dob cmc for cases missing year only.
do if (missing(f103yr) or f103yr=9998) and f103mo lt 98.
compute dobcmcflg=2.
compute dobcmc=(f103yrimp-1900)*12 + f103mo.
end if.
```

```
*dob cmc for cases missing month and year.
do if value(f103yr) gt 9000 and value(f103mo) gt 90.
compute dobcmcflg=3.
compute dobcmc=(f103yrimp-1900)*12 + f103moimp.
end if.
```

*Imputed month and year

```
*impute month=6 for missing month.
if missing(f103mo) or f103mo=98 or value(f103mo)=99 f103moimp=6.
```

```
*Impute year for missing year.
*year of birth= interview year -age.
if missing(f103yr) or f103yr=9998 or f103yr=9999 f103yrimp=year-f104.
```

*Education.

*highest level of education.

```
compute hieduc=f110.
if f109=2 hieduc=0.
if f109=1 and missing(f110) hieduc=9.
value labels hieduc 0 'none' 1 'primary' 2 'secondary' 3 'higher' 9
'missing'.
missing values hieduc (9).
```

*Currently in union.

```
recode f501 (1,2=1) (3=0) into inunion.
```

```
*zambia community women #1095- answers qs f504 on about current husband, but
invalid response on f501 and blank f502- code as inunion.
if survey='ZCW' and fqn=1095 inunion=1.
```

```
*missing on f501.
if sysmis(inunion) and missing(f501) inunion=9.
```

```
*didn't answer section 5.
if sysmis(inunion) and anssect5=0 inunion=9.
```

```
value labels inunion 1 'yes' 0 'no' 9 'missing'.
missing values inunion (9).
```

```
*****
```

```
*Ever married.
```

```
*****
```

```
if inunion=1 evermar=1.
if f502 le 2 evermar=1.
if f502=3 evermar=0.
```

```
*didn't answer section 9.
if sysmis(evermar) and anssect5=0 evermar=9.
```

```
*nigeria community women.
if survey='NCW' and (fqn=1045 or fqn=3490) evermar=9.
```

```
value labels evermar 0 'no' 1 'yes' 9 'missing'.
missing values evermar(9).
```

```
*****.
```

```
*In union/nvr married/formerly married/.
```

```
*****.
```

```
if inunion=1 inunion2=1.
if f502 le 2 inunion2=2.
if f502=3 inunion2=3.
if sysmis(inunion2) and (value(inunion)=9 and value(evermar)=9) inunion2=9.
```

```
*didn't answer section 9.
if sysmis(inunion2) and anssect5=0 inunion2=9.
```

```
*missing current marital status.
if sysmis(inunion2) and value(inunion)=9 inunion2=9.
```

```
value labels inunion2 1 'in union' 2 'formerly in union' 3 'never' 9
'missing'.
missing values inunion2 (9).
```

```
*****.
```

```
*Partner's education.
```

```
*****.
```

```
*highest level of education.
```

```
compute hieduc_p=f508.
```

```
if f508=4 hieduc_p=0.
```

```
if f508=9 and f509 ge 88 hieduc_p=99.
```

```
value labels hieduc_p 0 'none' 1 'primary' 2 'secondary' 3 'higher' 99
'missing' 88 'not in union'.
```

missing values hieduc_p (88,99).

```
*****.
*Number of unions.
*****.
***6 cases reply married more than once on f519, but only report 1 partner on
f521.
*4 cases are blank on f519 and f521.
compute uniontot=$sysmis.
if f519=1 uniontot=1.
if (f519=2 or sysmis(f519) or value(f519)=99) and (f521 lt 88) uniontot=f521.
if evermar=0 uniontot=0.
value labels uniontot 0 'never married' 99 'missing' 88 'NA'.
missing values uniontot (99,88).
if missing(uniontot) and inunion=1 and f519=1 uniontot=1.

*****.
*CMC of first marriage.
*****.
compute agfmcmc=$sysmis.

*currently married and married only once-- use f517 date and birth CMC.
do if (inunion=1 and uniontot=1) or( (inunion=0 and uniontot=1) or (f520=601
and value(uniontot)=99)).

*century month of first sex.
*have month and year.
if (f517mo le 12) and (f517yr lt 8888) agfmcmc=(f517yr-1900)*12 +f517mo.

*missing month- impute as june.
if (value(f517mo)=98 or sysmis(f517mo) or value(f517mo)=99 ) and (f517yr lt
8888) agfmcmc=(f517yr-1900)*12 + 6.
end if.

*undetermined.
if evermar=1 and uniontot=1 and (value(f517yr) gt 8000) agfmcmc=9998.

*married more than once, but blank on f526.
if uniontot gt 1 and (sysmis(f526yr) or value(f526yr) gt 9000) agfmcmc=9998.

*number of unions unknown and blank on f526.
if value(uniontot)=99 and (sysmis(f526yr) or value(f526yr) gt 9000)
agfmcmc=9998.

*never married.
if evermar=0 agfmcmc=8888.

