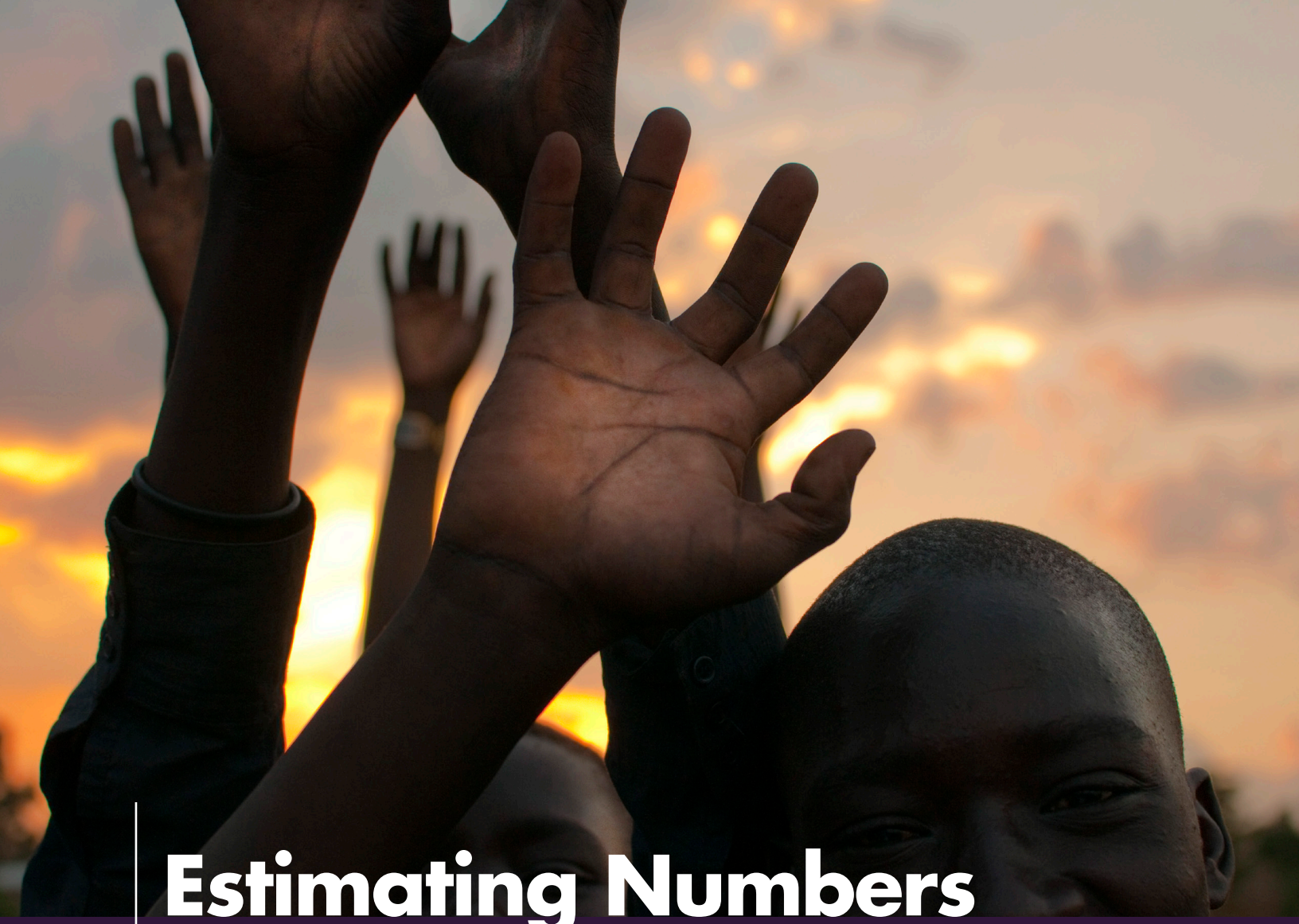


Estimating Numbers of Orphans and Vulnerable Children

A Test of Regression Modeling

Paul Brodish, Zulfiya Charyeva,
and **Karen Foreit**

February 2017

TR-17-154

USAID
FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE Evaluation

Estimating Numbers of **Orphans** and **Vulnerable Children**
A Test of Regression Modeling

Paul Brodish, MSPH, PhD
Zulfiya Charyeva, MA, MPH, PhD
Karen Foreit, MS, PhD

February 2017

Cover photo: Hands Color, Lisa Marie Albert, MEASURE Evaluation

MEASURE Evaluation
University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill, North Carolina 27517
Phone: +1-919-445-9350 • measure@unc.edu
www.measureevaluation.org

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. TR-17-154

ISBN: 978-1-9433-6434-3

USAID FROM THE AMERICAN PEOPLE
PEPFAR
MEASURE Evaluation

ACKNOWLEDGMENTS

We wish to acknowledge and express sincere gratitude for the contributions of the following people to the activity and preparation of this report:

- Christine Fu, Senior Research and Evaluation Advisor, Office of HIV/AIDS, United States Agency for International Development (USAID), for her technical assistance in the conceptualization and design of the activity, as well as her contributions to the analysis and her review of the final report

- Lisa Parker, MEASURE Evaluation/Palladium Senior Technical Advisor, for her technical assistance in the conceptualization and design of the activity

CONTENTS

ABBREVIATIONS

DHS	Demographic and Health Surveys
OVC	orphans and vulnerable children
PEPFAR	President's Emergency Plan for AIDS Relief
USAID	U.S. Agency for International Development

EXECUTIVE SUMMARY

Knowing how many orphans and vulnerable children (OVC) may require services can enhance the efficiency of programming. The OVC technical working group of the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) requested help from MEASURE Evaluation (funded by the U.S. Agency for International Development [USAID]) to develop procedures—preferably using published final reports of standardized surveys such as the Demographic and Health Surveys (DHS)—to estimate the sizes of OVC groups having high programmatic priority at the national and subnational levels.

An earlier study reported a tight linear fit between national adult HIV prevalence and the percentage of children living in a household with at least one HIV-positive adult. MEASURE Evaluation extended this analysis to all existing DHS data sets with HIV testing, to determine the feasibility of using regression modeling to estimate the size of two priority groups: (1) children living with at least one adult who is HIV-positive, and (2) orphans and coresident children living with at least one adult who is HIV-positive.

At the national level, we found reasonably tight linear relationships between HIV prevalence among adults and the proportion of children living with at least one HIV-positive adult and between adult HIV prevalence and the proportion of orphans and coresident children living with at least one HIV-positive adult. However, at the subnational level, owing to small sample sizes, we found greater variation at the same level of HIV prevalence in the proportion of children living with an HIV-positive adult. Although the slopes and intercepts of the national and subnational regression lines are almost the same, the confidence intervals for the subnational level estimates are very wide and many data points fall outside the prediction.

PEPFAR has given priority for OVC programs to 21 countries. Of these, 17 have a recent biomarker survey: Burundi, Cameroon, Cote d'Ivoire, Democratic Republic of Congo, Ethiopia, Haiti, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, Swaziland, Tanzania, Uganda, Zambia, and Zimbabwe. For these countries, we recommend using the results of our secondary DHS analyses for programmatic estimates. Our Excel workbook (*OVC estimates from DHS*, available here: https://www.measureevaluation.org/resources/files/ovc-estimates-from-dhs/at_download/file) includes point estimates and 95-percent confidence intervals at both the national and subnational levels of the proportions of children who live in a household with an HIV-positive adult and the proportion of children who are orphans or coresident children living in a household with an HIV-positive adult, and extrapolates the observed proportions to numbers of children.

At present, four countries in the PEPFAR OVC portfolio do not have a recent DHS survey with HIV testing: Botswana, Nigeria, South Africa, and South Sudan. For these countries, the linear regression is a reasonable way to estimate *at the national level* the proportion of children who live in a household with an HIV-positive adult and the proportion of children who are orphans or coresident children living in a household with an HIV-positive adult. We have created an Excel workbook to predict these two indicators at the national level and to extrapolate the predicted proportions to numbers of children (*OVC estimation calculator*, available here: https://www.measureevaluation.org/resources/files/ovc-estimation-calculator/at_download/file).

Although the calculator works for size estimation at the national level, we strongly recommend that it **not** be used at the subnational level, owing to low accuracy of the estimates.

BACKGROUND

PEPFAR's OVC programs address the needs of children who are most affected by HIV:

- Children orphaned by HIV
- Children living in a household with an HIV-positive person
- Children vulnerable to HIV or its socioeconomic effects in areas where the prevalence of HIV is high

Figure 1 presents an array of OVC classifications, many of which are overlapping.

Figure 1. OVC priority groups

Country programming is informed by the size of OVC populations at the national and subnational levels. PEPFAR's OVC technical working group requested the assistance of MEASURE Evaluation to develop procedures to estimate the sizes of as many priority groups as possible, preferably using published final reports of standardized population surveys such as the Demographic and Health Surveys (DHS).

A study published in 2015 reported a tight linear fit between national adult HIV prevalence and the percentage of children living in a household with at least one HIV-positive adult (Short & Goldberg, 2015). MEASURE Evaluation proposed extending this analysis to all existing DHS datasets that include HIV testing, to determine the feasibility of using regression modeling to estimate the size of the following vulnerable groups of children at the national and subnational levels: (1) children living with at least one HIV-positive adult, and (2) orphans and coresident children living with at least one HIV-positive adult.

