

---

Subject: Explanation of IPUMS-DHS

Posted by [boyle014](#) on Wed, 02 Dec 2015 13:57:26 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

IPUMS-DHS ([www.idhsdata.org](http://www.idhsdata.org)) is a new tool to help researchers use the DHS to study issues over time and across countries. Created at the Minnesota Population Center, with the support of ICF International, USAID, and NICHD, IPUMS-DHS eliminates the need to download and merge or append multiple DHS files. Instead, researchers select the variables, countries and years of interest to them, and create a single, customized dataset within minutes. The tool is completely free to users.

Currently, IPUMS-DHS includes surveys from Egypt, India, and many African countries: Benin, Burkina Faso, Cote D'Ivoire, Ethiopia, Ghana, Guinea, Kenya, Malawi, Mali, Mozambique, Niger, Nigeria, Tanzania, Uganda, Zambia, and Zimbabwe. It allows researchers to select WOMEN or CHILDREN as their unit of analysis and includes more than 2000 variables, including geographic variables that are harmonized over time and a number of variables that, with the regular DHS, would require downloading and merging the separate HR (household) files.

Anyone can browse the variables in IPUMS-DHS, where it is easy to identify comparability issues, such as universe or question wording differences across surveys. To download a dataset, users need to login using their DHS usernames and passwords. They will have access to data from the countries for which they are registered through The DHS Program.

If the tool does not currently include the variables or surveys that most interest you, please check back. New countries, units of analysis, and variables will be added annually. The next update is scheduled for March 2016. It will add three more African countries, BIRTHS as a unit of analysis, anthropometric information for children, and many other variables.

---

Subject: Re: Explanation of IPUMS-DHS

Posted by [Tolu238](#) on Mon, 30 Jan 2017 04:12:49 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Thank you for your help with the IPUMS data. I want to merge IPUMS-DHS data with the original child recode dataset for Nigeria for the 03, 08 and '13 surveys.

On the original dhs dataset, I created sample and was able to create 'idhspid' for 03 and 08 surveys but the 2013 survey turns out differently - does

not leave any space between the sample identifier and the caseid. I will appreciate your advice.  
TA

---

Subject: Re: Explanation of IPUMS-DHS

Posted by [boyle014](#) on Mon, 30 Jan 2017 12:16:31 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Thanks for the query. Can you please tell me what variables you need that are not in IPUMS-DHS?

I will forward this to one of my colleagues for more information on how to work with the IDs for Nigeria.

By the way, we are greatly expanding the number of variables in the next release of IPUMS-DHS, scheduled for April.

Liz Boyle  
Professor, Sociology  
Principal Investigator, IPUMS-DHS  
University of Minnesota

---

---

Subject: Re: Explanation of IPUMS-DHS  
Posted by [Tolu238](#) on Mon, 30 Jan 2017 16:21:41 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Thanks for your prompt response.  
I want to use the two variables which indicate household ownership of radio and television (v120 and v121) from the children's recode file.  
I wish I could wait but April is a bit far off for this project. However, I appreciate the efforts of IPUMS in making DHS data more usable. You have helped to drastically reduce time spent searching and linking files. Keep up the good job.

Tolu

---

---

Subject: Re: Explanation of IPUMS-DHS  
Posted by [gange026](#) on Tue, 31 Jan 2017 19:51:14 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Merging between the original DHS files and IPUMS-DHS data extracts will become less necessary as we add more variables to IPUMS-DHS. We expect to double the number of IPUMS-DHS variables in our April 2017 data release. Below is some guidance on linking between an IPUMS-DHS extract file and the original DHS files, using the latest Nigerian sample as an example.

Merging IPUMS-DHS data with original DHS data requires all data files to have a linking key, a unique identifier that is identical in name and in character length. To merge individual-level data, the original DHS files need to have a variable called "IDHSPID" as the linking key. This is a unique identifier for each respondent. In IPUMS-DHS, IDHSPID is a concatenation of SAMPLE (a 4-digit number representing the country and year of the survey) and CASEID, which is a sample-specific unique identifier for the respondent.

To create IDHSPID, CASEID is assumed to be right-justified. This means that there are leading

blanks in the data that cause CASEID to occupy the full variable width in the data, even if there are not numbers or letters to fill it. If you look at the data browser for the original 2003 or 2008 Nigerian DHS files, you should be able to see that there are a few spaces in every line of CASEID before the numbers begin. In the 2013 Nigeria children's recode, CASEID is not right-justified, and these spaces are not there.

An easy way to change this is to run the following command in Stata in your file for Nigeria 2013:  
`replace caseid=substr(" ", 1, 15 -length(caseid)) + caseid`  
There should be 15 spaces/blanks in the above quotation marks. If you then generate IDHSPID, there should be the appropriate number of spaces between the sample identifier and the CASEID.

A couple of other notes that may help you as you get further along in the merging process:

1. You will need to make sure IDHSPID is the same length in both the original DHS and IPUMS-DHS files. IPUMS-DHS stores IDHSPID in a 22-column string variable (str22), and you will want to make sure that when you create IDHSPID, that also has a width of 22. To do this you can add spaces at the beginning of the variable when you create it:

```
gen idhspid = " " + string(sample) + caseid
```

There should be 3 spaces/blanks in the quotation marks above to make up for differences in the width of CASEID in the original DHS data (15 columns) and IPUMS-DHS data (18 columns). The above code also assumes SAMPLE is a number rather than a string. If your version of SAMPLE is a string, you can simply run:

```
gen idhspid = " " + sample + caseid
```

2. If you are merging the original child recode dataset with children's data from IPUMS-DHS, you should keep in mind that IDHSPID is a unique identifier for the respondent, but not necessarily for children; children with the same mother will have the same value for IDHSPID. If you want to merge children's data with children's data, you should generate another variable that is a concatenation of IDHSPID and BIDX (Child's birth history index number) in both your original DHS data and IPUMS-DHS data. This will be a unique identifier for each child in the data. For example:

```
gen idhspidk = idhspid + string(bidx)
```

When you merge, this is the variable you should use as your linking key:

```
merge 1:1 idhspidk using [filename]
```

If the IPUMS-DHS data you are using is from the women's file, you do not need to create this additional variable, and you should merge your files using IDHSPID as your linking key.

I have also attached a sample do-file for Stata, that goes through every step of the merging process with data from the children's recode files from Nigeria 2003, 2008, and 2013, and an IPUMS-DHS extract using children as the unit of analysis.

Please let us know if you have any more questions.

-IPUMS-DHS staff

## File Attachments

1) [merging\\_instructions\\_ng.do](#), downloaded 1023 times

---

Subject: Re: Explanation of IPUMS-DHS  
Posted by [Michaelo](#) on Mon, 21 Feb 2022 10:20:58 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Thank you IPUMS team for the great work done in helping us pooling DHS data.  
I wish to use the most recent children dataset for a multi country analysis for sub Sahara Africa.  
However I am unable to access the sample for the following countries from your website.

Cameroon 2018 DHS  
Congo 2011-12 DHS  
Gabon 2012 DHS  
Gambia 2019-20 DHS  
Liberia 2019-20 DHS  
Rwanda 2019-20 DHS  
Senegal 2019 DHS  
Sierra Leone 2019 DHS  
Togo 2013-14 DHS.

I would be most grateful if you kindly assist me  
Thank you  
Michael