*married more than once use f526.
```



```
if uniontot gt 1 and value(uniontot) lt 80 and ( (f526yr lt 8888) and (f526mo
lt 88)) agfmcml=(f526yr-1900)*12 +f526mo.
```

```
value labels agfmcml 9998 'not determined' 8888 'never married'.
missing values agfmcml(8888,9998).
```

```
*not currently married, previously married more than once use f526 date and
birth CMC.
```

```
*century month of 1st sex,
```

```
*have month and year.
```

```
do if (inunion2=2 or inunion2=1) and (uniontot gt 1 or f519=2).
```

```
if (f526mo le 12) and (f526yr lt 8888) agfmcml=(f526yr-1900)*12 +f526mo.
```

```
*missing month- impute as june.
```

```
if (value(f526mo)=98 or missing(f526mo)) and (f526yr lt 8888)
```

```
agfmcml=(f526yr-1900)*12 + 6.
```

```
end if.
```

```
execute.
```

```
*****.
```

```
*calculate age at first union.
```

```
*****.
```

```
compute agefmar=$sysmis.
```

```
do if agfmcml lt 9998.
```

```
compute agefmar=trunc((agfmcml- dobcml)/12).
```

```
end if.
```

```
*for cases missing date, but provided age.
```

```
*married more than once.
```

```
if missing(agefmar) and (uniontot gt 1) and (f527 lt 88) agefmar=f527.
```

```
*married only once.
```

```
if missing(agefmar) and (uniontot=1) and (f518 lt 88) agefmar=f518.
```

```
*impute age if have year.
```

```
if missing(agefmar) and (missing(uniontot) and f520=601) and (f518 lt 88)
```

```
agefmar=f518.
```

```
*undetermined.
```

```
if evermar=1 and sysmis(agefmar) agefmar=77.
```

```
*never married.
```

```
if evermar=0 agefmar=88.
```

```
*missing- missing on marital q's.
```

```
if survey='NCM' and any(fqn,57,82,101,1383,3433,7033) agefmar=99.
```

```
value labels agefmar 77 'undetermined' 88 'never married'.
```

```
missing values agefmar (77,88).
```

*****.

*Ever had sex.

*****.

compute eversex=0.

if f602 gt 0 eversex=1.

*IF NO INFO ON FIRST SEX , BUT EVER MARRIED CODE EVERSEX AS 1 added 7/6 2011.

IF EVERSEX=0 AND EVERMAR=1 EVERSEX=1.

*code cases that answered questions about sexuality in the past 12 months as sexually experienced.

if (missing(f603day) or value(f603day) gt 800 or f603day=300) and f616 le 2 eversex=1.

*Community Zam women- one case where f602 is blank, but responds to other qs ab sexual activity.

if sysmis(f602) and (value(f603day) lt 800 or f604 lt 80) eversex=1.

value labels eversex 1 'yes' 0 'no'.

*****.

*Age at first sex.

*****.

compute agefsx=f602.

recode agefsx (0=88).

*first sex, at 1st cohab.

if agefsx=95 agefsx=agefmar.

*1st sex at first cohab, but age at 1st cohab not determined.

recode agefsx (77,sysmis=99).

if agefsx=95 and value(agefmar)=77 agefsx=99.

value labels agefsx 88 'never had sex' 99 'missing'.

missing values agefsx (99,88).

*****.

*Sex in past X months.

*****.

compute sex3mo=0.

if (m1stsex ge 1 and m1stsex le 3) sex3mo=1.

*if inunion assume active in past 3 mos.

if inunion=1 sex3mo=1.

value labels sex3mo 1 "yes" 0 "Not sexually active" .

```

*past 12 mos.
compute sex12mo=$sysmis.
if m1stsex ge 1 and m1stsex le 12 sex12mo=1.
if m1stsex=0 or m1stsex=13 sex12mo =0.

*missing actual months, but in union, code as active.
if sysmis(m1stsex) or value(m1stsex)=99 and inunion=1 sex12mo=1.

if m1stsex=14 sex12mo=1.

value labels sex12mo 1 "yes" 0 "Not sexually active" 98 'don"t know'.

*****.
*Parity
*****.
compute parity=f208.

*missing no of children on f208, use calculated.
if sysmis(parity) and sysmis(f208) and (sumbirs gt 0 and sumbirs lt 70)
parity=sumbirs.

*parity not known, yes to having children but no counts on f203-f208,
provides date of first birth.
if sysmis(parity) and (f212yr lt 7000 or f210 lt 70) parity=77.

if value(sumbirs)=77 parity=77.

*cases that did not answer section 2.
if f201_218cnt=0 parity=99.

value labels parity 99 'missing' 77 '# not determined'.
missing values parity (99,77).

*****
*Age at last birth
*****
*compute age at last birth by subtracting respondent dob cmc from cmc of last
birth.
compute aglstbir=trunc((lsbircmc-dobcmc)/12).

*no births.
if hadchild=0 aglstbir=88.

*missing.
if sysmis(aglstbir) and value(lsbircmc) gt 9000 aglstbir=99.