METHODS

Objectives

The objectives were to estimate the size of the following OVC target populations at the national and subnational levels: (1) children living with at least one HIV-positive adult, and (2) orphans and coresident children living with at least one HIV-positive adult.

Definitions for the proportions of children:

Indicator	Numerator	Denominator
Proportion of children living with at least one HIV-positive adult	Number of children ages 0–17 years who live with at least one HIV-positive adult	Total number of children ages 0–17 years in the households tested for HIV
Proportion of children who are orphans or coresident children living with at least one HIV-positive adult	Number of orphans and children ages 0–17 years living with orphans in a household with at least one HIV-positive adult	Total number of children ages 0–17 years in the households tested for HIV

Data Sources

We obtained household rosters and HIV test results of 60 DHS surveys that included HIV testing. HIV prevalence data at national and subnational levels for all countries were obtained from the DHS Program STATcompiler (http://www.statcompiler.com/en/). Appendix 1 lists the data sets included in the analyses. They cover 34 countries, including non-PEPFAR countries (such as the Dominican Republic), with some countries contributing multiple surveys (for example, Lesotho, Rwanda, and Tanzania with three surveys each).[1]

Analysis and Findings

Secondary data analysis was conducted in Stata Version 14.2.

Children living with at least one HIV-positive adult

DHS household rosters include information on every person usually living in the household at the time of interview (*de jure* household members) or who slept in the household the night before the interview (*de facto* members). The household member file includes age and sex, and for children ages 17 years and younger, whether the member's biological mother and biological father were still alive. The HIV file includes test results for all persons ages 15 and older who were tested for HIV.

We merged household member files of all *de jure* members with the HIV files. This allowed us to determine which children lived in a household with at least one HIV-positive adult. We counted as "orphans" all children whose mother and/or father was declared to be dead (single and double orphans), as well as children

[1] We included all countries and all surveys with HIV biomarkers to increase the number of observations and make the analyses as robust as possible.

whose parents' status (living or dead) was declared to be unknown. We counted as "coresident" all children living with orphans (but whose own parents were both alive). Households with one or more *de jure* members ages 18 or older who tested positive for HIV were considered to be a household with an HIV-positive adult.

Each of the 60 surveys with biomarker data (HIV test results) was analyzed using survey design effects and the appropriate weights provided by DHS. We calculated point estimates and 95-percent confidence intervals for the following two indicators at the national and subnational levels:[2]

1) Percentage of children ages 0–17 years living with at least one HIV-positive adult

2) Percentage of orphans and coresident children ages 0–17 years living with at least one HIV-positive adult

Linear regressions

Following Short and Goldberg (2015), we conducted linear regression modeling to examine the relationship between adult HIV prevalence and (1) the proportion of children living with at least one HIV-positive adult, and (2) the proportion of orphans and coresident children living with at least one HIV-positive adult. We regressed point estimates for each of the above two indicators on adult HIV prevalence obtained from STATcompiler. Our Stata program generates predicted values, standard errors of the prediction, and confidence bands for both of the above populations for a given HIV prevalence, and plots the regression line with 99% confidence bands.[3] Regressions were run at the national and subnational levels.

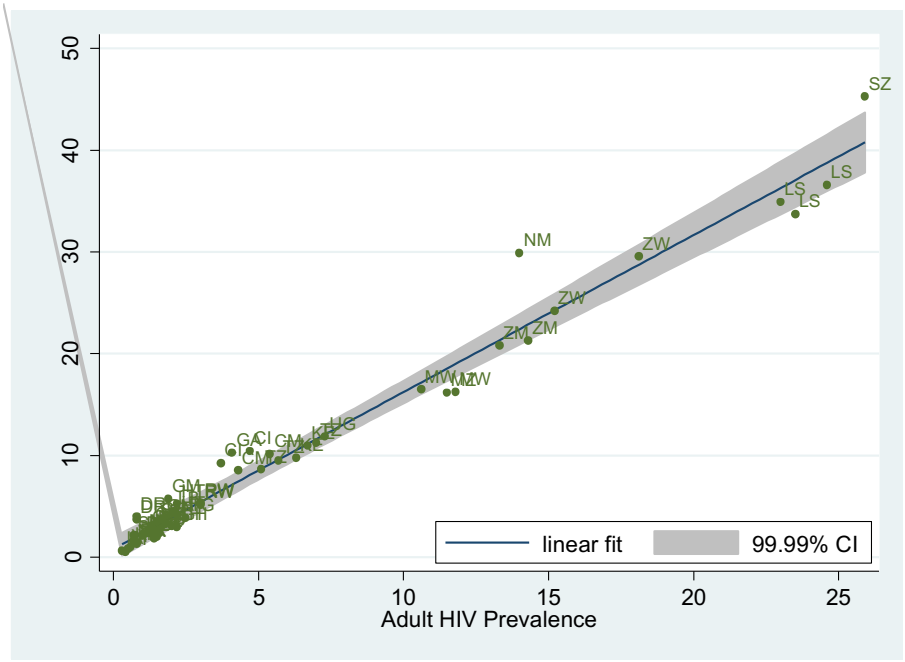
At the national level, we found a reasonably tight linear relationship between adult HIV prevalence and proportion of children living with at least one HIV-positive adult and between adult HIV prevalence and proportion of orphans and coresident children living with at least one HIV-positive adult.

Figure 2 presents the relationship between the proportion of children living with at least one HIV-positive adult and adult HIV prevalence for 60 DHS surveys with the HIV testing component.

[2] See Appendix for more information on the analysis (Stata_Program1.do and Stata_Program3.do).

[3] See Appendix 2 (Stata_Program2.do).

Figure 2. Proportion of children living with at least one HIV-positive adult, by national adult HIV prevalence for 60 DHS surveys with biomarkers



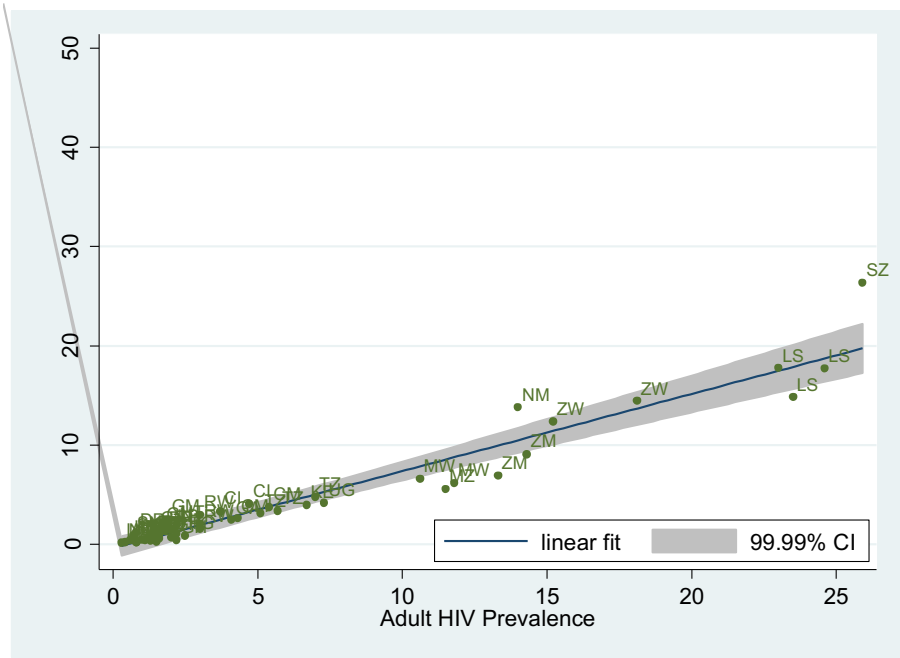
Source: DHS

A one-percent change in adult HIV prevalence is associated with a 1.54-percent change in the proportion of children living with at least one HIV-positive adult ($p<0.001$, $R^2 = 0.97$; see Appendix 2, Table 1).

Figure 3 presents the relationship between the proportion of orphans and coresident children living with at least one HIV-positive adult and adult HIV prevalence for 58 DHS surveys with the HIV testing component.[4]

[4] Two surveys did not include information on orphaning.

Figure 3. Proportion of children who are orphans or coresident children living with at least one HIV-positive adult, by national adult HIV prevalence for 58 DHS surveys[5] with biomarkers



Source: DHS

A one-percent change in adult HIV prevalence is associated with a 0.78-percent change in the proportion of orphans and coresident children living with at least one HIV-positive adult ($p<0.001$, $R^2 = 0.93$; see Appendix 2, Table 3).

