

Documentation file for Do-file and Datasets

“Education, HIV, and Early Fertility: Experimental Evidence from Kenya”

By Duflo, Dupas and Kremer

Do File

The do-file that generates the figures and tables in the paper is called
`analysis_public.do`

Data Sets

There are 6 datasets needed to run the do file. Each of them is described in details below.

DATASET 1 : studysample_allmerged

This database merge all the individual-level datasets collected from individuals in the sample:

- (1) Long-run follow-up survey: variables starting with Q
 Survey instrument is attached and called “Survey_Set_Eng.pdf”.
 All variables in the dataset are labelled.
- (2) A few variables from tracking checklists, all labeled
- (3) Roll call data: the data collected at each school visit. 9 school visits were made:
 - A pre-program baseline visit in 2003 (labeled 03v1)
 - Two visits in 2004 (labeled 04v1 and 04v2)
 - Three visits in 2005 (labeled 05v1, 05v2 and 05v3)
 - Two visits in 2006 (labeled 06v1 and 06v2)
 - One visit in 2007 (labeled 07v2)

At each visit, the following variables were collected:

<i>sch</i>	Primary school ID number in which student enrolled at time of visit
<i>std</i>	Grade in which student enrolled at time of visit 5 grade 5 6 grade 6 7 grade 7 8 grade 8 9 in secondary school 10 completed primary school, now at home 11 training with local artisan 20 training in polytechnic 21 training in vocational school 55 dropped out of school 77 dead 88 enrolled in primary school but grade unknown 99 pupil unknown
<i>age</i>	Age or year of birth were recorded if the information was missing
<i>pres</i>	If enrolled in school, presence at school on day of visit 1 Student present 2 Student absent 3 Student was present earlier in the day but went home
<i>secondary</i>	Name of secondary school in which student is enrolled (if applicable)
<i>preg</i>	(girls only): Pregnant? 1 Yes 2 No
<i>mar</i>	Married? 1 Yes 2 No
<i>child</i>	Number of children as of the visit
<i>abort</i>	When questions about pregnancy of children were asked, people sometime mentioned an abortion. If so, it was recorded as: 1 abortion mentioned

DATASET 2: hometracking.dta

Survey administered in Spring 2006 at the homes of a subset of female respondents. Survey instrument is attached and called “home tracking.pdf”.

DATASETS 3&4: KAPgirls.dta and KAPboys.dta

Anonymous in-class survey self-administered by students in grades 7 and 8 in 2005. Survey instrument varied slightly by gender. The two versions are attached and called PupKAP05_Girls.pdf and PupKAP05_Boys.pdf.

DATASET 5: school_info.dta

This database contains school-level characteristics that are all labeled.

DATASET 6: sampling_frame.dta

This database contains the list of all students enrolled in one of the 328 study schools at the onset of the study (2003 visit 1, or 03v1), in grades 5 to 8. It includes information on their gender (female=1 if female), school in 2003 (sch03v1) and that school treatment assignments (HIVtreat04v1 and Utreat03v1), grade in 2003 (std03v1), age in 2003 (age03), and their childbearing status as of visit 3 of 2003 and visit 1 of 2006. This frame was used to select students visited for the “hometracking” survey.