
Using the IMAGE Databases

1. On this CD is all the documentation and data you need to guide your use of the IMAGE Databases.
2. We recommend you use the documentation held here regularly – all the questionnaires, fieldwork manuals, a data dictionary, monographs, protocols and descriptive information on data distributions are held on the CD.
3. The data stored on this CD includes;
 - a. All information from the Participatory Wealth Ranking Process conducted in 2001 and used a sampling frame for the study
 - b. All Senior Female, Young Person and Household Questionnaires for Cohort 1 and 2 collected at baseline and follow up, as well as process information on attendance at IMAGE group meetings and loan cycles for SEF clients.
 - c. All Young Person and Household Questionnaires for Cohort 3 collected at baseline and follow up
4. The data is stored in full in Access databases. However, additionally, the most important data is held in two large Stata files and we expect people to use these most often. Below are some instructions for using these databases.
5. When you first receive the CD, make a full copy of its contents to your hard drive. Keep the CD safe. When you copy data from a CD it is often labeled by your computer as “read only” – meaning you cannot change it. In most cases, this might be OK – but it is a problem if you want to use the Access databases. If you wish to change this setting you need to go into Windows Explorer, right click the file, and go to Properties. At the bottom of the window that opens is a check box marked “Read only”. Uncheck this.
6. There are two Stata files. One contains all the main information relevant to Cohorts I and II, while the other contains all relevant information for Cohort III.
7. If you are going to use the Stata data, after you have opened Stata you need to increase the memory in order to open the file. When you first open the files they are very large. They contain much more information than you are ever likely to need when conducting a single analysis. However, in a few easy steps using Stata commands it is possible to trim the file down so that it contains only the people and the data that you need. Before doing this we suggest that you save a copy of the file with a new name so you don't overwrite the primary information.

Thus, the first three lines of a Stata do file will probably read ...

```
set mem 256m
use “[File pathway and name]”
save “[File pathway and name_changed]”, replace (1)
```

8. Now you are ready to tailor the file for your own needs. You will almost certainly need to do two things.
 - a. To trim down the files so it contains only people (or households) you are interested in. This is **essential**.
 - b. To trim down the file so it contains only relevant data on those individuals (or households). This is not essential but **recommended**.

9. To trim down the files so it contains only people (or households) you are interested in

Stata has three main commands that are useful here – *drop if*, *keep if* and *duplicates drop*. By combining the use of these commands with a few special fields in the database it is possible in just a few lines to limit the dataset to the **population** you need. This might be all young people, all senior females, all household members or a number of other potentially useful groups.

10. To trim down the file so it contains only the relevant data on those individuals (or households)

Stata has two commands that are useful here – *drop* and *keep*. By combining the use of these commands it is possible in just a few lines to limit the file to the **data** you need.

11. **Appendix I** contains examples of the code you would need to use to limit your dataset to a number of relevant **populations** and **data sets**. As a starting point you can identify the group you are interested in and copy the code held in Appendix I to a Stata do file.
12. You are now ready to go about your analysis. We recommend using Stata do files to keep a record of the commands you run, particularly any commands where you generate or recode variables. This has a number of advantages that make data analysis easier.
13. Before closing Stata you may want to save a data file with any new variables you have created in it. Note that as long as you have kept a record of the commands you have run, you will always be able to get your codings back from the original data so saving a file may not be essential. If you do save a file we would recommend you re-save the file with a different file name.
14. Questions and queries about the data can be directed to RADAR through james.hargreaves@lshtm.ac.uk.

James, Edwin and Rico
28th Oct 2005

The IMAGE Study Database Populations at a glance

1) Participatory Wealth Ranking

Total Households in Sampling Frame	9824
Total Number of PWR Reference Groups	79
Characteristic Statements	8865
Descriptive statements	3671
Codes	880

2) Cohort I & II

	Eligible for inclusion	Data	
		Baseline	Follow Up
Senior females (I)	857	843	750 (746 interviewed at baseline)
Households	924 (838 eligible at baseline)	833	888 (808 interviewed at baseline)
Household Members	-	5816 (from 833 households)	6802
Young people (II)	1835	1456	1059 (of 1456 eligible)*

3) Cohort III

	Eligible for inclusion	Data	
		Baseline	Follow Up
Households	1816 (1640 eligible at baseline)	1482 (of 1640 eligible)	1543 (1390 interviewed at baseline)
Household Members		9653 (from 1482 households)	11570
Young people (III)	3882	2858	2325 (1967 interviewed at baseline)*

* Note that there were slightly different eligibility criteria for follow up for Cohort II and Cohort III. Only those successfully interviewed at baseline were eligible to be followed up in Cohort II, while all those eligible at baseline were eligible for follow up in Cohort III.

Appendix I

This document contains Stata code that could be used in do files to trim down the IMAGE data stata files so they contain only people (or households) and data you are interested in. In each case the code contains two sections.