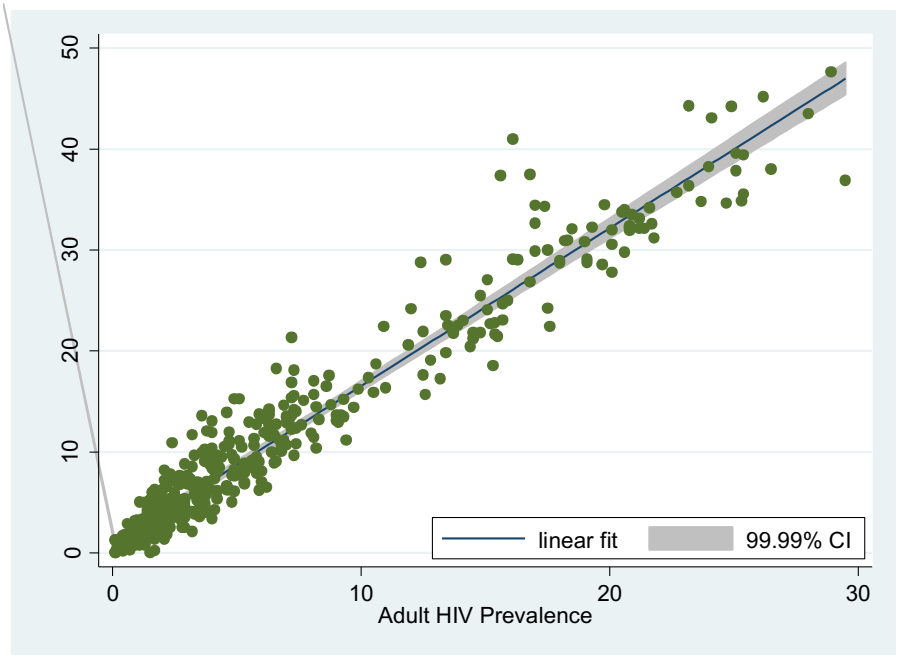
We then used the slope and intercept from the regression analyses to predict the child indicators from adult HIV prevalence and compared the predicted values with the observations from the DHS secondary analyses. Overall, roughly half of the predictions fell within the 99 percent confidence interval of the direct observations, and equal numbers of predictions fell above and below the confidence range.

We repeated the analyses at the subnational level for 53 surveys conducted in sub-Saharan Africa. Owing to small sample sizes, we found greater variation in the proportion of children living with an HIV-positive adult at the same level of adult HIV prevalence. While the slopes and intercepts of the national and subnational regression lines are almost the same, the confidence intervals for the subnational level estimates are very wide and many data points fall outside the prediction.

Figures 4 and 5 present the relationship between each of the two child indicators and adult HIV prevalence at the subnational level.

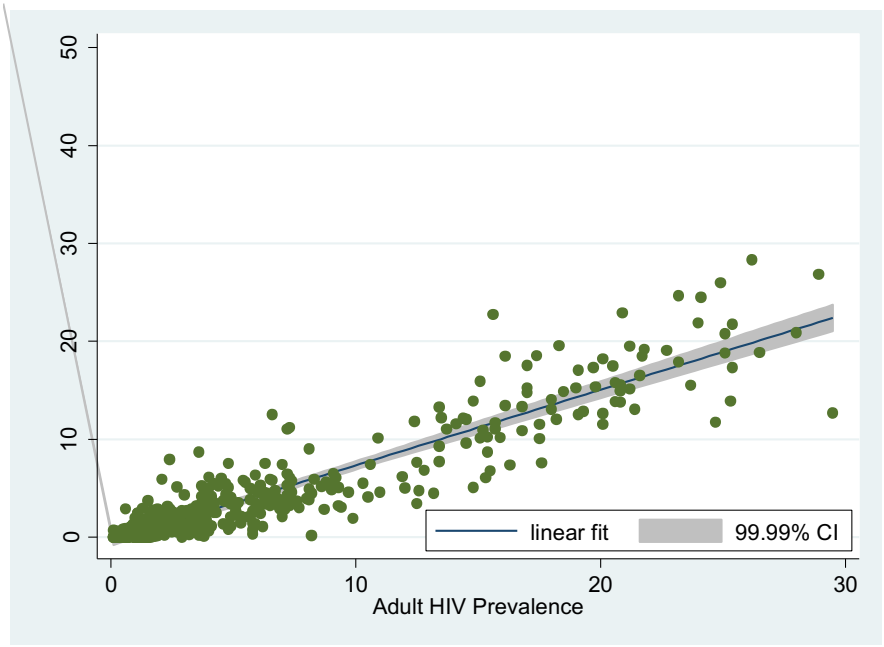
[5] Two of the HIV biomarker surveys did not include information on the orphan status of resident children.

Figure 4. Proportion of children living with at least one HIV-positive adult, by subnational adult HIV prevalence for 504 subnational areas in sub-Saharan Africa



Source: DHS

Figure 5. Proportion of children who are orphans or coresident children living with at least one HIV-positive adult, by subnational adult HIV prevalence for 496 subnational areas in sub-Saharan Africa



Source: DHS

RECOMMENDATIONS

For PEPFAR OVC countries that do not have a recent biomarker survey (DHS/AIDS Indicator Survey [AIS] with HIV testing), the linear regression provides a reasonable way to estimate *at the national level* the proportions of children who live in a household with an HIV-positive adult and the proportion of children who are orphans or coresident children living in a household with an HIV-positive adult. At present, four countries in the PEPFAR OVC portfolio do not have recent DHS surveys with HIV testing: Botswana, Nigeria, South Africa, and South Sudan. For these countries, we provide an Excel workbook to predict these two indicators at the national level and to extrapolate the predicted proportions to numbers of children *(OVC estimation calculator*, available here: https://www.measureevaluation.org/resources/files/ovc-estimation-calculator/at_download/file). We recommend that programs in Botswana, Nigeria, South Africa, and South Sudan use this calculator for size estimation at the national level. In addition, as the 2009 Kenya DHS survey did not collect information on orphans, the national calculator can be used to estimate the numbers of orphans and coresident children living with at least one HIV-positive adult in Kenya.

While the calculator works for size estimation at the national level, we strongly recommend that it **not** be used at the subnational level owing to low accuracy of the estimates.

We recommend using the results of our secondary DHS analyses for PEPFAR OVC countries that do have a recent biomarker survey: Burundi, Cameroon, Cote d'Ivoire, Democratic Republic of Congo, Ethiopia, Haiti, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, Swaziland, Tanzania, Uganda, Zambia, and Zimbabwe. For these countries, we provide an Excel workbook (*OVC estimates from DHS*, available here: https://www.measureevaluation.org/resources/files/ovc-estimates-from-dhs/at_download/file). The workbook gives point estimates and 95-percent confidence intervals at both the national and subnational levels of the proportion of children who live in a household with an HIV-positive adult and the proportion of children who are orphans or coresident children living in a household with an HIV-positive adult. It extrapolates the observed proportions to numbers of children.

REFERENCE

Short, S. E., & Goldberg, R. E. (2015). Children living with HIV-infected adults: Estimates for 23 countries in sub-Saharan Africa. *PLoS One*, 10(11), e0142580.

APPENDIX 1. DHS/AIS DATA SETS INCLUDED IN REGRESSION ANALYSES

Country	Year(s)	Country	Year(s)
Burkina Faso	2003, 2010	Liberia	2007, 2013
Burundi	2010	Malawi	2004, 2010
Cambodia	2005	Mali	2006, 2013
Cameroon	2004, 2011	Mozambique	2009
Chad	2015	Namibia	2013
Côte d'Ivoire	2005, 2012	Niger	2006, 2012
Democratic Republic of Congo	2007, 2014	Rwanda	2005, 2010, 2015
Dominican Republic*	2007 (2), 2013 (2)	São Tomé and Príncipe	2009
Ethiopia	2005, 2011	Senegal	2005, 2011
Gabon	2012	Sierra Leone	2008, 2013
Gambia	2013	Swaziland	2007
Ghana	2003, 2014	Tanzania	2004, 2008, 2012
Guinea	2005, 2012	Togo	2014
Haiti*	2006, 2012	Uganda	2011
India*	2006	Vietnam*	2005**
Kenya	2003, 2009**	Zambia	2007, 2014
Lesotho	2004, 2009, 2014	Zimbabwe	2006, 2011

* Excluded from the analyses of subnational areas, which was restricted to sub-Saharan Africa

** Did not collect information on orphans (Kenya 2009) or counts of orphans in households with an HIV-positive adult were 0 owing to low HIV prevalence (<1%) and missing data on orphaning (>65%) (Vietnam 2005)

APPENDIX 2. STATA OUTPUTS AND PROGRAMS

Table 1 presents the STATA output with the results of regression analysis of children living in a household with at least one HIV-positive adult.

Table 1. STATA output with the results of regression analysis of children living in a household with at least one HIV-positive adult at the national level

Source	SS	df	MS	Number of obs	=	60
				F(1, 58)	=	2228.84
Model	6216.87388	1	6216.87388	Prob > F	=	0.0000
Residual	161.77831	58	2.7892812	R-squared	=	0.9746
				Adj R-squared	=	0.9742
Total	6378.65219	59	108.112749	Root MSE	=	1.6701

percent2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
AdultHIVPrevalence	1.542752	.0326781	47.21	0.000	1.47734	1.608164
_cons	.8016792	.2773095	2.89	0.005	.2465839	1.356774

Table 2 presents the STATA output with the results of regression analysis of children living in a household with at least one HIV-positive adult at the subnational level.

