printcase: A Stata command for visualizing single observations

Max Weinreb, Austin, TX

Jenny Trinitapoli, Chicago, IL, jennytrini@uchicago.edu

**Abstract.** In this report, we introduce the printcasecommand for outputting data from a specific observation into an easy-to-read Microsoft Word or PDF document. printcaseallows analysts to focus on a single observation within a cross-sectional dataset and view that observation in its entirety. The output displays all fields associated in table format, with all variables identified by their corresponding labels and all responses their corresponding value labels. We 1) explain how *printcase* works, 2) give examples of circumstances under which this type of table-based quasi-questionnaire would be useful, and 3) provide Stata code for “printing” single observations.

**Keywords.** survey research, fieldwork, data quality, interviewer training, printcase

# Introduction

printcaseis a Stata command analysts can use to generate a table of variables and responses, specific to a uniquely identified observation from any .dta file.

There are times when examining a single survey from cross-sectional data is valuable to improve comprehension and generate new insights. But in the era of e-tablet modes of data collection, producing something that allows one to look at a single survey in a format resembling a paper questionnaire is surprisingly difficult. *printcase* addresses this need by providing researchers with an abbreviated quasi-questionnaire generated from responses for a single observation in a data file. When survey items (questions and value labels) are fully labeled, the printed case can proxy a completed survey, much like what we used in the days of pencil-and-paper questionnaires.

We can think of at least four reasons why researchers would want to skim or study responses from a particular, single observation.

First, examining individual surveys in their entirety is useful for data cleaning and making judgment calls about unusual values. One of the ways this is done is through examining other responses in the questionnaire to aid in data cleaning to make sure that the answers are inherently consistent. Researchers disagree about whether and how to go about editing data (Sana & Weinreb 2008), but some argue that leveraging information provided by respondents themselves is superior to even the most sophisticated approaches to imputation (Leahey 2008; Leahey, Entwistle, and Einaudi 2003; Waal, Pannekoek, and Schotus 2011). By looking at the complete answers, it can become clear how to recode an outlier. An example comes from our own data-collection effort in Balaka, Malawi, is that of a woman who said she had never had sex and was “not at all worried about HIV” reported that she had been tested six times in the past month. The value was unusual and seemed to be a mistake. However, upon closer examination of the questionnaire, we learned that this woman was part of a peer-to-peer counseling group, in which she would encourage friends to get tested, accompany them to the testing facility, and go through the entire process with them as part of a district-wide effort to increase voluntary testing. This shows us how reading a particular observation vertically can generate insights that are concealed when we only examine data using measures of central tendency. When the underlying dataset has been de-identified, researchers can produce a case for study and contemplation, and that case will also be fully anonymized.

Another reason a researcher would want to study all the answers from one observation is to look at numeric responses in conjunction with the interviewers’ notes or other open-ended responses, which may have been collected during interviews or after, to help analysts geographically and temporally separated from the interview understand other aspects of the moment or the interaction that informed the answers gathered. Reflecting on more than a decade of fieldwork and data analysis, Bledsoe et al. (1998) remark that their analytical efforts were enhanced when they were able to juxtapose “open-ended commentary as variables alongside… quantifiable responses.” Sometimes these open-ended responses provide information that can subsequently be coded up into close-ended responses or may in fact require changing a response value. Take, for example, a survey question that asked respondents how many hours of television they watched per day in the past week. The responses range between 0 and 11, with an average of 2.2. The 95th percentile is 5, and analysts are left with questions about how to manage the values of 6 and over. By examining the particular observation, the analyst may be able to identify an erroneous response (the “fat fingers” problem) in which the 11 should have been a 1 = “the 8 o’clock news every night,” from a true value of 11, e.g. “Since accident, respondent is bed-ridden & watches all day.” Given a well-labeled dataset, *printcase* provides similar functionality, in which analysts can leverage interviewer notes, write-in responses, and other qualitative descriptions as a complement to the quantitative data.