*correction for cases with parity=1 where age at first birth and last birth
are different.
if parity=1 and (dif_agelst_agelast=-1 or dif_agelst_agelast=-2)
aglstbir=agefbir.

```

```
value label aglstbir 88 'never had a birth' 99 'missing'.
missing values aglstbir (88,99).
```

```
*****.
*Currently Pregnant.
*****.
compute curpreg=f301.
recode curpreg (2,8=0).
*code those that didn't answer the section as missing.
if sysmis(curpreg) and (sysmis(f301) or value(f301)=9 ) curpreg=9.

value labels curpreg 1 'yes' 0 'no' 9 'missing'.
missing value curpreg (9).
```

```
*****.
*Current Method
*****.
compute cumeth=0.
*fster.
if f405a=1 cumeth=1.
*mster.
if f405b=1 and cumeth=0 cumeth=2.
*pill.
if f405c=1 and cumeth=0 cumeth=3.
*iud.
if f405d=1 and cumeth=0 cumeth=4.
*injectables.
if f405e=1 and cumeth=0 cumeth=5.
*implants.
if f405f=1 and cumeth=0 cumeth=6.
*condom.
if f405g=1 and cumeth=0 cumeth=7.
*fcondom.
if f405h=1 and cumeth=0 cumeth=8.
*diaphragm.
if f405i=1 and cumeth=0 cumeth=9.
*foam/jelly.
if f405j=1 and cumeth=0 cumeth=10.
*LAC.
if f405k=1 and cumeth=0 cumeth=11.
*periodic abstinence.
if f405l=1 and cumeth=0 cumeth=12.

*rhythm(periodic abstinence).
if (survey='ZFM') and fqn=27 cumeth=12.
if survey='NCM' and fqn=1255 cumeth=12.
if survey='NCM' and (fqn=7065 or fqn=7013) cumeth=12.
```

```
*withdrawal.
if f405m=1 and cumeth=0 cumeth=13.
*other.
*includes emergency cp.
if f405x=1 and cumeth=0 cumeth=14.
if survey='NCM' and fqn=3468 cumeth=14.
*not using.
if cumeth=0 and (f404=2 or f402=408 or f404=8 or f405cky=0) cumeth=77.
*currently preg.
if curpreg=1 cumeth=66.
*blank.
if (f404=1 or missing(f404)) and cumeth=0 cumeth=99.
```

```
variable label cumeth 'Current method use- most effective[f405]'.
value label cumeth 1 'fem sterilizaiton' 2 'male sterilzation' 3 'pill' 4
'iud' 5 'injectable' 6 'implant'
7 'male condom' 8 'fem condom' 9 'diaphragm' 10 'foam/jelly' 11 'LAM' 12
'Periodic abst' 13 'withdrawal' 14 'other' 66 'pregnant' 77 'not using'
88 'NA' 99 'blank'.
```

```
*2nd recode incorporating 'other specify'.
compute cumeth1=cumeth.
```

```
*traditional herbs.
if (survey='ZCW' and cumeth1=14) and any(fqn,764,795,914,935,4485)
cumeth1=15.
execute.
```

```
*other medications.
if (survey='ZCW' and cumeth1=14) and any(fqn, 5,112,119,172,450,455)
cumeth1=16.
```

```
*emergency cp.
if (survey='ZCW' and cumeth1=14) and fqn=436 cumeth1=17.
```

```
*pregnant.
if survey = "ZCM" and fqn=945 cumeth1=66.
```

```
value labels cumeth1 1 'fem sterilizaiton' 2 'male sterilzation' 3 'pill' 4
'iud' 5 'injectable' 6 'implant'
7 'male condom' 8 'fem condom' 9 'diaphragm' 10 'foam/jelly' 11 'LAM' 12
'Periodic abst' 13 'withdrawal' 14 'other unspecified' 15 'trad herbs' 16
'other medications' 17 'emergency cp' 66 'pregnant' 77 'not using'
88 'NA' 99 'blank'.
```

```
*****.
*Currently using traditional.
*****.
```

```
recode cumeth ( 11 thru 14=1) (else=0) into cutrad.
```

```
*exclude emergency cp.
if cumeth1=17 cutrad=0.
```

```
*traditional from other specify.
if cumeth1=15 cutrad=1.
```

```
*code cafemol as traditional.
if survey='ZCW' and any(fqn, 5,112,172) cutrad=1.
```

```
value labels cutrad 1 'yes' 0 'no'.
freq vars=cutrad.
```

```
*****.
```

```
*Current METHOD TYPE
```

```
*****.
```

```
*modern only.
if f405mod gt 0 and f405trad=0 cutype=1.
```

```
*traditional only.
if f405mod =0 and f405trad gt 0 cutype=2.
```

```
*rhythm.
if f405mod =0 and cumeth=12 cutype=2.
```

```
*both.
if f405mod gt 0 and f405trad gt 0 cutype=3.
```

```
*other- includes other drugs and emergency cp.
if cumeth1=16 or cumeth1=17 cutype=4.
```

```
if cumeth1=14 cutype=4.
```

```
*Zambia community women.
if survey='ZCW' and fqn=119 cutype=3.
```

```
*none.
if curuse=0 cutype=5.
```

```
value labels cutype 1 'modern' 2 'trad' 3 'both' 4 'other' 5 'not using'.
```