Table 2. STATA output with the results of regression analysis of children living in a household with at least one HIV-positive adult at the subnational level

Source	SS	df	MS	Number of obs	=	504
				F(1, 502)	=	9136.02
Model	53085.7418	1	53085.7418	Prob > F	=	0.0000
Residual	2916.92038	502	5.81059836	R-squared	=	0.9479
				Adj R-squared	=	0.9478
Total	56002.6621	503	111.3373	Root MSE	=	2.4105

percent	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
HIVprev	1.562114	.0163431	95.58	0.000	1.530005	1.594223
_cons	.9089405	.1431995	6.35	0.000	.6275963	1.190285

Table 3 presents the STATA output with the results of regression analysis of orphans and coresident children living in a household with at least one HIV-positive adult at the national level.

Table 3. STATA output with the results of regression analysis of orphans and coresident children living in a household with at least one HIV-positive adult at the national level

Source	SS	df	MS	Number of obs	=	58
				F(1, 56)	=	802.26
Model	1566.77762	1	1566.77762	Prob > F	=	0.0000
Residual	109.365182	56	1.95294968	R-squared	=	0.9348
				Adj R-squared	=	0.9336
Total	1676.14281	57	29.4060141	Root MSE	=	1.3975

percent2	Coef.	Std. Err.	t	P>|t|	[99.99% Conf.	Interval]
AdultHIVPrevalence	.7781564	.0274732	28.32	0.000	.6630511	.8932618
_cons	-.4204177	.2360273	-1.78	0.080	-1.40931	.5684745

Table 4 presents the STATA output with the results of regression analysis of orphans and coresident children living in a household with at least one HIV-positive adult at the subnational level.

Table 4. STATA output with the results of regression analysis of orphans and coresident children living in a household with at least one HIV-positive adult at the subnational level

Source	SS	df	MS	Number of obs	=	496
				F(1, 494)	=	3080.90
Model	12935.8457	1	12935.8457	Prob > F	=	0.0000
Residual	2074.16993	494	4.19872456	R-squared	=	0.8618
				Adj R-squared	=	0.8615
Total	15010.0156	495	30.3232639	Root MSE	=	2.0491

percent	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
HIVprev	.7729363	.0139253	55.51	0.000	.7455761	.8002964
_cons	-.4035773	.1224218	-3.30	0.001	-.6441088	-.1630457