Third, although survey data is almost always examined in the aggregate, the responses of particular individuals are critical for longitudinal data collection or other data collection which engages with the same respondent multiple times as it allows the interviewer to review what the respondent shared previously before reengaging with a respondent. An example of this can be found in Pearce’s (2002) approach to Systematic Anomalous Case Analysis. In this method, analysts analyze aggregate data using traditional, regression-based methods, identifying patterns and selecting cases that deviate from the trends for in-depth follow-up research, especially in-depth interviews and ethnography. A careful read of the completed questionnaire is an essential step for preparing to conduct a valuable follow-up interview with the same individual. In the absence of a paper questionnaire to consult, *printcase* can be used to generate a file that sketches the earlier conversation between interviewer and respondent. This would serve as the basis upon which new questions for a follow-up conversation could be generated.

Finally, paper questionnaires are valuable for training interviewers and enumerators. Most studies still use a blank, paper version of their questionnaire for interviewer training, emphasizing the scripts that structure transitions between modules and the introductions that cue particular questions and clarify whether responses should be read. Paper versions are easier to browse and skim as a full document, rather than item-by item. This is important for teaching skip-patterns and familiarizing interviewers with the overarching goals of the particular study they are fielding.

*printcase* cannot replace the designed questionnaire, but it can be used to quickly produce a set of responses – actual or theoretical (i.e., from synthetic data) – that interviewers can study as part of interviewer training. In our experience, it is particularly valuable to have interviewers study a completed questionnaire collected while piloting the instrument; this exercise helps prepare interviewers for the kinds of responses they might encounter in the field, and it also helps train them to think about the internal consistency of a narrative during the interview.

Fieldwork supervisors, responsible for ensuring data quality, may also want to browse printed cases to check the quality of interviewers’ work and provide additional support and training where necessary. For example, if one interviewer is entering more “refused to answer” responses than others, they may need to introduce a particular topic with more sensitivity or learn how to probe more effectively. By browsing *printed cases* with a focus on the interviewer’s work, field supervisors can catch and remedy interviewer-specific errors before they are manifest too deeply in the entire data-collection enterprise.

# The printcase command

*printcase* is meant to be used with a dataset that has been adequately labeled with variable names and corresponding variable labels. The dataset must be organized by unique id (numeric, of any number format) assigned to each individual/case. The output of *printcase* is a table three columns wide, which displays: 1) variable name, 2) variable label, and 3) response value. The first row contains the column labels, and one row is generated for every variable in the dataset, unless exclusions are specified (see below).

printcase builds on the following stata commands– Max will add a paragraph giving lots of credit, as the editor requested.

### Syntax

printcase using “filename” if *id\_variable==`value’ [, options]*

where *id\_variable* is the name of the unique identifying variable in the dataset, and *value* is the value of *id\_variable* to print. The *id\_variable­* must uniquely identify variables in the dataset; id\_*variable* must be a numeric variable; it can cannot contain spaces. If no using file is specified, *printcase* saves the file to the home directory and names the file by concatenating the specified *id\_variable* and *value*. For example: *printcase if id==100* generates a file called “id100.docx" and saves it in the root directory.

### 2. 2 Options are the following:

noempty suppresses all empty and system missing responses and their variables from the resulting table, if the value label is an empty string (i.e., “”) or a Stata missing value

code (“.”, “.d”, etc.). If not specified, all empty responses will be included. In longitudinal cases, variable rows are only suppressed if variables are empty for all observations.

ignore(“string1” “string2”…) allows users to specify variables to be ignored based on the values, for example missing strings “” general missing “.” or skipped values “.s” if the dataset distinguishes. (Other datasets use “99” or other codes.) The result in the output document will not include those variables or their responses. There is no limit on the number of responses to ignore.

longitudinal indicates that the case id is non-unique and that nested values (time or observations) should be presented in columns.

unit(“label”) can be used with longitudinal to label the columns within a case.

replace overwrites an existing printed case.

addnotes includes the first note on any variable in the variable label column. (If the dataset is documented to such a degree, this may help the analyst discern skip patterns or heed cautions about variables with known problems.) Although multiple notes can be attached to a variable, only the first note is included in the printed case.

