



USAID
FROM THE AMERICAN PEOPLE

USAID POVERTY ASSESSMENT TOOLS (PAT) DATA ANALYSIS GUIDE

April 2013

This publication was produced for review by the United States Agency for International Development. It was prepared by FHI360.

DISCLAIMER

The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

Table of Contents

Introduction	2
PAT analysis: output and applications	3
Data analysis in Epi Info	3
Data analysis in CPro	4
Data analysis in other software	5
Overview: calculating poverty rates using PAT data.....	5
Detailed instructions for calculating poverty rates using PAT data.....	6
Reviewing Indicators.....	8
Quality Control.....	9

Introduction

This manual provides an overview of the 2 basic methods of analyzing PAT data:

1. Using the data entry template developed by the PAT team in Epi Info or CPro
2. Using the software of your choice (SPSS, STATA, Excel, Access, etc.).

The software typically used to analyze PAT data (developed by the PAT team in CPro or Epi Info) are user-friendly and allow for advanced analysis beyond basic PAT results (share of households above and below given poverty lines). However in recent years, a number of PAT users have requested information on analyzing PAT data in other software, often for data management purposes or to integrate PAT into existing, larger survey efforts (programmed in another statistical package). While certification documents for each PAT (found on each country's webpage) include most of the information needed, this Data Analysis Manual (and others to follow) provides detailed instructions on analyzing PAT data. Because development organizations use a variety of statistical software, analysis instructions are generic (rather than providing software-specific code) and presume that your PAT team has developed a method for entering the data into your system.

Please note that it is important to decide how you will enter AND analyze your data *before* collecting it so you can prepare the best format for your survey *and* data entry template. The majority of this document will be focused on providing instructions for analyzing data in your own software since CPro and Epi Info are covered in other materials.

PAT analysis: output and applications

An analysis of PAT data will tell you the percent of your target population living above or below one or more given poverty lines. For example, the poverty lines available for analysis in the Madagascar PAT are:

<i>Value: USD</i>	<i>Value: Local currency</i>	<i>Line description</i>
\$1.25 per day per capita	3060.695 Malagasy francs per day per capita (2001 prices)	International extreme poverty line (“dollar a day line”)
\$2.50 per day per capita	6121.39 Malagasy francs per day per capita (2001 prices)	International poverty line (“two dollars a day line”)

Poverty lines available for analysis are listed on each country’s PAT webpage. PAT data can be used for any number of purposes:

<i>Purpose</i>	<i>Question to answer</i>	<i>What the answer can help you understand</i>
Point-in-time	What is the poverty status of this group right now?	Are our services or products appropriate? Do outreach efforts align with goals?
Change over time	Has the poverty level of this group changed over [this time period]?	Are we seeing the effects we think we should see?
Comparison	How does the poverty level of this group compare to that group?	What do these different groups need? Can we explain the differences?
Data validation	How does information about poverty levels compare to what we’ve learned from other data?	Is the information consistent? Does it help us make decisions or understand the way things are?

Data analysis in Epi Info

Learning Epi

Use any of the following resources to learn how to analyze data in Epi Info.

USAID PAT Manual, Chapter 6 (English, Spanish & French)	povertytools.org/implementation.html
PAT Website	povertytools.org/EpiInfo.html
PAT Online Training, Section 9	povertytools.org/moodle

Important notes

1. Data Entry Templates for all PATs are available on the PAT website: www.povertytools.org
2. Epi Info is a free software available for download at: <http://wwwn.cdc.gov/epiinfo/>
 - a. DO NOT use Epi version 7. Click the “Downloads” button to download Version 3.5.4.
3. Data entered into Epi info *can* be exported into CSV (Excel) and other formats.
4. Data entered into Excel *can* be imported into Epi Info with some careful formatting. To do so, enter at least one survey into Epi using the provided data entry template and then export the data. You will see the column headings used in the exported data and can copy this format to your data. Note that some long variable names may be truncated (nownrefrigerat~); you will have to fix these truncations. To import data, you will have to first save it in a compatible

format. For example, select “Save as” and choose “Microsoft Excel 5.0 / 95”. When you Read/Import your data into Epi, select the appropriate data format (ie “Excel 5.0”).

5. Questions can easily be added to the data entry template if your PAT questionnaire is integrated into a longer survey. You will have to develop your own analysis programs for additional analysis. Questions added to the data entry template will NOT be included in the poverty calculation.

Data analysis in CPro

Learning CPro

Use any of the following resources to learn how to analyze data in CPro.

USAID PAT Manual, Chapter 6 (English, Spanish & French)	povertytools.org/implementation.html
PAT Website	povertytools.org/CPro.html
PAT Online Training, Section 9	povertytools.org/moodle

Important notes

1. Data Entry Templates for all PATs are available on the PAT website: www.povertytools.org
2. CPro is a free software available for download at: www.povertytools.org/CPro.html
 - a. DO NOT use CPro version 5.0 or higher. It is not compatible with PAT templates. Use version 4.1.002 available on the PAT website or here: csprouers.org/software
3. Data entered into CPro *can* be exported into CSV (Excel) and other formats.
4. The PAT team is NOT aware of a method for importing data into CPro from other applications (Excel, etc.).