Stata_Program 1

The Stata_Program1.do analyzes DHS household member files previously merged with the HIV files to classify the household as having or not having an HIV-positive adult. It calculates proportions for both groups of children at the national level. We had to run 10 of these 60 files separately using slightly modified code because standard errors were not produced without the modification to the survey command in Stata. The results of the analysis are directly output to the Excel file named "results."

```
version 14.2
capture log close
clear all
set more off,perm
```

```
cd C: /* Indicate the directory here */
log using Stata_Program1.log, replace
/*****************************************************************************/
/* What project this is: OVC size estimation activity to develop a calculator */
/*  to estimate the size of the following populations at national & sub-       */
/*  national levels: 1) children living with at least one HIV positive adult  */
/*                   2) orphans and co-resident children living with at least */
/*                      one HIV-positive adult.                                */
/*  Inputs to the calculator would be: (1) adult HIV prevalence from the DHS   */
/*  (2) number of children ages 0-17 from the census or population projection. */
/*  Inside the black box, the calculator would first estimate the % of         */
/*  children living with an HIV positive adult based on the slope and          */
/*  intercept of the regression line, and then multiply the % by the           */
/*  population of kids.                                                        */
/*  The analysis plan will have two phases.                                    */
/*  Phase 1:                                                                   */
/*  1.Start with the analysis at the national level for 60 surveys with HIV data */
/*  2.Validate the calculator with the OVC priority countries.                 */
/*    For each country, compare the actual % of kids living with an HIV        */
/*    positive adult against the prediction from the calculator.               */
/*  3.Share the results with USAID.                                            */
/*  Phase 2:                                                                   */
/*  4.If everyone is satisfied with the validation at the national level, do   */
/*    the same with sub-national levels—usually province or region. If that    */
/*    works out, we could try disaggregating the sub-national units by         */
/*    urban/rural. We will have the same analysis plan for orphans and         */
/*    co-resident children 0-17 years old living with at least one HIV+ adult. */
/* What this do-file does:                                                     */
/*  1) get 2009 Lesotho files and attempt to replicate PLOS One paper results */
/*     in Table 2: % children living with at least one HIV+ adult             */
/* Who wrote it: Paul Henry Brodish                                            */
/* When you wrote it: November 17, 2016                                        */
/*     NOTE: trying a different approach to coding                             */
/*     11/23- try a second country to verify the approach- Zambia gives        */
/*          similar slightly inflated results vs. the model paper              */
/*     11/28- add code to identify orphans and co-resident children            */
/*     11/29- add code to automate process of pulling in datasets and putting  */
/*          out results; clean up code if possible                             */
/*     11/30- define orphans to include and exclude the missing values on the  */
/*          orphan status variables; include an indicator on orphans and       */
/*          co-resident children living in tested households                   */
/*     12/5 - re-enter the datasets now that an Excel file was prepared of     */
/*          all countries and datasets by year, matching correct pr (hh        */
/*          member file to ar (HIV) file                                       */
/*     12/13- subtracted four files that did not yield CIs                     */
/*     12/14- replaced HH weight with male survey/HIV testing weight           */
/*     12/15- substituted HH weight for male survey/HIV testing weight for 6   */
/*          datasets that had male survey/HIV testing weight = 0 or missing    */
/*     01/06- combines ovc2.do with ovc3.do for a single program to generate   */
/*          national-level results for all 60 countries/surveys                */
/* Where it is: indicate the directory here                                    */
/*****************************************************************************/
```

```
***get individual-level HIV data, keep variables needed, rename them
**fix unique id on two hiv datasets
use ciar50fl.dta,clear
isid hivclust hivstruct hivnumb hivline
tostring hivstruct,gen(hivstructs)
tostring hivnumb,gen(hivnumbs)
gen hivnumbs2=hivstructs + hivnumbs
gen float hivnumbr=real(hivnumbs2)
drop hivstructs hivnumb hivnumbs hivnumbs2
rename hivnumbr hivnumb
save ciar50flr.dta,replace
order hivclust hivnumb hivline
sort hivclust hivnumb hivline
isid hivclust hivnumb hivline

use snar4afl.dta,clear
isid hivclust hivstruct hivnumb hivline
tostring hivstruct,gen(hivstructs)
tostring hivnumb,gen(hivnumbs)
gen hhs=hivstructs + hivnumbs
gen float hhr=real(hhs)
drop hivnumb
rename hhr hivnumb
save snar4aflr.dta,replace
order hivclust hivnumb hivline
sort hivclust hivnumb hivline
isid hivclust hivnumb hivline

local id=0
foreach file in  ///
  "bfar61fl.dta" "bfar41fl.dta" "buar61fl.dta" "khar51fl.dta" "cmar61fl.dta" ///
  "cmar42fl.dta" "tdar71fl.dta" "cdar61fl.dta" "cdar50fl.dta" "ciar61fl.dta" ///
  "ciar50flr.dta" "drar61fl.dta" "drar6afl.dta" "drar51fl.dta" "drar5afl.dta" ///
  "etar61fl.dta" "etar51fl.dta" "gaar60fl.dta" "gmar60fl.dta" "ghar71fl.dta" ///
  "ghar4afl.dta" "gnar61fl.dta" "gnar51fl.dta" "htar61fl.dta" "htar51fl.dta" ///
  "iaar51fl.dta" "kear51fl.dta" "kear42fl.dta" "lsar71fl.dta" "lsar61fl.dta" ///
  "lsar41fl.dta" "lbar6afl.dta" "lbar51fl.dta" "mwar61fl.dta" "mwar4afl.dta" ///
  "mlar6afl.dta" "mlar51fl.dta" "mzar51fl.dta" "nmar61fl.dta" "niar61fl.dta" ///
  "niar51fl.dta" "rwar71fl.dta" "rwar61fl.dta" "rwar51fl.dta" "star50fl.dta" ///
  "snar61fl.dta" "snar4aflr.dta" "slar61fl.dta" "slar51fl.dta" "szar51fl.dta" ///
  "tzar6afl.dta" "tzar51fl.dta" "tzar4afl.dta" "tgar61fl.dta" "ugar6afl.dta" ///
  "vnar51fl.dta" "zmar62fl.dta" "zmar51fl.dta" "zwar61fl.dta" "zwar51fl.dta" {
use `file', clear
keep hivclust hivnumb hivline hiv03
rename hivclust cluster
rename hivnumb  household
rename hivline  linenum
rename hiv03    hivresult
sort cluster household linenum
isid cluster household linenum
local ++id
```

```
save hivfile`id', replace
}
***get HH member file, keep variables needed, rename them
**unique identifiers for member file are hhid(householdid) hv001(cluster) hv002(hhnumber)
hvidx(linenumber)
** hhid is a combination of hv001 and hv002 so don't need it
**check on a few datasets that are not merging, adjust so they merge correctly to HIV data
use cipr50fl.dta,clear
duplicates report
duplicates list hv001 hv002 hvidx
*keep hv001 hv002 hvidx hhid
order hv001 hv002 hvidx
sort hv001 hv002 hvidx
unique hv001 hv002 hvidx

gen hv002r1=substr(hhid,8,5)
gen hv002r2=trim(hv002r1)

gen y = subinstr(hv002r2," ","",.)
gen float y2=real(y)
sort hv001 y2 hvidx
drop hv002 y hv002r1 hv002r2 //hv002r4
rename y2 hv002
order hv001 hv002 hvidx
sort hv001 hv002 hvidx
save cipr50flr.dta,replace
isid hv001 hv002 hvidx

/***NOTE: does not merge to HIV data, duplicates on identifiers*/
use snpr4hfl.dta,clear
duplicates report
duplicates list hv001 hv002 hvidx
*keep hv001 hv002 hvidx hhid
order hv001 hv002 hvidx
sort hv001 hv002 hvidx
unique hv001 hv002 hvidx

gen hv002r1=substr(hhid,8,5)
gen hv002r2=trim(hv002r1)

gen y = subinstr(hv002r2," ","",.)
gen float y2=real(y)
sort hv001 y2 hvidx
drop hv002 y hv002r1 hv002r2 //hv002r4
rename y2 hv002
order hv001 hv002 hvidx
sort hv001 hv002 hvidx
save snpr4hflr.dta,replace
isid hv001 hv002 hvidx

** NOTE: Also check India and Vietnam to see why Orphans=0
*India
```

```
use iapr52fl,clear
keep hv001 hv002 hvidx hv005 hv021 hv022 hv102 hv103 hv105 hv111 hv113
tab1 hv111 hv113,m // a lot of missing data --61%

*Vietnam
use vnpr53fl,clear
keep hv001 hv002 hvidx hv005 hv021 hv022 hv102 hv103 hv105 hv111 hv113
tab1 hv111 hv113,m // a lot of missing data --66%

use vnar51fl,clear
tab hiv03,m //only 9 out of 1,675 tested +

**fix India data psu
use iapr52fl,clear
drop hv021
rename sh021 hv021
save iapr52flr,replace

***fix 9 datasets where male survey/HIV testing weight is missing or 0 or 1
use drpr61fl.dta,clear
replace hv028=hv005
save drpr61flr.dta,replace

use drpr6afl.dta,clear
replace hv028=hv005
save drpr6aflr.dta,replace

use drpr52fl.dta,clear
replace hv028=hv005
save drpr52flr.dta,replace

use drpr5afl.dta,clear
replace hv028=hv005
save drpr5aflr.dta,replace

use mzpr51fl.dta,clear
replace hv028=hv005
save mzpr51flr.dta,replace

use tzpr6afl.dta,clear
replace hv028=hv005
save tzpr6aflr.dta,replace

use tzpr4afl.dta,clear
replace hv028=hv005
save tzpr4aflr.dta,replace

use ugpr6afl.dta,clear
replace hv028=hv005
save ugpr6aflr.dta,replace

use zmpr61fl.dta,clear
```

```
replace hv028=hv005
save zmpr61flr.dta,replace

local id=0
foreach file in  ///
  "bfpr62fl.dta" "bfpr44fl.dta" "bupr61fl.dta" "khpr51fl.dta" "cmpr60fl.dta" ///
  "cmpr45fl.dta" "tdpr71fl.dta" "cdpr61fl.dta" "cdpr50fl.dta" "cipr61fl.dta" ///
  "cipr50flr.dta" "drpr61flr.dta" "drpr6aflr.dta" "drpr52flr.dta" "drpr5aflr.dta" ///
  "etpr61fl.dta" "etpr51fl.dta" "gapr60fl.dta" "gmpr60fl.dta" "ghpr71fl.dta" ///
  "ghpr4bfl.dta" "gnpr61fl.dta" "gnpr53fl.dta" "htpr61fl.dta" "htpr52fl.dta" ///
  "iapr52flr.dta" "kepr52fl.dta" "kepr42fl.dta" "lspr71fl.dta" "lspr61fl.dta" ///
  "lspr41fl.dta" "lbpr6afl.dta" "lbpr51fl.dta" "mwpr61fl.dta" "mwpr4dfl.dta" ///
  "mlpr6hfl.dta" "mlpr53fl.dta" "mzpr51flr.dta" "nmpr61fl.dta" "nipr61fl.dta" ///
  "nipr51fl.dta" "rwpr70fl.dta" "rwpr61fl.dta" "rwpr53fl.dta" "stpr50fl.dta" ///
  "snpr61fl.dta" "snpr4hflr.dta" "slpr61fl.dta" "slpr51fl.dta" "szpr52fl.dta" ///
  "tzpr6aflr.dta" "tzpr51fl.dta" "tzpr4aflr.dta" "tgpr61fl.dta" "ugpr6aflr.dta" ///
  "vnpr53fl.dta" "zmpr61flr.dta" "zmpr51fl.dta" "zwpr62fl.dta" "zwpr52fl.dta" {
use `file', clear
keep hv001 hv002 hvidx hv005 hv021 hv022 hv028 hv102 hv103 hv105 hv111 hv113
rename hv001 cluster
rename hv002 household
rename hvidx linenum
rename hv005 hhweight
rename hv021 psu
rename hv022 strata
rename hv028 maleweight
rename hv102 dejurehhmember
rename hv103 defactohhmember
rename hv105 age
rename hv111 motheralive
rename hv113 fatheralive
sort cluster household linenum
order cluster household linenum
isid cluster household linenum
local ++id
save memberfile`id',replace
}
***merge HH member file with HIV file
forvalues i = 1/60 {
use memberfile`i',clear
merge 1:1 cluster household linenum using hivfile`i'

***create individual-level dummy indicators for adult hh member testing and adult hh member hiv status
gen tested=0
replace tested=1 if hivresult!=. & (age>=18 & age<.) & dejurehhmember==1 & dejurehhmember!=.

gen testedpos=0
replace testedpos=1 if hivresult==1 & (age>=18 & age<.) & dejurehhmember==1 & dejurehhmember!=.

***create individual-level dummies identifying kids
gen kid=0
replace kid=1 if age<=17
```

```
**add an indicator for orphans and co-resident children in household
**define orphans as including unknown status of mother or father
gen orphan=0
replace orphan=1 if kid==1 & (inlist(motheralive,0,8) | inlist(fatheralive,0,8))

***aggregate totals at HH level
bysort cluster household: egen numtested=total(tested)
bysort cluster household: egen numtestedpos=total(testedpos)
bysort cluster household: egen numorphans=total(orphan)
bysort cluster household: egen numkids=total(kid)
bysort cluster household: egen numkids2=total(kid) if numtestedpos>0
bysort cluster household: egen numkids3=total(kid) if numtestedpos>0 & numorphans>0

**replacing missings with 0
replace numkids2=0 if numkids2==.
replace numkids3=0 if numkids3==.

**create subpop indicator variable
gen hivtestedhh=0
replace hivtestedhh=1 if numtested>0

***collapse to a hh level dataset, keeping only the variables needed
collapse (mean) psu strata maleweight hivtestedhh numkids numkids2 numkids3,by(cluster household)
save hhfile`i' , replace
}

/* NOTE: These files problematic */
*use hhfile2,clear // no strata variable
*use hhfile9,clear // stratum with single sampling unit
*use hhfile13.dta,clear // POOR MATCHING, but using anyway
*use hhfile47.dta,clear // strata==0
*use hhfile53.dta,clear // strata==.
*use hhfile56.dta,clear // Vietnam

*****************************************************************************************************
*****************************************************************************************************
****Code for outputing to Excel
*** Set the Excel File and Spreadsheet *** (NOTE: user needs to create a file called results.xlsx in current
directory)
putexcel set "results.xlsx",sheet("Sheet1") modify

**NOTE: 10 datasets have single unit sampling strata or missing strata so commented out and run below
without strata statement
**NOTE: file 56 has omitted 106 strata because no subpop members--Vietnam
local id=1
foreach file in "hhfile1"  /*"hhfile2"*/  "hhfile3"  "hhfile4"  "hhfile5"  /*"hhfile6"*/ "hhfile7"  /*"hhfile8"
"hhfile9"*/ "hhfile10" ///
"hhfile11" "hhfile12" /*"hhfile13"*/ "hhfile14" "hhfile15" "hhfile16" "hhfile17" "hhfile18" "hhfile19"
"hhfile20" ///
"hhfile21" "hhfile22" "hhfile23" "hhfile24" "hhfile25" "hhfile26" "hhfile27" "hhfile28" "hhfile29" "hhfile30"
///
```

```
"hhfile31" "hhfile32" /*"hhfile33"*/ "hhfile34" "hhfile35" "hhfile36" "hhfile37" "hhfile38" "hhfile39"
"hhfile40" ///
"hhfile41" "hhfile42" "hhfile43" "hhfile44" /*"hhfile45"*/ "hhfile46" /*"hhfile47"*/ "hhfile48" "hhfile49"
"hhfile50" ///
"hhfile51" "hhfile52" /*"hhfile53"*/ "hhfile54" "hhfile55" "hhfile56" "hhfile57" "hhfile58" "hhfile59"
/*"hhfile60"*/ {
use `file', clear
***svyset the data and get percentage
svyset, clear
gen weight=(maleweight/1000000)
gen sumweight=sum(weight) //checking that sum of weights = N
svyset psu [pweight=weight],strata(strata)
*svydescribe
svy, subpop(if hivtestedhh==1): ratio numkids2/numkids
regsave,ci
*** Output to Columns ***
        putexcel C`id'=(round(_b[_ratio_1],.0001))
        putexcel D`id'=(round(ci_lower,.0001))
        putexcel E`id'=(round(ci_upper,.0001))

local ++id
}

local id=1
foreach file in "hhfile1"  /*"hhfile2"*/  "hhfile3"  "hhfile4"  "hhfile5"  /*"hhfile6"*/  "hhfile7"  /*"hhfile8"
"hhfile9"*/ "hhfile10" ///
"hhfile11" "hhfile12" /*"hhfile13"*/ "hhfile14" "hhfile15" "hhfile16" "hhfile17" "hhfile18" "hhfile19"
"hhfile20" ///
"hhfile21" "hhfile22" "hhfile23" "hhfile24" "hhfile25" "hhfile26" "hhfile27" "hhfile28" "hhfile29" "hhfile30"
///
"hhfile31" "hhfile32" /*"hhfile33"*/ "hhfile34" "hhfile35" "hhfile36" "hhfile37" "hhfile38" "hhfile39"
"hhfile40" ///
"hhfile41" "hhfile42" "hhfile43" "hhfile44" /*"hhfile45"*/ "hhfile46" /*"hhfile47"*/ "hhfile48" "hhfile49"
"hhfile50" ///
"hhfile51" "hhfile52" /*"hhfile53"*/ "hhfile54" "hhfile55" "hhfile56" "hhfile57" "hhfile58" "hhfile59"
/*"hhfile60"*/ {
use `file', clear
***svyset the data and get percentage
svyset, clear
gen weight=(maleweight/1000000)
gen sumweight=sum(weight) //checking that sum of weights = N
svyset psu [pweight=weight],strata(strata)
*svydescribe
svy, subpop(if hivtestedhh==1): ratio numkids3/numkids
regsave,ci
*** Output to Columns ***
        putexcel G`id'=(round(_b[_ratio_1],.0001))
        putexcel H`id'=(round(ci_lower,.0001))
        putexcel I`id'=(round(ci_upper,.0001))

local ++id
}
```

```
local id=1
foreach file in "hhfile1"  /*"hhfile2"*/  "hhfile3"  "hhfile4"  "hhfile5"  /*"hhfile6"*/  "hhfile7"  /*"hhfile8"
"hhfile9"*/ "hhfile10" ///
"hhfile11" "hhfile12" /*"hhfile13"*/ "hhfile14" "hhfile15" "hhfile16" "hhfile17" "hhfile18" "hhfile19"
"hhfile20" ///
"hhfile21" "hhfile22" "hhfile23" "hhfile24" "hhfile25" "hhfile26" "hhfile27" "hhfile28" "hhfile29" "hhfile30"
///
"hhfile31" "hhfile32" /*"hhfile33"*/ "hhfile34" "hhfile35" "hhfile36" "hhfile37" "hhfile38" "hhfile39"
"hhfile40" ///
"hhfile41" "hhfile42" "hhfile43" "hhfile44" /*"hhfile45"*/ "hhfile46" /*"hhfile47"*/ "hhfile48" "hhfile49"
"hhfile50" ///
"hhfile51" "hhfile52" /*"hhfile53"*/ "hhfile54" "hhfile55" "hhfile56" "hhfile57" "hhfile58" "hhfile59"
/*"hhfile60"*/ {
use `file', clear
***svyset the data and get percentage
svyset, clear
gen weight=(maleweight/1000000)
gen sumweight=sum(weight) //checking that sum of weights = N
svyset psu [pweight=weight],strata(strata)
*svydescribe
svy, subpop(if hivtestedhh==1): total numkids numkids2 numkids3
*ereturn list
*** Output to Columns ***
        putexcel A`id'=(round(_b[numkids]))
        putexcel B`id'=(round(_b[numkids2]))
        putexcel F`id'=(round(_b[numkids3]))
local ++id
}
*****************************************************************************************************
***What are these 10 files commented out above?: BF2003, CM2004, CD2014, CD2007, DR2013b, LB2007
ST2009, SN2005, TZ2004, ZW2006
***Running these 10 without the strata statement
putexcel set "results.xlsx",sheet("Sheet2") modify
local id=1
foreach file in "hhfile2" "hhfile6" "hhfile8" "hhfile9" "hhfile13" "hhfile33" "hhfile45" "hhfile47" "hhfile53"
"hhfile60" {
use `file', clear
***svyset the data and get percentage
svyset, clear
gen weight=(maleweight/1000000)
gen sumweight=sum(weight) //checking that sum of weights = N
svyset psu [pweight=weight] //,strata(strata)
*svydescribe
svy, subpop(if hivtestedhh==1): ratio numkids2/numkids
regsave,ci
*** Output to Columns ***
        putexcel C`id'=(round(_b[_ratio_1],.0001))
        putexcel D`id'=(round(ci_lower,.0001))
        putexcel E`id'=(round(ci_upper,.0001))

local ++id
```

```
}

local id=1
foreach file in "hhfile2" "hhfile6" "hhfile8" "hhfile9" "hhfile13" "hhfile33" "hhfile45" "hhfile47" "hhfile53"
"hhfile60" {
use `file', clear
***svyset the data and get percentage
svyset, clear
gen weight=(maleweight/1000000)
gen sumweight=sum(weight) //checking that sum of weights = N
svyset psu [pweight=weight] //,strata(strata)
*svydescribe
svy, subpop(if hivtestedhh==1): ratio numkids3/numkids
regsave,ci
*** Output to Columns ***
        putexcel G`id'=(round(_b[_ratio_1],.0001))
        putexcel H`id'=(round(ci_lower,.0001))
        putexcel I`id'=(round(ci_upper,.0001))

local ++id
}

local id=1
foreach file in "hhfile2" "hhfile6" "hhfile8" "hhfile9" "hhfile13" "hhfile33" "hhfile45" "hhfile47" "hhfile53"
"hhfile60" {
use `file', clear
***svyset the data and get percentage
svyset, clear
gen weight=(maleweight/1000000)
gen sumweight=sum(weight) //checking that sum of weights = N
svyset psu [pweight=weight] //,strata(strata)
*svydescribe
svy, subpop(if hivtestedhh==1): total numkids numkids2 numkids3
*ereturn list
*** Output to Columns ***
        putexcel A`id'=(round(_b[numkids]))
        putexcel B`id'=(round(_b[numkids2]))
        putexcel F`id'=(round(_b[numkids3]))
local ++id
}
log close
exit
```