pdf sets the output of *printcase* to be a PDF file instead of the default Microsoft Word file (.docx). Page numbers and footers are not generated in PDF files, whereas they are in Microsoft Word. All other options are unaffected by specifying a *pdf* file as the output.

font(string) sets the font to be used for the entire document of the output of printcase. Any installed font can be specified. If not set, the default is Arial.

landscape changes the paper orientation from portrait (the default) to landscape.

width(value, value, value) allows analyst to control column width for the table. Values must sum to 100 and must correspond to the number of columns in the table – 3 columns for a default cross-sectional case (variable name, variable label, response). For longitudinal cases the number of columns is equal to 2+n, where n is the number of nested observations.

Examples of -printcase- use

In this example, the researcher first loads the model births recode dataset from the DHS (ICF 2020). The researcher would like to vertically read the answers provided by the 100th respondent using *printcase*, but the dataset’s identifier variable is a string variable. The user generates a unique, numeric ID variable called “id” by simply sequentially numbering each observation in the dataset and then makes that variable the first in the dataset.

cd "/Users/RedactedForPeerReview/Box/printcase/"

use "./printcase\_resubmission/ZZBR62FL.DTA", clear

\*create unique identifier for each case

gen id=\_n

\*make this the first variable

order id

Because this dataset (like many others) doesn’t contain any notes, we add two notes to demonstrate the value of the *includenotes* feature.

\*\*Adding some notes to the dataset to demonstrate the value of includenotes

notes v007: check if this is ethiopian, nepali, or gregorian calendar

notes v130: religious groups are ambiguous in the model dataset

In Example 1, the analyst calls *printcase*, using the common *if* arguments to specify the id variable and relying on the default filename. This user chose to analyze observation 100. The options specified here will generate a pdf output format (rather than MS Word document) using Ariel font.

\*\*EXAMPLE 1: print the 100th case as a pdf

printcase if id==100, pdf

\*^document is 43 pages long, a little bit too much with many blank cells

In Example 2a, the user generates a MS Word Document using Georgia font (rather than Ariel), designates a file name, and suppresses empty variables from the output document using the *noempty* option, and includes notes. Given the length of the full DHS dataset, this shortens the printed case from 43 to 9 pages.

/\*EXAMPLE 2a: print 100th case with some options

1) as a word document to demonstrate the footer;

2) suppressing empty variables to shorten the overall file;

3) adding in notes; \*/

printcase using "./printcase\_submission3/zzz\_example2" if id==100, noempty addnotes

In example 2b, we improve legibility of the output by specifying a serif font (Georgia) and controlling the column width to reduce white space and generate an even more compact case for browsing. The *replace* option allows the user to write over the previously printed case for this observation.

/\*4) now customize the font;

5) control width of columns;

6) use replace option to manage the file conflict \*/

printcase using "./printcase\_submission3/zzz\_example2" if id==100, font("Georgia") width(20, 40, 40) noempty addnotes replace

Example 3 shows how analysts can use printcase to view longitudinal or nested data. To illustrate, we use the DHS children’s model dataset in which the units are children, and they are nested within mothers (id) but the same logic can apply to longitudinal datasets where the respondent (id) contributes multiple observations over time. First, the researcher prepares the dataset, reducing the number of variables to be examined, deleting rows that indicate children (up to 14) that are not reported in the birth histories (up to 4), and ensuring that the order reflects the logic of the dataset (i.e., children young to old or survey wave, first to last).

/\*EXAMPLE 3: A LONGITUDINAL EXAMPLE\*/

\*https://dhsprogram.com/data/Download-Model-Datasets.cfm?flag=1 : children's model dataset

use "/Users/JennyTrinitapoli/Box/printcase\_jt/printcase\_submission3/ZZBR62FL.DTA", clear

\*1) make a functioning woman-level ID from the sampling pieces

egen id=concat(v001 v002 v003)

destring id, replace

order id

\*2) Sort the data as necessary. If data is truly longitudinal, user should order the dataset by the j variable,

\*as responses will be arrayed in the order they appear in the dataset itself. In this example childrenare ordered by year of birth \*/

sort id bidx

\*3) drop all woman-specific variables that do not vary across children and were already examined above

drop aw\* v\*

\*4) reduce dataset, dropping empty rows for children who aren't indexed in the birth history

drop if midx==.

