
Subject: Re: matching WHO weight/height scores to older surveys

Posted by [Janet-DHS](#) on Thu, 28 Apr 2022 14:56:42 GMT

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

Following is response from DHS Research & Data Analysis Director, Tom Pullum:

With Benin 2001 (BJ41) and some other surveys the HW file includes a variable hwhhid that is a string containing hv001 and hv002. You have to extract hv001 and hv002, which requires figuring out which columns of hwhhid contain them. The following code will do this:

```
use "...BJHW41FL.DTA", clear
```

```
tab hwlevel
```

```
* household file; the merge is with the PR file
```

```
* Look at the values of hwhhid
```

```
list hwhhid if _n<=20, table clean
```

```
describe hwhhid
```

```
* We see that hwhhid is a 12-character string; must unpack to get cluster number and household number
```

```
forvalues lc=1/12 {
```

```
gen c_`lc'=substr(hwhhid,`lc',1)
```

```
}
```

```
list c_* hwhhid if _n<=20, table clean compress
```

```
* It appears that the cluster id is in cols 1-10 and the household id is in cols 11-12
```

```
gen hv001=substr(hwhhid,1,10)
```

```
gen hv002=substr(hwhhid,11,2)
```

```
* convert from string to numeric
```

```
destring hv001 hv002, replace
```

```
drop c_*
```

```
rename hwline hvidx
```

```
sort hv001 hv002 hvidx
```

I think that's what you are looking for. In the HW file, hwline is the line number of the child. You can then merge with the child in the PR file. The mother's line number, if you need it, is hv112 on the child's record.

For the CAR 1994-95 data (CF31), the merge is with the KR file, not the PR file.

```
use "...CFHW31FL.DTA", clear
```

tab hwlevel

* individual file rather than household file; the merge is with the KR file

tab hwline

* It appears that hwline is bidx

* Look at the values of hwcaseid

list hwcaseid hwline if _n<=20, table clean

describe hwcaseid

* We see that hwcaseid is a 15-character string; must unpack to get cluster number and household number

```
forvalues lc=1/15 {  
gen c_`lc'=substr(hwcaseid,`lc',1)  
}
```

list c_* hwcaseid hwline if _n<=20, table clean compress

* It appears that the cluster id is in cols 1-10 and the household id is in 11-12, the mother's line is in 13-15

```
gen v001=substr(hwcaseid,1,10)
```

```
gen v002=substr(hwcaseid,11,2)
```

```
gen v003=substr(hwcaseid,13,3)
```

* convert from string to numeric
destring v001 v002 v003, replace

drop c_*

rename hwline bidx

sort v001 v002 v003 bidx

* Next merge with the KR file

Here, hwline is bidx, the sequence of the child (bidx=1 is the youngest child).