Stata_Program 2

The Stata_Program2.do file conducts linear regressions of the results of the Stata_Program_1 analysis with adult HIV prevalence taken from the DHS STATcompiler. It generates predicted values, standard errors of the prediction, and confidence bands for both groups of children living with an HIV-positive adult for a given HIV prevalence, and plots the regression line with 99-percent confidence bands.

```
version 14.2
capture log close
clear all
set more off,perm
cd C:\  /* Indicate the directory here   */
log using Stata_Program2.log, replace
/*******************************************************************************/
/* What project this is: OVC size estimation activity- develop a calculator    */
/*  to estimate the size of the following populations at national & sub-        */
/*  national levels: 1) children living with at least one HIV positive adult   */
/*                   2) orphans and co-resident children living with at least   */
/*                        one HIV-positive adult.                               */
/* What this do-file does:                                                     */
/*        1) Regresses hiv prevalence on percent of children living with at least */
/*           one HIV+ adult and saves predicted values and confidence bands    */
/*        2) Regresses hiv prevalence on percent of orphans and co-resident children */
/*           living with at least one HIV+ adult and saves predicted values and CIs */
/* Who wrote it: Paul Henry Brodish                                            */
/* When you wrote it: January 9, 2017                                          */
/* Where it is: indicate the directory here\                                   */
/*******************************************************************************/
***get data
import excel "C:\Users\brodish\Zulfiya\Results_Jan_06_17.xlsx", sheet("HIV+ HH") cellrange(A1:J61)
firstrow clear
gen percent2=percent*100  //percent actual
gen lb=lower_ci*100 //lower ci actual
gen ub=upper_ci*100 //upper ci actual

reg percent2 AdultHIVPrevalence
predict xb // predicted percentage from regression
gen diff=abs(percent2-xb) //absolute differnence in actual and predicted
predict syhat,stdp //standard error of prediction
gen lowerpred = xb - 1.96*syhat //lower ci for syhat
gen upperpred = xb + 1.96*syhat //upper ci for syhat
gen lowerpred2 = xb - 2.576*syhat //lower ci for syhat
gen upperpred2 = xb + 2.576*syhat //upper ci for syhat
save regressiondata_rev3,replace

set level 99.99
***NOTE: stdp is the default for the CI calculation for lfitci
graph twoway (lfitci percent2 AdultHIVPrevalence)  ///
        (scatter percent2 AdultHIVPrevalence, mlabel(Code) mlabp(1) msize(small) mlabsize(small) ), ///
        ytitle("% of all children ages 0-17") ///
        legend(ring(0) pos(5) order(2 "linear fit" 1 "99.99% CI"))

import excel "C:\Users\brodish\Zulfiya\ Results_Jan_06_17.xlsx", sheet("Orphans + unknowns HIV+
HH") cellrange(A1:J61) firstrow clear
drop if Country=="Kenya" & Year==2009
drop if Country=="Vietnam"

gen percent2=percent*100  //percent actual
gen lb=lower_ci*100 //lower ci actual
```

```
gen ub=upper_ci*100 //upper ci actual
reg percent2 AdultHIVPrevalence
predict xb // predicted percentage from regression
gen diff=abs(percent2-xb) //absolute differnence in actual and predicted
predict syhat,stdp //standard error of prediction
gen lowerpred = xb - 1.96*syhat //lower ci for syhat
gen upperpred = xb + 1.96*syhat //upper ci for syhat
gen lowerpred2 = xb - 2.576*syhat //lower ci for syhat
gen upperpred2 = xb + 2.576*syhat //upper ci for syhat
save regressiondata_rev4,replace

set level 99.99
***NOTE: stdp is the default for the CI calculation for lfitci
graph twoway (lfitci percent2 AdultHIVPrevalence) ///
        (scatter percent2 AdultHIVPrevalence, mlabel(Code) mlabp(1) msize(small) mlabsize(small) ), ///
        ytitle("% of all children ages 0-17") ///
        legend(ring(0) pos(5) order(2 "linear fit" 1 "99.99% CI"))

log close
exit
```