\*5) sort dataset by i and j variables to ensure correct ordering of columns

Once the dataset is prepped, this analyst prints a case that includes all the corresponding observations in columns using the longitudinal option and specifying that the units for the columns should be labeled “Child”. Because this requires more horizontal space, we use the landscape option to rotate the page orientation and the width option to control column width for the three children (observations) reported by this mother (non-unique id).

\*6) print a case that allows the researcher to view reports on all children reported by one woman

printcase using "./printcase\_submission3/zzz\_example3" if id==1151, longitudinal unit("Child") landscape width(12, 22, 22, 22, 22) noempty

sort id midxprintcase using "./printcase\_submission3/zzz\_example2" if id==100, font("Georgia") width(20, 40, 40) noempty addnotes replace

Table

Description automatically generated

Figure 1. Screenshot of *printcase* output from Example 2b.

# 4 Conclusions

There are many reasons that reading questionnaires vertically has never caught on as standard practice in survey research. Questionnaires need to be stored carefully and kept confidential. Paper is heavy and difficult to transport. Oftentimes, the paper questionnaires have already been destroyed as part of the data-security protocol. And the quantitative scholar’s goal of making inferences fundamentally rests on our ability to identify statistical regularities – not to over-interpret particular cases. Still, the ability to look closely at a single observation in cross-sectional data is sometimes valuable. In particular, being able to produce a neat, readable quasi-questionnaire directly from a dataset – without headache – when necessary will enhance the workflow of data collection for many fieldworkers.

# 5 Acknowledgements

We’re grateful to the following people for feedback on the printcase command and the text of this article: Kathleen Broussard, Abdallah Chilungo, Ann Moore, Alex Weinreb, and the editors and reviewers from *The Stata Journal*. The Demographic and Health Surveys Program kindly shared DHS model datasets to produce a realistic example.

# 6 References

Bledsoe, Caroline, Fatoumatta Banja, and Allan G. Hill. 1998. “Reproductive Mishaps and Western Contraception: An African Challenge to Fertility Theory.” *Population and Development Review* 24(1):15–57.

Ebert, Jonas Fynboe, Linda Huibers, Bo Christensen, and Morten Bondo Christensen. 2018. “Paper- or Web-Based Questionnaire Invitations as a Method for Data Collection: Cross-Sectional Comparative Study of Differences in Response Rate, Completeness of Data, and Financial Cost.” *Journal of Medical Internet Research* 20(1):e24.

Hassler, Kendyl, Kelly J. Pearce, and Thomas L. Serfass. 2018. “Comparing the Efficacy of Electronic-Tablet to Paper-Based Surveys for on-Site Survey Administration.” *International Journal of Social Research Methodology* 21(4):487–97.

ICF. “Download Model Datasets.” The DHS Program website. Funded by USAID. http://www.dhsprogram.com. [Accessed June 20, 2020]. https://dhsprogram.com/data/Download-Model-Datasets.cfm

Kusumoto, Yasuaki, Yoshihiro Kita, Satomi Kusaka, Yoshinori Hiyama, Junko Tsuchiya, Toshiki Kutsuna, Hiroyuki Kameda, Saori Aida, Masaru Umeda, and Tetsuya Takahashi. 2017. “Difference between Tablet Methods and Paper Questionnaire Methods of Conducting a Survey with Community-Dwelling Elderly.” *Journal of Physical Therapy Science* 29(12):2100–2102.

Leahey, Erin. 2008. “Overseeing Research Practice : The Case of Data Editing.” *Science, Technology & Human Values* 33(5):605–30.

Leahey, Erin, Barbara Entwisle, and Peter Einaudi. 2003. “Diversity in Everyday Research Practice The Case of Data Editing.” *Sociological Methods & Research* 32(1):64–89.