The first section removes unwanted records so that your data set contains only those records you are interested in – this stage is **essential**.

The second section limits the data to the most relevant fields and is **optional**.

A detailed example, with explanation is held below.

<pre>*13. All SFs eligible at baseline keep if senfem!=1 & (fvfcode>0.5 & fvfcode<9999) duplicates drop ypidl, force count keep villnum intcom ypidl f* drop ff*</pre>	<p>Stage 1 This line uses fields in the database that identify the particular group we are interested in to limit the data</p> <p>This command removes duplicates of the records of interest, leaving you with the population you are interested in</p> <p>Its good to always check that the count of records you get here agrees with what you expect. In this case, it would be 857. (See IMAGE Study Populations at a glance)</p> <p>Stage 2 This command keeps only key index fields and fields beginning with 'f' (ie those relevant to senior females)</p> <p>Following this, all remaining fields starting 'ff' are removed from the data as these contain follow up information but the group we are interested in here is SFs eligible at baseline</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Note that the first part of the code is **essential** because you will get strange results unless you make sure you have the right number of individuals or households in your data set relevant to the group you are interested in.

The second part of the code is **recommended**. You don't need to limit the amount of data that you have in your data set, although it may prove more efficient to do so. Data limitation is achieved very simplistically in the code held below – limiting information to that held in the specific questionnaires aimed at the group of interest. Sometimes you might want to keep, for example, household level information in a dataset of young people. You can do this by manipulating the code we have provided – or at the simplest level you could use a single line of code to keep the variables you need.

keep [List of all the variables you need].

Good luck !

Cohort 1 and 2

*1. All dwellings identified in PWR

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear  
drop if hhidpwr==.  
duplicates drop hhidpwr, force  
count  
keep villnum hhidpwr refgpd1 bassurv score1 rank1 score2 rank2 score3  
rank3 scoreav scoreav_c incons eligible selecsf povstate povstact  
order villnum refgpd1 hhidpwr
```

*2. All HHs eligible for interview at baseline only

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear  
keep if hh_sampled==3  
duplicates drop hhid, force  
count  
drop f* y* s* h100* hh100*
```

*3. All HHs successfully interviewed at baseline

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear  
keep if hh_sampled==3 & (hvfcode==1)  
duplicates drop hhid, force  
count  
drop h100*  
drop hh100*  
keep villnum h* i*  
rename hhidpwr temp  
rename hhid temp1  
drop hh*  
rename temp hhidpwr  
rename temp1 hhid
```

*4. All HHs eligible for interview at both time points

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear  
keep if hh_sampled!=1 & hhid!=.  
duplicates drop hhid, force  
count  
drop h100* hh100*  
keep villnum hhidpwr hhid h* i*
```

*5. All HHs successfully interviewed at follow up

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear  
keep if hh_sampled!=1 & (hhvfcode==1)  
duplicates drop hhid, force  
count
```

```
drop h100* hh100* hhomdate hhom1 hhom2 hhdate_died hhd1 hhimdate hhim1
hhim2 hhdateofbirth hhb1
keep villnum hh*
```

*6. All HHs succesfully interviewed at both time points

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear
keep if hh_sampled!=1 & hhid!=. & hvfcode==1 & hhvfcode==1
duplicates drop hhid,force
count
drop h100* hh100*
keep villnum hhidpwr hhid h* i*
```

*7. All HHmems enumerated at baseline

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear
drop if h100a==.
duplicates drop ypidd1,force
count
keep villnum hhid ypidd1 h100*
```

*8. All HHmems enumerated at either timepoint

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear
drop if hhid==.
duplicates drop ypidd1, force
count
keep villnum hhid ypidd1 h100* hh100* hhomdate hhom1 hhom2 hhdate_died
hhd1 hhimdate hhim1 hhim2 hhdateofbirth hhb1
```

*9. All YPs eligible at baseline

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear
keep if young_p==3
duplicates drop ypidd1,force
count
keep villnum ypidd1 dateofbirth sex y*
```

*10. All YPs succesfully interviewed at baseline

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear
keep if young_p==3
keep if (yvfcode==1 | yvfcode==5)
duplicates drop ypidd1,force
count
drop yy*
keep villnum ypidd1 dateofbirth sex y*
drop young_p
```

*11. All YPs succesfully interviewed at follow up

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear  
keep if young_p==3  
keep if (yyvfcode==1 | yyvfcode==5)  
duplicates drop ypiddl, force  
count  
keep villnum ypiddl dateofbirth sex yy*
```

*12. All YPs succesfully interviewed at baseline and follow up

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear  
keep if young_p==3  
keep if (yvcode==1 | yvcode==5) & (yyvfcode==1 | yyvfcode==5)  
duplicates drop ypiddl, force  
count  
keep villnum ypiddl dateofbirth sex y*  
drop young_p
```