Stata_Program 3

The Stata_Program3.do file parallels the analysis of Stata Program 1. Using the DHS household member files merged with the HIV files, it calculates proportions for both groups of children at the subnational level. Owing to the complexity of accommodating varying numbers of subnational units across different countries, rather than directly outputting the subnational results to Excel, we instead computed regional estimates country by country within Stata and manually copied and pasted the output into Excel.

```
version 14.2
capture log close
clear all
set more off,perm
cd C:\Users\brodish\Zulfiya
log using Stata_Program3.log, replace
/*****************************************************************************/
/* What project this is: OVC size estimation activity- develop a calculator  */
/*  to estimate the size of the following populations at the national & sub-  */
/*  national level: 1) children living with at least one HIV positive adult  */
/*                  2) orphans and co-resident children living with at least  */
/*                     one HIV-positive adult.                                */
/*  Inputs to the calculator would be: (1) adult HIV prevalence from the DHS  */
/*  (2) number of children ages 0-17 from the census or population projection.*/
/*  Inside the black box, the calculator would first estimate the % of        */
/*  children living with an HIV positive adult based on the slope and         */
/*  intercept of the regression line, and then multiply the % by the          */
/*  population of kids.                                                       */
/*  The analysis plan will have two phases.                                   */
/*  Phase 1:                                                                  */
/*  1.Start with the analysis at the national level for 59 countries with HIV */
/*    data                                                                    */
```

```
/*  2.Validate the calculator with the OVC priority countries.                  */
/*    For each country, compare the actual % of kids living with an HIV         */
/*    positive adult against the prediction from the calculator.                */
/*  3.Share the results with USAID.                                             */
/*  Phase 2:                                                                    */
/*  4.If everyone is satisfied with the validation at the national level, do    */
/*    the same with sub-national levels—usually province or region. If that     */
/*    works out, we could try disaggregating the sub-national units by          */
/*    urban/rural. We will have the same analysis plan for orphans and          */
/*    co-resident children 0-17 years old living with at least one HIV+ adult.  */
/* What this do-file does:                                                      */
/*  1) get 2009 Lesotho files and attempt to replicate PLOS One paper results  */
/*     in Table 2: % children living with at least one HIV+ adult              */
/* Who wrote it: Paul Henry Brodish                                             */
/* When you wrote it: November 17, 2016                                         */
/*     NOTE: trying a different approach to coding                              */
/*     11/23- try a second country to verify the approach- Zambia gives         */
/*        similar slightly inflated results vs. the model paper                 */
/*     11/28- add code to identify orphans and co-resident children            */
/*     11/29- add code to automate process of pulling in datasets and putting  */
/*           out results; clean up code if possible                            */
/*     11/30- define orphans to include and exclude the missing values on the  */
/*           orphan status variables; include an indicator on orphans and      */
/*           co-resident children living in tested households                  */
/*     12/5 - re-enter the datasets now that an Excel file was prepared of     */
/*           all countries and datasets by year, matching correct pr (hh       */
/*           member file to ar (HIV) file                                      */
/*     12/13- subtracted four files that did not yield CIs                     */
/*     12/14- replaced HH weight with male survey/HIV testing weight           */
/*     12/15- substituted HH weight for male survey/HIV testing weight for 6   */
/*           datasets that had male survey/HIV testing weight = 0 or missing   */
/*     12/20- look at Lesotho 2009 subnational regions, then other countries   */
/* Where it is: C:\Users\brodish\Zulfiya\                                       */
/******************************************************************************/

***get individual-level HIV data, keep variables needed, rename them
**fix unique id on this hiv dataset
use ciar50fl.dta,clear
isid hivclust hivstruct hivnumb hivline
tostring hivstruct,gen(hivstructs)
tostring hivnumb,gen(hivnumbs)
gen hivnumbs2=hivstructs + hivnumbs
gen float hivnumbr=real(hivnumbs2)
drop hivstructs hivnumb hivnumbs hivnumbs2
rename hivnumbr hivnumb
save ciar50flr.dta,replace
order hivclust hivnumb hivline
sort hivclust hivnumb hivline
isid hivclust hivnumb hivline

use snar4afl.dta,clear
isid hivclust hivstruct hivnumb hivline
```

```
tostring hivstruct,gen(hivstructs)
tostring hivnumb,gen(hivnumbs)
gen hhs=hivstructs + hivnumbs
gen float hhr=real(hhs)
drop hivnumb
rename hhr hivnumb
save snar4aflr.dta,replace
order hivclust hivnumb hivline
sort hivclust hivnumb hivline
isid hivclust hivnumb hivline

local id=0
foreach file in ///
"bfar61fl.dta"  "bfar41fl.dta" "buar61fl.dta" "cmar61fl.dta" "cmar42fl.dta"  ///
"tdar71fl.dta"  "cdar61fl.dta" "cdar50fl.dta" "ciar61fl.dta" "ciar50flr.dta" ///
"etar61fl.dta"  "etar51fl.dta" "gaar60fl.dta" "gmar60fl.dta" "ghar71fl.dta"  ///
"ghar4afl.dta"  "gnar61fl.dta" "gnar51fl.dta" "htar61fl.dta" "htar51fl.dta"  ///
"kear51fl.dta"  "kear42fl.dta" "lsar71fl.dta" "lsar61fl.dta" "lsar41fl.dta"  ///
"lbar6afl.dta"  "lbar51fl.dta" "mwar61fl.dta" "mwar4afl.dta" "mlar6afl.dta"  ///
"mlar51fl.dta"  "mzar51fl.dta" "nmar61fl.dta" "niar61fl.dta" "niar51fl.dta"  ///
"rwar71fl.dta"  "rwar61fl.dta" "rwar51fl.dta" "star50fl.dta" "snar61fl.dta"  ///
"snar4aflr.dta" "slar61fl.dta" "slar51fl.dta" "szar51fl.dta" "tzar6afl.dta"  ///
"tzar51fl.dta"  "tzar4afl.dta" "tgar61fl.dta" "ugar6afl.dta" "zmar62fl.dta" ///
"zmar51fl.dta"  "zwar61fl.dta" "zwar51fl.dta" {

use `file', clear
keep hivclust hivnumb hivline hiv03
rename hivclust cluster
rename hivnumb  household
rename hivline  linenum
rename hiv03    hivresult
sort cluster household linenum
isid cluster household linenum
local ++id
save hivfile`id', replace
}

***get HH member file, keep variables needed, rename them
**unique identifiers for member file are hhid(householdid) hv001(cluster) hv002(hhnumber)
hvidx(linenumber)
** hhid is a combination of hv001 and hv002 so don't need it
**check on a few datasets that are not merging, adjust so they merge correctly to HIV data

**check on a few datasets that are not merging, adjust so they merge correctly to HIV data
use cipr50fl.dta,clear
duplicates report
duplicates list hv001 hv002 hvidx
*keep hv001 hv002 hvidx hhid
order hv001 hv002 hvidx
sort hv001 hv002 hvidx
unique hv001 hv002 hvidx
```

```
gen hv002r1=substr(hhid,8,5)
gen hv002r2=trim(hv002r1)

gen y = subinstr(hv002r2," ","",.)
gen float y2=real(y)
sort hv001 y2 hvidx
drop hv002 y hv002r1 hv002r2 //hv002r4
rename y2 hv002
order hv001 hv002 hvidx
sort hv001 hv002 hvidx
save cipr50flr.dta,replace
isid hv001 hv002 hvidx

/***NOTE: does not merge to HIV data, duplicates on identifiers*/
use snpr4hfl.dta,clear
duplicates report
duplicates list hv001 hv002 hvidx
*keep hv001 hv002 hvidx hhid
order hv001 hv002 hvidx
sort hv001 hv002 hvidx
unique hv001 hv002 hvidx

gen hv002r1=substr(hhid,8,5)
gen hv002r2=trim(hv002r1)

gen y = subinstr(hv002r2," ","",.)
gen float y2=real(y)
sort hv001 y2 hvidx
drop hv002 y hv002r1 hv002r2 //hv002r4
rename y2 hv002
order hv001 hv002 hvidx
sort hv001 hv002 hvidx
save snpr4hflr.dta,replace
isid hv001 hv002 hvidx

***fix 9 datasets where male survey/HIV testing weight is missing or 0 or 1
use mzpr51fl.dta,clear
replace hv028=hv005
save mzpr51flr.dta,replace

use tzpr6afl.dta,clear
replace hv028=hv005
save tzpr6aflr.dta,replace

use tzpr4afl.dta,clear
replace hv028=hv005
save tzpr4aflr.dta,replace

use ugpr6afl.dta,clear
replace hv028=hv005
save ugpr6aflr.dta,replace
```

```
use zmpr61fl.dta,clear
replace hv028=hv005
save zmpr61flr.dta,replace

local id=0
foreach file in  ///
"bfpr62fl.dta"  "bfpr44fl.dta"  "bupr61fl.dta" "cmpr60fl.dta"  "cmpr45fl.dta"  ///
"tdpr71fl.dta"  "cdpr61fl.dta"  "cdpr50fl.dta" "cipr61fl.dta"  "cipr50flr.dta" ///
"etpr61fl.dta"  "etpr51fl.dta"  "gapr60fl.dta" "gmpr60fl.dta"  "ghpr71fl.dta"  ///
"ghpr4bfl.dta"  "gnpr61fl.dta"  "gnpr53fl.dta" "htpr61fl.dta"  "htpr52fl.dta"  ///
"kepr52fl.dta"  "kepr42fl.dta"  "lspr71fl.dta" "lspr61fl.dta"  "lspr41fl.dta"  ///
"lbpr6afl.dta"  "lbpr51fl.dta"  "mwpr61fl.dta" "mwpr4dfl.dta"  "mlpr6hfl.dta"  ///
"mlpr53fl.dta"  "mzpr51flr.dta" "nmpr61fl.dta" "nipr61fl.dta"  "nipr51fl.dta"  ///
"rwpr70fl.dta"  "rwpr61fl.dta"  "rwpr53fl.dta" "stpr50fl.dta"  "snpr61fl.dta"  ///
"snpr4hflr.dta" "slpr61fl.dta"  "slpr51fl.dta" "szpr52fl.dta"  "tzpr6aflr.dta" ///
"tzpr51fl.dta"  "tzpr4aflr.dta" "tgpr61fl.dta" "ugpr6aflr.dta" "zmpr61flr.dta" ///
"zmpr51fl.dta"  "zwpr62fl.dta"  "zwpr52fl.dta" {
use `file', clear
keep hv001 hv002 hvidx hv005 hv021 hv022 hv024 hv028 hv102 hv103 hv105 hv111 hv113
rename hv001 cluster
rename hv002 household
rename hvidx linenum
rename hv005 hhweight
rename hv021 psu
rename hv022 strata
rename hv024 region
rename hv028 maleweight
rename hv102 dejurehhmember
rename hv103 defactohhmember
rename hv105 age
rename hv111 motheralive
rename hv113 fatheralive
sort cluster household linenum
order cluster household linenum
isid cluster household linenum
local ++id
save memberfile`id',replace
}
***merge HH member file with HIV file
forvalues i = 1/53 {
use memberfile`i',clear
merge 1:1 cluster household linenum using hivfile`i'

***create individual-level dummy indicators for adult hh member testing and adult hh member hiv status
gen tested=0
replace tested=1 if hivresult!=. & (age>=18 & age<.) & dejurehhmember==1 & dejurehhmember!=.

gen testedpos=0
replace testedpos=1 if hivresult==1 & (age>=18 & age<.) & dejurehhmember==1 & dejurehhmember!=.

***create individual-level dummies identifying kids
gen kid=0
```

```
replace kid=1 if age<=17

**add an indicator for orphans and co-resident children in household
**define orphans also as including unknown status of mother or father
gen orphan=0
replace orphan=1 if kid==1 & (inlist(motheralive,0,8) | inlist(fatheralive,0,8))

***aggregate totals at HH level
bysort cluster household: egen numtested=total(tested)
bysort cluster household: egen numtestedpos=total(testedpos)
bysort cluster household: egen numorphans=total(orphan)
bysort cluster household: egen numkids=total(kid)
bysort cluster household: egen numkids2=total(kid) if numtestedpos>0
bysort cluster household: egen numkids3=total(kid) if numtestedpos>0 & numorphans>0

**replacing missings with 0
replace numkids2=0 if numkids2==.
replace numkids3=0 if numkids3==.

**create subpop indicator variable
gen hivtestedhh=0
replace hivtestedhh=1 if numtested>0

***collapse to a hh level dataset, keeping only the variables needed
collapse (mean) psu strata region maleweight hivtestedhh numkids numkids2 numkids3,by(cluster household)
save hhfile`i', replace
}

use hhfileXX,clear //NOTE: Enter the hhfile number of interest from the above 53 files called into the
program
svyset, clear
gen weight=(maleweight/1000000)
gen sumweight=sum(weight) //checking that sum of weights = N
svyset psu [pweight=weight] ,strata(strata)
svy, subpop(if hivtestedhh==1): ratio numkids2/numkids,over(region) //NOTE: Enter the ratio of interest
here

log close
exit
```