Newell, Steve M., Henrietta L. Logan, Yi Guo, John G. Marks, and James A. Shepperd. 2015. “Evaluating Tablet Computers as a Survey Tool in Rural Communities.” *The Journal of Rural Health : Official Journal of the American Rural Health Association and the National Rural Health Care Association* 31(1):108–17.

Pearce, Lisa D. 2002. “Integrating Survey and Ethnographic Methods for Systematic Anomalous Case Analysis.” *Sociological Methodology* 32(1):103–32.

Sana, Mariano and Alexander A. Weinreb. 2008. “Insiders, Outsiders, and the Editing of Inconsistent Survey Data.” *Sociological Methods Research* 36(4):515–41.

Waal, Ton de, Jeroen Pannekoek, and Sander Scholtus. 2011. *Handbook of Statistical Data Editing and Imputation*. John Wiley & Sons.

# 7 About the authors

Max Weinreb is an undergraduate student at the University of Texas at Austin pursuing a B.S. in computer science with a minor in business. He is the author of the printcase command.

Jenny Trinitapoli is associate professor of sociology at the University of Chicago. Since 2009, she has been PI of the Tsogolo La Thanzi project, spearheading the collection of 12 rounds of data from over 3000 young adults living in Balaka, Malawi. She envisioned and commissioned the printcase command out of necessity, when transitioning from paper-based to e-tablet data collection with an experienced data collection team.