```
*****.
```

```
*Contraception Source
```

```
*****.
```

```
*based on crosstab of source for current users and former users.
```

```
*no cases provide source on cpsorcnu and cpsorcnu because they are mutually exclusive.
```

```
* to avoid multiple if statements- commands below will compute the source variable by taking the valid values 1,2,3 as compared to invalid and missing.
```

```
compute cpsource=min(cpsorcu,cpsorcnu).
recode cpsource (9=8).
if value(cpsorcu) ge 8 and value(cpsorcnu) ge 8 and missing(cpsource)
cpsource=8.
freq vars=CPSORCu cpsorcnu cpsource.

value label cpsource 1 'public' 2 'private' 3 'other' 8 'NA/missing'.
missing values cpsource (8).
```

```
*****.
*Preferred method for future use.
*****.
compute metfudur=0.

*fster.
if f414a=1 metfudur=1.

*mster.
if f414b=1 and metfudur=0 metfudur=2.

*pill.
if f414c=1 and metfudur=0 metfudur=3.

*iud.
if f414d=1 and metfudur=0 metfudur=4.

*injectables.
if f414e=1 and metfudur=0 metfudur=5.

*implants.
if f414f=1 and metfudur=0 metfudur=6.

*condom.
if f414g=1 and metfudur=0 metfudur=7.

*fcondom.
if f414h=1 and metfudur=0 metfudur=8.

*diaphragm.
if f414i=1 and metfudur=0 metfudur=9.

*foam/jelly.
if f414j=1 and metfudur=0 metfudur=10.

*LAC.
if f414k=1 and metfudur=0 metfudur=11.

*periodic abstinence.
if f414l=1 and metfudur=0 metfudur=12.
```

```

*withdrawal.
if f414m=1 and metfutur=0 metfutur=13.

*traditional herbs from other specify.
if survey='ZCW' and any(fqn, 4721,873,912) metfutur=14.
sort cases by fqn.

*don't know - from other specify.
if metfutur=0 and survey='ZCW' and ANY(fqn,311,14,70,1084,1131) metfutur=16.
if survey='ZCW' and fqn=14 metfutur=16.

*all other.
if metfutur=0 and survey='ZCW' and f414x=1 metfutur=15.

*won't use/NA-currently using.
if metfutur=0 and (f413=2 or f413=8 or f414cky=0) metfutur=88.

value labels metfutur 1 'fem sterilization' 2 'male sterilization' 3 'pill' 4
'iud' 5 'injectable' 6 'implant'
7 'male condom' 8 'fem condom' 9 'diaphragm' 10 'foam/jelly' 11 'LAM' 12
'Periodic abst' 13 'withdrawal' 14 'traditional herbs' 15 'other'
16 'don't know/any meth' 88 'NA/missing' 99 'blank'.

*****.
*Reason for not using in the future.
*****.
*code reason for future nonuse into numeric var.
recode f415 (11 thru 16=1) (21 thru 26=2) ( 31 thru 33=3) into reasnfut.

*other specify.
if survey='ZCW' and any(fqn,568,1294) reasnfut=4.

*zcw can become infertile- code as method related.
if survey='ZCW' and fqn=709 reasnfut=1.

*zcw my age is gone-code as fertility related.
if survey='ZCW' and (fqn=458 or fqn=1053) reasnfut=2.

*NA.
if f413=1 or f413=8 reasnfut=8.
If (f413=9 or f413=2) and missing(reasnfut) reasnfut=9.

value labels reasnfut 1 'method-related' 2 'fertility' 3 'opposition' 4
'other' 8 'NA' 9 'missing'.
missing values reasnfut (8,9).