Stata_Program 4

The Stata_Program4.do parallels the analysis of Stata Program 2, but at the subnational level. This file conducts linear regressions of the results of the Stata_Program_3 analysis with adult HIV prevalence taken from the DHS STATcompiler. It generates predicted values, standard errors of the prediction, and confidence bands for both groups of children living with an HIV-positive adult for a given HIV prevalence, and plots the regression line with 99-percent confidence bands.

```
version 14.2
capture log close
clear all
set more off,perm
```

```
cd C:\Users\brodish\Zulfiya
log using Stata_Program4.log, replace
/*****************************************************************************/
/* What project this is: OVC size estimation activity- develop a calculator  */
/*  to estimate the size of the following populations at national & sub-      */
/*  national levels: 1) children living with at least one HIV positive adult */
/*                   2) orphans and co-resident children living with at least */
/*                      one HIV-positive adult.                              */
/* What this do-file does:                                                   */
/*       1) Regresses hiv prevalence on percent of children living with at least */
/*          one HIV+ adult and saves predicted values and confidence bands   */
/*       2) Regresses hiv prevalence on percent of orphans and co-resident children */
/*          living with at least one HIV+ adult and saves predicted values and CIs */
/* Who wrote it: Paul Henry Brodish                                          */
/* When you wrote it: February 6, 2017                                       */
/* Where it is: C:\Users\brodish\Zulfiya\                                    */
/*****************************************************************************/
***get data
*import excel "C:\Users\brodish\Zulfiya\results.xlsx", sheet("HIV+ HH") cellrange(A1:J61) firstrow clear
import excel "C:\Users\brodish\Zulfiya\Results_proportion of children & orphans+co-
resident_children_with HIV_plus_adult_sub-national_Feb06_2017.xlsx", sheet("Regression1")
cellrange(F1:G505) firstrow clear
*import excel "C:\Users\brodish\Zulfiya\Results_proportion of children & orphans+co-
resident_children_with HIV_plus_adult_sub-national_Feb09_2017.xlsx", sheet("Regression2")
cellrange(C1:D497) firstrow clear
gen percent=Proportionofchildrenlivingwi*100  //percent actual
*gen percent=Proportionoforphansunknown*100 //percent actual
*gen lb=lower_ci*100 //lower ci actual
*gen ub=upper_ci*100 //upper ci actual

reg percent HIVprev

predict xb // predicted percentage from regression
gen diff=abs(percent-xb) //absolute differnence in actual and predicted
predict syhat,stdp //standard error of prediction
gen lowerpred = xb - 1.96*syhat //lower ci for syhat
gen upperpred = xb + 1.96*syhat //upper ci for syhat
gen lowerpred2 = xb - 2.576*syhat //lower ci for syhat
gen upperpred2 = xb + 2.576*syhat //upper ci for syhat
*save regressiondata_rev3,replace

set level 99.99
***NOTE: stdp is the default for the CI calculation for lfitci
graph twoway (lfitci percent HIVprev)  ///
        (scatter percent HIVprev, /*mlabel(Code) mlabp(1) msize(small) mlabsize(small)*/ ), ///
        /*ytitle("% children living w/ at least 1 HIV+ adult") */ ///
           ytitle("% orphans & cores. children living w/ at least 1 HIV+ adult") ///
legend(ring(0) pos(5) order(2 "linear fit" 1 "99.99% CI"))

log close
exit
```

MEASURE Evaluation
University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill, North Carolina 27517
Phone: +1-919-445-9350 • measure@unc.edu
www.measureevaluation.org

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. TR-17-154

USAID
FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE
Evaluation