# 8 Appendix A: Full example from the Demographic and Health Surveys Program (Example 2b)

id 100

# User: RedactedForPeerReview

# ./printcase\_submission3/ZZBR62FL.DTA

# Date Printed: 26 Apr 2022

|  |  |  |
| --- | --- | --- |
| Variable Name | Variable Label | Response |
| id |  | 100 |
| caseid | case identification | 1 30 2 |
| bidx | birth column number | 2 |
| v000 | country code and phase | ZZ6 |
| v001 | cluster number | 1 |
| v002 | household number | 30 |
| v003 | respondent's line number | 2 |
| v004 | ultimate area unit | 1 |
| v005 | women's individual sample weight (6 decimals) | 1057703 |
| v006 | month of interview | 6 |
| v007 | year of interview N: check if this is ethiopian, nepali, or gregorian calendar | 2015 |
| v008 | date of interview (cmc) | 1386 |
| v009 | respondent's month of birth | 3 |
| v010 | respondent's year of birth | 1981 |
| v011 | date of birth (cmc) | 975 |
| v012 | respondent's current age | 34 |
| v013 | age in 5-year groups | 30-34 |
| v014 | completeness of age information | month and year - information complete |
| v015 | result of individual interview | completed |
| v016 | day of interview | 29 |
| v017 | cmc start of calendar | 1321 |
| v018 | row of month of interview | 15 |
| v019 | length of calendar | 66 |
| v019a | number of calendar columns | 2 |
| v020 | ever-married sample | all woman sample |
| v021 | primary sampling unit | 1 |
| v022 | sample strata for sampling errors | 26 |
| v023 | stratification used in sample design | region 2 - rural |
| v024 | region | region 2 |
| v025 | type of place of residence | rural |
| v027 | number of visits | 1 |
| v028 | interviewer identification | 225 |
| v029 | keyer identification | 7 |
| v030 | field supervisor | 221 |
| v031 | field editor | 222 |
| v032 | office editor | 1 |
| v034 | line number of husband | 1 |
| v042 | household selected for hemoglobin | selected |
| v044 | selected for domestic violence module | woman not selected |
| v101 | region | region 2 |
| v102 | type of place of residence | rural |
| v106 | highest educational level | no education |
| v113 | source of drinking water | protected well |
| v115 | time to get to water source | 60 |
| v116 | type of toilet facility | pit latrine with slab |
| v119 | household has: electricity | no |
| v120 | household has: radio | yes |
| v121 | household has: television | no |
| v122 | household has: refrigerator | no |
| v123 | household has: bicycle | no |
| v124 | household has: motorcycle/scooter | no |
| v125 | household has: car/truck | no |
| v127 | main floor material | earth, sand |
| v128 | main wall material | bricks |
| v129 | main roof material | metal sheets |
| v130 | religion N: religious groups are ambiguous in the model dataset | religion 2 |
| v131 | ethnicity | ethnic group 4 |
| v133 | education in single years | 0 |
| v135 | usual resident or visitor | usual resident |
| v136 | number of household members (listed) | 4 |
| v137 | number of children 5 and under in household (de jure) | 1 |
| v138 | number of eligible women in household (de facto) | 1 |
| v139 | de jure region of residence | region 2 |
| v140 | de jure type of place of residence | rural |
| v149 | educational attainment | no education |
| v150 | relationship to household head | wife |
| v151 | sex of household head | male |
| v152 | age of household head | 42 |
| v153 | household has: telephone (land-line) | no |
| awfactt | all woman factor - total | 100 |
| awfactu | all woman factor - urban/rural | 100 |
| awfactr | all woman factor - regional | 100 |
| awfacte | all woman factor - educational | 100 |
| awfactw | all woman factor - wealth index | 100 |
| v155 | literacy | cannot read at all |
| v157 | frequency of reading newspaper or magazine | not at all |
| v158 | frequency of listening to radio | at least once a week |
| v159 | frequency of watching television | not at all |
| v160 | toilet facilities shared with other households | yes |
| v161 | type of cooking fuel | wood |
| v167 | number of trips in last 12 months | 3 |
| v168 | away for more than one month in last 12 months | yes |
| v190 | wealth index | middle |
| v191 | wealth index factor score (5 decimals) | -22522 |
| ml101 | type of mosquito bed net(s) slept under last night | no net |
| v201 | total children ever born | 2 |
| v202 | sons at home | 1 |
| v203 | daughters at home | 1 |
| v204 | sons elsewhere | 0 |
| v205 | daughters elsewhere | 0 |
| v206 | sons who have died | 0 |
| v207 | daughters who have died | 0 |
| v208 | births in last five years | 1 |
| v209 | births in past year | no births |
| v210 | births in month of interview | 0 |
| v211 | date of first birth (cmc) | 1276 |
| v212 | age of respondent at 1st birth | 25 |
| v213 | currently pregnant | no or unsure |
| v215 | time since last menstrual period | 105 |
| v216 | menstruated in last six weeks | yes |
| v217 | knowledge of ovulatory cycle | don't know |
| v218 | number of living children | 2 |
| v219 | living children + current pregnancy | 2 |
| v220 | living children + current pregnancy (grouped) | 2 |
| v221 | marriage to first birth interval (months) | 90 |
| v222 | last birth to interview (months) | 36 |