*13. All SFs eligible at baseline

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear  
keep if senfem!=1 & (fvfcode>0.5 & fvfcode<9999)  
duplicates drop ypiddl, force  
count  
keep villnum intcom ypiddl f*  
drop ff*
```

*14. All SFs succesfully interviewed at baseline

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear  
keep if senfem!=1 & (fvfcode==1 | fvfcode==5)  
duplicates drop ypiddl, force  
count  
keep villnum intcom ypiddl f*  
drop ff*
```

*15. All SFs succesfully interviewed at follow up

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear  
keep if senfem!=1 & (ffvfcode==1 | ffvfcode==5)  
keep villnum intcom ypiddl ff*  
count  
duplicates drop ypiddl, force
```

*16. All SFs succesfully interviewed at baseline and follow up

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 1 and 2\IMAGE Cohort 1 and 2.dta", clear
```

```
keep if senfem!=1 & (fvfcode==1 | fvfcode==5) & (ffvfcode==1 |  
ffvfcode==5)  
duplicates drop ypiddl, force  
count  
keep villnum intcom ypiddl f*
```


Cohort 3

*1. All dwellings identified in PWR

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 3\IMAGE Cohort 3.dta", clear  
drop if hhidpwr==.  
duplicates drop hhidpwr, force  
count  
keep villnum hhidpwr refigpdl bassurv score1 rank1 score2 rank2 score3  
rank3 scoreav scoreav_c incons eligible selecsf povstate povstact  
order villnum refigpdl hhidpwr
```

*2. All HHs eligible for interview at baseline only

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 3\IMAGE Cohort 3.dta", clear  
keep if hh_sampled==3  
duplicates drop hhid, force  
count  
drop f* y* s* h100* hh100*
```

*3. All HHs successfully interviewed at baseline

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 3\IMAGE Cohort 3.dta", clear  
keep if hh_sampled==3 & (hvfcode==1)  
duplicates drop hhid, force  
count  
drop h100*  
drop hh100*  
keep villnum h* i*  
rename hhidpwr temp  
rename hhid temp1  
drop hh*  
rename temp hhidpwr  
rename temp1 hhid
```

*4. All HHs eligible for interview at both time points

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 3\IMAGE Cohort 3.dta", clear  
keep if hh_sampled!=1 & hhid!=.  
duplicates drop hhid, force  
count  
drop h100* hh100*  
keep villnum hhidpwr hhid h* i*
```

*5. All HHs successfully interviewed at follow up

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE  
Cohort 3\IMAGE Cohort 3.dta", clear  
keep if hh_sampled!=1 & (hhvfcode==1)  
duplicates drop hhid, force  
count
```

```
drop h100* hh100* hhomdate hhom1 hhom2 hhdate_died hhd1 hhimdate hhim1
hhim2 hhdateofbirth hhb1
keep villnum hh*
```

*6. All HHs succesfully interviewed at both time points

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE
Cohort 3\IMAGE Cohort 3.dta", clear
keep if hh_sampled!=1 & hhid!=. & hvfcode==1 & hhvfcode==1
duplicates drop hhid,force
count
drop h100* hh100*
keep villnum hhidpwr hhid h* i*
```

*7. All HHmems enumerated at baseline

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE
Cohort 3\IMAGE Cohort 3.dta", clear
drop if h100a==.
duplicates drop ypiddl,force
count
keep villnum hhid ypiddl h100*
```

*8. All HHmems enumerated at either timepoint

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE
Cohort 3\IMAGE Cohort 3.dta", clear
drop if hhid==.
duplicates drop ypiddl, force
count
keep villnum hhid ypiddl h100* hh100* hhomdate hhom1 hhom2 hhdate_died
hhd1 hhimdate hhim1 hhim2 hhdateofbirth hhb1
```

*9. All YPs eligible at baseline

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE
Cohort 3\IMAGE Cohort 3.dta", clear
keep if young_p==3
duplicates drop ypiddl,force
count
keep villnum ypiddl dateofbirth sex y*
```

*10. All YPs succesfully interviewed at baseline

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE
Cohort 3\IMAGE Cohort 3.dta", clear
keep if young_p==3
keep if yvfcode==1
duplicates drop ypiddl,force
count
```

```
drop yy*
keep villnum ypiddl dateofbirth sex y*
drop young_p
```

*11. All YPs succesfully interviewed at follow up

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE
Cohort 3\IMAGE Cohort 3.dta", clear
keep if young_p==3
keep if yyvfcode==1
duplicates drop ypiddl,force
count
keep villnum ypiddl dateofbirth sex yy*
```

*12. All YPs succesfully interviewed at baseline and follow up

```
use "H:\Data\ac\image\Final Databases\Distribution CD\Databases\IMAGE
Cohort 3\IMAGE Cohort 3.dta", clear
keep if young_p==3
keep if yvfcode==1 & yyvfcode==1
duplicates drop ypiddl,force
count
keep villnum ypiddl dateofbirth sex y*
drop young_p
```