* Zambia community women- method related- side effects, can become infertile,
can become infertile.
if survey = "ZCW" and any(fqn, 4554, 4575,709,1053) reasnfut=1.

```


* Zambia community women- fertility- not married, stays away for sex, don't have partner, have no partner.

if survey = "ZCW" and any(fqn, 133, 1086, 702, 903) reasnfut=2.

* Zambia community women- fertility- menopausal.

if survey = "ZCW" and any(fqn, 606) reasnfut=2.

* Zambia community women- fertility- can't get pregnant, cannot conceive, takes years to conceive,

* difficult to conceive, thinks she can't get pregnant, she is closed, my age is gone.

if survey = "ZCW" and any(fqn, 97, 923, 925, 993, 1205, 4487, 4561, 1053) reasnfut=2.

* Zambia community women- opposition- don't tempting god.

if survey = "ZCW" and any(fqn, 1088) reasnfut=3.

* Zambia community women- opposition- wants to be pregnant, wants to get pregnant.

if survey = "ZCW" and any(fqn, 4518, 733) reasnfut=3.

* Zambia community women- opposition- don't want, don't want to, not interested.

if survey = "ZCW" and any(fqn, 81, 71, 324) reasnfut=3.

* Zambia community women- opposition- not used.

if survey = "ZCW" and any(fqn, 89) reasnfut=3.

* Zambia community women- other-, contaminated with a disease, grown up,

* I have grown, no feelings, not ready, nothing much, she just feel safe,

* unsure.

if survey = "ZCW" and any(fqn, 1068, 571, 918, 292, 312, 1075, 1077, 1222, 4551) reasnfut=4.

*****.

*Fertility preferences.

*****.

*create ferpref type variable for respondent.

*want no more.

if f701=2 ferpref=5.

*undecided.

if f701=8 ferpref=4.

*want in less than 2 yrs.

if (f701=1 or value(f701)=9) and (f703=993 or f703 le 202) ferpref=1.

if (missing(f701) or value(f701)=9) and f703=993 ferpref=1.

*want in 2+ yrs.

if f701=1 and (f703 gt 202 and f703 lt 900) ferpref=2.

*wants unsure of timing- includes after marriage, when health improves (when feel better, when cd4 count rises) other and don't know/blank on f703.
if f701=1 and (f703=995 or f703=996 or f703=998 or value(f703)=999)
ferpref=3.

*reported infertile?.
if f701=3 ferpref=6.

*missing.
if sysmis(ferpref) and ((value(f703) gt 800 or sysmis(f703)) and
(value(f701) gt 8 or value(f701)=1) and f702cnt=1) ferpref=9.

*missing on f701 and f703.
if sysmis(ferpref) and missing(f701) and missing(f703) ferpref=9.
if sysmis(ferpref) and value(f701)=9 ferpref=9.

value labels ferpref 1 'want in 2 yrs' 2 ' want after 2 yrs' 3 'wants unsure timing' 4 'undecided' 5 'want no more' 6 'infertile' 9 'missing'.
missing values ferpref (9).

*****.

*Fertility preference of partner.

*****.

*create ferpref type variable for partner of respondent.

*want no more.

if f709=2 ferpref_p=5.

*undecided.

if f709=8 ferpref_p=4.

*want in less than 2 yrs.

if f709=1 and (f710=993 or f710 lt 202) ferpref_p=1.

*want in 2+ yrs.

if f709=1 and (f710 ge 202 and f710 lt 900) ferpref_p=2.

*wants unsure of timing- includes after marriage, other and don't know on f710.

if f709=1 and (f710=995 or f710=996 or f710=998) ferpref_p=3.

*reported infertile?.

if f709=3 ferpref_p=6.

*not in union.

if inunion=0 ferpref_p=8.

if value(f709)=88 and value(f710)=888 ferpref_p=8.

*missing.

if sysmis(ferpref_p) and inunion=1 and (missing(f709) and missing(f710))
ferpref_p=9.

if sysmis(ferpref_p) and inunion=1 and missing(f710) ferpref_p=9.

value labels ferpref_p 1 'want in <2 yrs' 2 ' want after 2 yrs' 3 'wants
unsure timing' 4 'undecided' 5 'want no more' 6 'infertile' 8 'NA' 9
'missing'.

CMC date of last abortion

*****.

*have month and year.

if (f809yr lt 9000 and f809mo lt 13) abrtcmc=(f809yr-1900)* 12 + f809mo.

*if only have year, impute month=june.

if (f809yr lt 9000 and value(f809mo)=98) abrtcmc=(f809yr-1900)* 12 + 6.

*missing date, but have age, impute.

if sysmis(abrtcmc) and f809a lt 80 abrtcmc=(dobcmc+ (f809a*12)) +6.

*code missing.

if sysmis(abrtcmc) and abort=1 abrtcmc=9999.

*code NA.

if sysmis(abrtcmc) and abort=0 abrtcmc=8888.

formats abrtcmc (f4.0).

value label abrtcmc 9999 'missing' 8888 'NA'.

missing values abrtcmc (8888,9999) .

age last abortion.

*****.

if f809a lt 80 abrtage=f809a.

if sysmis(abrtage) abrtage=trunc((abrtcmc- dobcmc)/12).

*missing.

if abort=1 and sysmis(abrtage) abrtage=99.

*na.

if abort=0 and sysmis(abrtage) abrtage=88.

value labels abrtage 88 'NA' 99 'missing'.

missing values abrtage (99,88).

*****.

*f816/830 who did you see for first method to terminate pregnancy.

*****.

*****these constructions have to be done separately, because first
step/method not asked on male surveys.

*new main variable.

compute f830r=f830.

```

APPLY DICTIONARY from *
/SOURCE VARIABLES = f830
/TARGET VARIABLES = f830r.

*recode.
recode f830r ( 1 thru 3=1) ( 4 thru 8,96=2) (99=3) into abrwhol .

*na--- last attempt.
if sysmis(abrwhol) and abort=0 abrwhol=8.

*missing- last attempt.
if sysmis(abrwhol) and abort=1 abrwhol=9.

value labels abrwhol 1 'health prof' 2 'other person ' 3 'no one' 8 'na' 9
'missing' .

*****.
*f817/831 where went to have 1st and last methods done.
*****.
*new main variable.
compute f831r=f831.

APPLY DICTIONARY from *
/SOURCE VARIABLES = f831
/TARGET VARIABLES = f831r.

*recode.
recode f831r ( 11 thru 15=1) ( 21 thru 25=2) (31=3) (32,33=4) into abrwhr1 .

value labels abrwhr1 1 'public' 2 'private ' 3 'respondent home' 4 'other
home' 8 'na' 9 'missing' .