| v224 | entries in birth history | 2 |
| v226 | time since last period (comp) (months) | 0 |
| v227 | flag for last period | no flag |
| v228 | ever had a terminated pregnancy | no |
| v235 | index last child prior to maternity-health (calendar) | 2 |
| v237 | birth between last and interview | no |
| v238 | births in last three years | 0 |
| v312 | current contraceptive method | not using |
| v313 | current use by method type | no method |
| v361 | pattern of use | never used |
| v362 | intention to use | does not intend |
| v364 | contraceptive use and intention | does not intend to use |
| v367 | wanted last child | wanted then |
| v384a | heard family planning on radio last few months | yes |
| v384b | heard family planning on tv last few months | no |
| v384c | heard family planning in newspaper/magazine last few months | no |
| v393 | visited by family planning worker last 12 months | yes |
| v394 | visited health facility last 12 months | yes |
| v395 | at health facility, told of family planning | yes |
| v3a00a | source of family planning for non-users: government hospital | no |
| v3a00b | source of family planning for non-users: government health center | no |
| v3a00c | source of family planning for non-users: family planning clinic | no |
| v3a00d | source of family planning for non-users: public mobile clinic | no |
| v3a00e | source of family planning for non-users: public health worker | no |
| v3a00f | source of family planning for non-users: other public | no |
| v3a00j | source of family planning for non-users: private hospital/clinic | no |
| v3a00k | source of family planning for non-users: pharmacy | no |
| v3a00l | source of family planning for non-users: private doctor | no |
| v3a00m | source of family planning for non-users: private mobile clinic | no |
| v3a00n | source of family planning for non-users: private health worker | no |
| v3a00o | source of family planning for non-users: other private | no |
| v3a00s | source of family planning for non-users: shop | no |
| v3a00t | source of family planning for non-users: church | no |
| v3a00u | source of family planning for non-users: friend/relative | no |
| v3a00x | source of family planning for non-users: other | no |
| v3a00y | source of family planning for non-users: no source | yes: knows no source |
| v3a00z | source of family planning for non-users: any source | no |
| v401 | last birth a caesarean section | no |
| v404 | currently breastfeeding | no |
| v405 | currently amenorrheic | no |
| v406 | currently abstaining | no |
| v416 | heard of oral rehydration | heard of ors |
| v417 | entries in maternity table | 1 |
| v418 | entries in health table | 1 |
| v419 | entries in height/weight table | 1 |
| v426 | when child put to breast | 202 |
| v437 | respondent's weight in kilograms (1 decimal) | 417 |
| v438 | respondent's height in centimeters (1 decimal) | 1602 |
| v439 | height/age percentile | 2782 |
| v440 | height/age standard deviation | -59 |
| v441 | height/age percent ref. median | 9786 |
| v442 | weight/height percent ref. median (dhs) | 6660 |
| v443 | weight/height percent ref. median (fog) | 7910 |
| v444 | weight/height percent ref. median (who) | 8516 |
| v444a | weight/height standard deviation (dhs) | -278 |
| v445 | body mass index | 1625 |
| v446 | rohrer's index | 1014 |
| v447 | result of measurement - height/weight | measured |
| v447a | women's age in years (from household questionnaire) | 34 |
| v452a | under age 18 (from household questionnaire) | age 18 or older |
| v452c | read consent statement - hemoglobin | granted |
| v453 | hemoglobin level (g/dl - 1 decimal) | 137 |
| v454 | currently pregnant (from household questionnaire) | no/don't know |
| v455 | result of measurement - hemoglobin | measured |
| v456 | hemoglobin level adjusted for altitude and smoking (g/dl - 1 decimal) | 137 |
| v457 | anemia level | not anemic |
| v459 | have mosquito bed net for sleeping (from household questionnaire) | no |
| v460 | children under 5 slept under mosquito bed net last night (household questionnair N: children under 5 slept under mosquito bed net last night (household questionnaire) | no net in household |
| v461 | respondent slept under mosquito bed net | no |
| v463a | smokes cigarettes | no |
| v463b | smokes pipe | no |
| v463c | uses chewing tobacco | no |
| v463d | uses snuff | no |
| v463x | smokes other | no |
| v463z | does not use tobacco | yes, smokes nothing |
| v465 | disposal of youngest child's stools when not using toilet | put/rinsed in toilet/latrine |
| v467b | getting medical help for self: getting permission to go | no problem |
| v467c | getting medical help for self: getting money needed for treatment | no problem |
| v467d | getting medical help for self: distance to health facility | not a big problem |
| v467f | getting medical help for self: not wanting to go alone | no problem |
| v468 | record for last birth | last birth only |
| v473a | read consent statement - hiv | granted |
| v473b | result of measurement - hiv | blood taken |
| v477 | number of injections in last 12 months | none |
| v481 | covered by health insurance | no |