*Missing.
*na--- last attempt.
if sysmis(abrwhr1) and abort=0 abrwhr1=8.
*missing- last attempt.
if sysmis(abrwhr1) and abort=1 abrwhr1=9.

*****
CP method used when became pregnant.
*****.
*count modern users.
count f842mod=f842a to f842i (1).

*count traditional.
*includes traditional methods specified on other.
count f842trad=f842k to f842m(1).

```

```
*modern only.
if f842mod gt 0 and f842trad=0 abcpfail=1.
*traditional only.
if f842mod =0 and f842trad gt 0 abcpfail=2.
*both.
if f842mod gt 0 and f842trad gt 0 abcpfail=3.

**other- do freq of f842 other to determine if can be coded into existing
response categories.
If sysmis(abcpfail) and f842x=1 abcpfail=4.

*none.
if f842z=1 abcpfail=5.

**no abortions.
if abort=0 abcpfail=8.

*missing.
if sysmis(abcpfail) and abort=1 abcpfail=9.

value labels abcpfail 1 'modern' 2 'trad' 3 'both' 4 'other' 5 'didn't use'
8 'na' 9 'missing'.
missing values abcpfail (8,9).
```

```
Know cytotec.
*****.
recode f845 (1=1) (2,8=0) into cytoknow.
if value(f845)=99 cytoknow=9.
value labels cytoknow 1 'yes' 0 'no' 9 'missing'.
missing values cytoknow (9).
```

```
f846 know source for cytotec - recode?.
*****.
*zambia community women-on other response(f846ot) several respondents coded
as don't know source.
*create a don't know category on recode var.
```

*Note: fieldworker coded as other public.

```
*other-- includes source shop, church friend/relative and other specify..
count cytother=f846o to f846x (1).
recode cytother (1 thru hi=1).
```

```
*exclude 4 cases for Zambia community women that responded don't know( coded
below in separate recode var) and didn't reply yes
on any othe the 'other vars-o,p,q.
if survey='ZCW' AND (fqn=1251 or fqn=39 or fqn=265 or fqn=209) cytother=0.
```

```
if cytoknow=0 cytother=8.
if missing(cytoknow) and cytother=0 cytother=8.
value labels cytother 1 'yes' 0 'no' 8 'na/missing'.
```

Section 9 HIV

*****.

*****.

*Know HIV.

*****.

*Must look at cases that are blank on f901/f902- determine whether they know hiv by looking for any other questions answered in the section .

*select cases blank on f901/f902 in ORIGINAL DATA, save data to excel and transpose so variables appear in rows.

*Know HIV based on f901/f902 and responses to other sect 9 vars.

compute knohiv=0.

if f901=1 or f902=1 knohiv=1.

value labels knohiv 1 'yes' 0 'no'.

*****.

*CP methods received counseling on after last HIV test.

*****.

*f959- use coding for f405(CUTYPE)- response categories are the same.

*count modern .

count f959mod=f959a to f959i (1).

*count traditional.

count f959trad=f959k to f959m f959x(1).

*zambia community women- using cafemol.

if survey='ZCW' and fqn=455 f959trad=1.

*modern only.

if f959mod gt 0 and f959trad=0 tscpmetc=1.

*traditional only.

if f959mod =0 and f959trad gt 0 tscpmetc=2.

*both.

if f959mod gt 0 and f959trad gt 0 tscpmetc=3.

*none.

if f957=2 tscpmetc=4.

*na.

if value(f957)=88 tscpmetc=8.

*missing.

if (f957=1 or f957=8 or value(f957)=99) and sysmis(tscpmetc) tscpmetc=9.

```
value labels tscpmetc 1 'modern' 2 'trad' 3 'both' 4 'not counseled' 9
'missing' 8 'na'.
missing values tscpmetc (8,9).
```

```
*****.
*Recode 2 - Most effective.
*****.
compute tscpeff=0.

*fster.
if f961a=1 tscpeff=1.

*mster.
if f961b=1 and tscpeff=0 tscpeff=2.

*pill.
if f961c=1 and tscpeff=0 tscpeff=3.

*iud.
if f961d=1 and tscpeff=0 tscpeff=4.

*injectables.
if f961e=1 and tscpeff=0 tscpeff=5.

*implants.
if f961f=1 and tscpeff=0 tscpeff=6.

*condom.
if f961g=1 and tscpeff=0 tscpeff=7.

*fcondom.
if f961h=1 and tscpeff=0 tscpeff=8.

*diaphragm.
if f961i=1 and tscpeff=0 tscpeff=9.

*foam/jelly.
if f961j=1 and tscpeff=0 tscpeff=10.

*LAC.
if f961k=1 and tscpeff=0 tscpeff=11.

*periodic abstinence.
if f961l=1 and tscpeff=0 tscpeff=12.

*withdrawal.
if f961m=1 and tscpeff=0 tscpeff=13.

*other.
if f961x=1 and tscpeff=0 tscpeff=14.
```

```

*none.
if tscpeff=0 and f960=2 tscpeff=77.

*na- never tested.
if value(tscptype)=8 tscpeff=88.