| v481a | health insurance type: mutual/community organization | no |
| v481b | health insurance type: provided by employer | no |
| v481c | health insurance type: social security | no |
| v481d | health insurance type: private/commercially purchased | no |
| v481x | health insurance type: other | no |
| v501 | current marital status | married |
| v502 | currently/formerly/never in union | currently in union/living with a man |
| v503 | number of unions | once |
| v504 | currently residing with husband/partner | living with her |
| v505 | number of other wives | no other wives |
| v507 | month of first cohabitation | 10 |
| v508 | year of first cohabitation | 1998 |
| v509 | date of first cohabitation (cmc) | 1186 |
| v510 | completeness of date information in v509 | year - age/month imputed |
| v511 | age at first cohabitation | 17 |
| v512 | years since first cohabitation | 16 |
| v513 | cohabitation duration (grouped) | 15-19 |
| v525 | age at first sex | at first union |
| v527 | time since last sex | 106 |
| v528 | time since last sex (in days) | 6 |
| v529 | time since last sex (in months) | 0 |
| v530 | flag for v529 | no flag |
| v531 | age at first sex (imputed) | 17 |
| v532 | flag for v531 | no flag |
| v536 | recent sexual activity | active in last 4 weeks |
| v613 | ideal number of children | 5 |
| v614 | ideal number of children (grouped) | 5 |
| v621 | husband's desire for children | both want same |
| v623 | exposure | fecund |
| v625 | exposure (definition 2) | fecund |
| v625a | exposure to need for contraception (definition 3) | fecund |
| v627 | ideal number of boys | 2 |
| v628 | ideal number of girls | 3 |
| v629 | ideal number of either sex | 0 |
| v633b | reason for not having sex: husband has other women | no |
| v701 | husband/partner's education level | no education |
| v704 | husband/partner's occupation | market-oriented skilled agricultural and fishery workers |
| v705 | husband/partner's occupation (grouped) | agricultural - self employed |
| v714 | respondent currently working | yes |
| v715 | husband/partner's total number of years of education | 0 |
| v716 | respondent's occupation | market-oriented skilled agricultural and fishery workers |
| v717 | respondent's occupation (grouped) | agricultural - self employed |
| v719 | respondent works for family, others, self | self-employed |
| v729 | husband/partner's educational attainment | no education |
| v730 | husband/partner's age | 42 |
| v731 | respondent worked in last 12 months | currently working |
| v732 | respondent employed all year/seasonal | seasonal |
| v741 | type of earnings from respondent's work | not paid |
| v743a | person who usually decides on respondent's health care | husband/partner alone |
| v743b | person who usually decides on large household purchases | husband/partner alone |
| v743d | person who usually decides on visits to family or relatives | husband/partner alone |
| v743f | person who usually decides what to do with money husband earns | husband/partner alone |
| v744a | beating justified if wife goes out without telling husband | no |
| v744b | beating justified if wife neglects the children | no |
| v744c | beating justified if wife argues with husband | no |
| v744d | beating justified if wife refuses to have sex with husband | no |
| v744e | beating justified if wife burns the food | no |
| v745a | owns a house alone or jointly | jointly only |
| v745b | owns land alone or jointly | jointly only |
| bord | birth order number | 1 |
| b0 | child is twin | single birth |
| b1 | month of birth | 4 |
| b2 | year of birth | 2006 |
| b3 | date of birth (cmc) | 1276 |
| b4 | sex of child | female |
| b5 | child is alive | yes |
| b8 | current age of child | 9 |
| b9 | child lives with whom | respondent |
| b10 | completeness of information | month and year - information complete |
| b12 | succeeding birth interval (months) | 74 |
| b16 | child's line number in household | 3 |
| g100 | ever heard of female circumcision | yes |
| g102 | respondent circumcised | yes |
| g103 | flesh removed from genital area | yes |
| g105 | genital area sewn closed | yes |
| g106 | age at circumcision (in years) | 15 |
| g107 | person who performed circumcision | traditional |
| g108 | number of daughters circumcised | no daughter circumcised |
| g116 | intends to have daughter(s) circumcised in future | yes |
| g117a | cleanliness/hygiene: female circumcision benefit | no |
| g117b | social acceptance: female circumcision benefit | yes |
| g117c | better marriage prospects: female circumcision benefit | yes |
| g117d | virginity/prevent premarital sex: female circumcision benefit | no |
| g117e | greater sexual pleasure for men: female circumcision benefit | no |
| g117f | religious approval: female circumcision benefit | no |
| g117j | no opinion/don't know: female circumcision benefit | no |
| g117x | other: female circumcision benefit | no |
| g117y | no benefit from female circumcision | no |
| g118 | female circumcision required by religion | yes |
| g119 | female circumcision: continue or be stopped | continued |
| szone | geographical zone | central |
| shreg | health regions | periphery |