*missing.
if value(tscptype)=9 tscpeff=99.

value labels tscpeff 1 'fem sterilizaiton' 2 'male sterilization' 3 'pill' 4
'iud' 5 'injectable' 6 'implant'
7 'male condom' 8 'fem condom' 9 'diaphragm' 10 'foam/jelly' 11 'LAM' 12
'Periodic abst' 13 'withdrawal' 14 'other' 77 'none'
88 'NA' 99 'missing'.
missing values tscpeff (88,99).

*****.
*Recode 2 - Dual method(Condom combination).
*****.
*condom alone.
*value on modern count should=1 for condom use only.
compute tscpdual=$sysmis.
if f961g=1 and (f961trad=0 and f961mod=1 ) tscpdual=3.

*condom + other modern.
*value on modern count must be 2 or more-- condom + another method.
if f961g=1 and (f961trad=0 and f961mod gt 1) tscpdual=1.
*if using modern and traditional methods in addition to condom, code as using
modern(1 case for Zambia facility).
if f961g=1 and (f961trad=1 and f961mod gt 1) tscpdual=1.

*condom + traditional.
*value on modern count must=1(for condom).
if f961g=1 and (f961trad gt 0 and f961mod=1) tscpdual=2.
*not using a method.
if tscpeff=77 tscpdual=4 .
*using but not condom.
if tscpeff lt 77 and sysmis(tscpdual) tscpdual=4.

*na/missing- never tested.
*na- never tested.
if value(tscptype)=8 tscpdual=8.
*missing.
if value(tscptype)=9 tscpdual=9.

value label tscpdual 1 'condom + modern ' 2 'condom + trad' 3 'condom only' 4
'not using condom' 8 'never tested' 9 'missing'.
missing values tscpdual (8,9).

```



```
*****.
*Respondent shared test results.
*****f975/ffm965*****.
compute tsdisclos=f975.
recode tsdisclos (2=0).
if value(f975)=99 tsdisclos=99.
if value(f975)=88 tsdisclos=88.

*if yes on disclosed but all no's on f976 then code disclosure as no.
value labels tsdisclos 0 'no' 1 'yes' 88 'na' 99 'missing' .
missing values tsdisclos (88,99).
formats tsdisclos f975 (f2.0).

*****.
*HIV STATUS- FOR COMMUNITY SURVEYS ONLY.
*****.
*BASED ON F971 ONLY.
recode f971 (1=1) (2=2) ( 3 thru 8, missing=3) into hivstat1.
value labels hivstat1 1 'positive' 2 'negative' 3 'unknown'.

*based on f971, sealed envelope and perceived risk.
if f971=1 or f1107=1 or f915=5 hivstat2=1.
if sysmis(hivstat2) hivstat2=hivstat1.
value labels hivstat2 1 'positive' 2 'negative' 3 'unknown'.
```

Extended methodology

Data collection (Sampling)

This survey was conducted in the Southern, Lusaka, and Northern provinces of Zambia. The provinces were selected based on the levels of HIV, fertility and contraceptive use reported in the 2007 Zambia Demographic and Health Survey (ZDHS). The goal of selecting the three provinces was to include urban and rural diversity and capture regional or ethnic variation, although they are not necessary to the goals of this study. Besides, the HIV epidemic in Zambia was more or less evenly spread in major population groupings.

A target sample of 2,600 females (aged 18–49 years) of reproductive age was selected. This figure represents the combined samples for men and women. In Zambia there were 1,363 male respondents and 1,441 female respondents, yielding a total of 2,804 responses, well over the target. The 2,600 sample size estimate was based on an overall proportion of 30% (the approximate proportion of women who want no more children or women who are currently using any method in the 2007 ZDHS and a detectable difference of 5% (the approximate difference on these two outcomes between those who, in answer to a survey question, perceive themselves at high risk of HIV and those who perceive themselves at lower risk). An alpha of 0.05 and statistical power of 80% yields an approximate sample size of 2,600.

This sample was sub-sampled from the sample for the 2007 ZDHS. This minimized costs and optimized the usage of survey amenities in Zambia. The 2007 ZDHS sample was designed to

provide estimates of population and health indicators at the national and provincial levels. The sample was designed to provide specific indicators including reproductive health indicators and HIV prevalence for each of the nine provinces in Zambia. The sampling frame of the 2007 ZDHS was the listing of enumeration areas (EAs) from the 2000 Census of Population and Housing. In the 2000 Census programme, the whole of Zambia was divided into EAs of about equal size. Therefore, all households in Zambia had a chance of being selected into the ZDHS sample excluding those omitted purposively from restricted areas such as military barracks. There were 16,757 EAs in the 2000 Census of Population and Housing out of which 320 were selected for the 2007 ZDHS with probability proportional to the size of the EA. However, permission to access one EA was not given.

The sample for the 2007 ZDHS was stratified by separating every province into urban and rural areas. Therefore, the nine provinces were stratified into 18 sampling strata. EAs were selected independently in every stratum, first at the Census Supervisory Area (CSA) level and then at the EA level. The Central Statistical Office demarcated all political administrative areas such as provinces into CSAs. CSAs were either in rural or urban areas. CSAs were then divided into EAs. Implicit stratifications and proportional allocation was achieved by sorting the sample frame according to the geographical/administrative order and by using a probability proportional to size. Households in the 319 EAs were listed. Out of the listed households, 25 on average were selected for each EA by equal probability systematic sampling. Out of these EAs, those in Southern, Northern and Lusaka provinces were sub-sampled for the HIV status and achieving fertility desires community survey.

The sample was selected indirectly by first selecting the households. The number of households to be selected was determined by the ratio of households to complete interviews of the population 15 years and older observed in the 2007 ZDHS. In order to yield the sample size of 2,600 women and men, the number of EAs to be selected if 25 households on average were selected from each EA by equal probability systematic sampling was 60, of which 38 were in rural areas and 22 in urban areas. This also took into account different response rates for women and men. In Lusaka Province, 17 EAs were allocated, out of which 5 were in rural areas and 12 in urban areas. In Southern province, a total of 18 were allocated: 14 in rural areas and 4 in urban areas. In Northern Province, 25 were allocated: 19 in rural areas and 6 in urban areas. The total sample of 2,600 was allocated to the provinces proportional to its projected population of 2009 distributed by urban and rural areas. The projected populations were obtained from the Central Statistical Office.

Data Collection (Survey)

Separate questionnaires were used to collect data for women, and men. The questionnaires included the core set of topics covered in demographic and health surveys: demographic and socioeconomic characteristics of the respondent, sexual behaviour, union status, contraceptive use, fertility preferences, fecundity, and a pregnancy history (including all pregnancies: live births, miscarriages and induced abortions). Questions were asked to determine the respondents' perception of risk for HIV infection and status of HIV, if they knew it. Those who knew that they had HIV were asked additional questions, including how they first found out about their status and when.

Additionally, a “secret ballot” was conducted where the interviewer asked the respondent two questions that could be answered yes or no: One question about experiencing an abortion (or, for men, if a partner has experienced an abortion) and another about having two or more sex partners in the past 12 months. Respondents were instructed to circle a picture that represents their answer (yes or no) and to fold the paper and put it in an envelope and seal it and place it inside a larger envelope for the interviewer to take. The interviewer turned his/her back while reading out each question until the respondent finished sealing the envelope so that the respondent's actions were not observed.

The questionnaires and consent forms were developed in English and translated into one of the major languages of the provinces in which the survey was carried out (Tonga, Nyanja or Bemba). Nyanja questionnaires were used in Lusaka and Southern provinces. Bemba questionnaires were used in Northern and Lusaka provinces. The questionnaires were translated into the three languages by hired translators. The interviewers reverse translated the questionnaires into English in order to check that the translation retained its meaning. When it did not, the interviewers made an appropriate correction. The reverse translations were done during the pilot survey and during the training for the main survey.

Fieldwork started in the second week of October and was completed in February 2010. Follow-up interviews were carried out in May and June in Southern Province and in August and September in Northern Province in order to improve the response rates. Response rates were initially lower than expected due to seasonality factors such as the rains between October and February.

Informed consent was obtained from all survey respondents prior to interview. The questionnaires were mainly administered in the main languages of the provinces. A few were administered in English. No tokens of appreciation or compensation were given as part of the survey.

Bibliography of articles based on the dataset

- Bankole, Akinrinola, A.E. Biddlecom, J.O. Akinyemi, K. Dzekedzeke, O. Awolude and I.F. Adewole. "Does knowledge about antiretroviral therapy and mother-to-child transmission affect the relationships between HIV status and fertility preferences and contraceptive use: New evidence from Nigeria and Zambia." *Journal of Biosocial Science*, 2013, 16:1-20.
- Bankole, Akinrinola, Sarah Keogh, Odunayo Akinyemi, Kumbutso Dzekedzeke, Olutosin Awolude and Isaac Adewole. "Differences in unintended pregnancy, contraceptive use and abortion by HIV status among women in Nigeria and Zambia." *International Perspectives on Sexual and Reproductive Health*, 2014, 40(1): 28-38.
- Kavanaugh, Megan L., Ann M. Moore, Odunayo Akinyemi, Isaac Adewole, Kumbutso Dzekedzeke, Olutosin Awolude and Oyedunni Arulogun. "Community attitudes towards childbearing and abortion among HIV-positive women in Nigeria and Zambia." *Culture, Health, and Sexuality*, Vol. 15, No. 2, February 2013, 160–174.
- Moore, Ann M., Akinrinola Bankole, Olutoin Awolude, Suzette Audam, Adesina Oladokun & Isaac Adewole. "Attitudes of women and men living with HIV and their healthcare providers towards pregnancy and abortion by HIV-positive women in Nigeria and Zambia," *African Journal of AIDS Research*, 14:1, 2015, 29-42.
- Moore, Ann M., Sarah Keogh, Megan Kavanaugh, Akinrinola Bankole, Chishimba Mulambia, and Namuunda Mutombo. "Bucking social norms: Examining anomalous fertility aspirations in the face of HIV in Lusaka, Zambia." *Social Science & Medicine*, 119, 2014, 88e97.