Data analysis in other software

Analyzing PAT results in your own software (Excel, SPSS, etc.) is fairly straightforward assuming you know the basics of programming in that software. These instructions and the details that follow presume you have already entered your data using the software with which you will be analyzing it.

Overview: calculating poverty rates using PAT data

STEP 1: Review the list of PAT indicators

- Transform relevant PAT survey responses into their corresponding indicators (see indicator lists in each PAT Certification Document)

STEP 2: Analyze household-level data. For each household (survey)...

- Multiply each indicator by its corresponding coefficient. Include the listed intercept (value = 1 * its coefficient).
- Add these products together.
- Take the inverse log of that sum. This is the household's per capita consumption.
- Calculate whether that consumption is greater than (above) or less than (below) the value of the poverty line(s) in question.

STEP 3: Aggregate your results

- Calculate the total % of households living above and/or below the given poverty line(s).

Variable	Coef.
Intercept	6.0427
Household size	-0.3713
Household size squared	0.0177
Household head age	0.0178
Household head age squared	-0.0002
Household lives in Cochabamba	0.0020
Household lives in Oruro	0.0294
Household lives in Potosí	0.0126

Example of variable / coefficient listing in PAT certification document

Detailed instructions for each step are on the following pages.

IF YOU PREFER MATHEMATICAL TERMS...

1. For each Household (HH) j calculate the total consumption:

$$HH_j \text{ consumption} = e^{\sum_{i=1}^X \text{Coef}(i) * \text{Var}(i)}$$

2. Determine if that consumption puts them above or below a given poverty line (poverty line X)

- If (HH consumption) > (poverty line X), $HH_j = a$ ("above poverty line X ")
- If (HH consumption) < (poverty line X), $HH_j = b$ ("below poverty line X ")

3. Aggregate the results to determine the % of the survey population above/below the poverty line

- % of HHs above poverty line $X = n(a)/n$
- % of HHs below poverty line $X = n(b)/n$

Where

- $n(a)$ is the number of households assigned a ,
- $n(b)$ is the number of households assigned b , and
- n = the total number of households surveyed

Detailed instructions for calculating poverty rates using PAT data

STEP 1: Review the list of PAT indicators

PAT indicators are listed in each country's certification document, available on each country's page at povertytools.org. Determine which survey responses will need to be transformed into new indicators and create the appropriate variables. Here's an example (taken from the Madagascar PAT).

PAT survey question

7. What is the main construction material used for the roof of your dwelling?

- 1 Bark, leaves, twigs
- 2 Clay loam, cob
- 3 Wood (boards, plywood, hardboard)
- 4 Cement/concrete
- 5 Mat
- 6 No Roof
- 7 Other

Corresponding indicator (variable) – from PAT Certification document

PAT Indicator (variable)	Coef.
Roof is made of wood (boards, plywood, hardboard)	0.1548

The indicator tells us if the answer to question 7 was option 3 (wood) or not. The easiest way to do this is generally to create a new variable that has a value of 1 if the answer to the question is 3 and a value of 0 if the answer was something else (in this case: 1,2,4,5,6 or 7). Then multiply this new variable (0 or 1) by the coefficient.

For most PATs, this process will be very intuitive and simple. For other PATs, there may be a small number of indicators that are a bit more complex to construct. [See *Reviewing Indicators* on page 8 for more details on the types of indicators used in the PAT and how to construct them from PAT survey data.](#)

STEP 2: Analyze household-level data

For each household (survey)...

1. Multiply each PAT indicator (variable) by its corresponding coefficient.
 - a. $\text{Coef1} * \text{Var1}$, $\text{Coef2} * \text{Var2}$, $\text{Coef3} * \text{Var3}$, etc...
 - b. Notice that the list of PAT Indicators includes an intercept which is an essentially a variable with a value of 1 ($1 * 8.6952$ in the example below)

PAT Indicator (variable)	Coef.
Intercept	8.6952

2. Add each of these products together.
 - a. $\text{Coef1} * \text{Var1} + \text{Coef2} * \text{Var2} + \text{Coef3} * \text{Var3}$, etc...

- b. The sum of these parts equals the natural log of the household consumption.
 - i. $\text{Coef1*Var1} + \text{Coef2*Var2} + \text{Coef3*Var3} \dots = \ln(\text{hh consumption})$
3. Take the inverse log of this sum to get the household's estimated daily per capita consumption. Recall that PATs use consumption as a proxy for income.
 - a. $\text{HH consumption} = e^{\ln(\text{hh consumption})}$
 - b. In Excel, the function "EXP(x)" takes the inverse log of the inputted value (x)
4. Compare the consumption value to the relevant poverty line (is it above or below the line?).
 - a. In Excel you might use an equation like this:
 - i. $=\text{IF}(\text{AH2} \leq 1910.27, \text{"below line"}, \text{"above line"})$
 - ii. Cell AH2 contains the value of the HH consumption
 - iii. 1910.27 is the value of the poverty line in question in local currency
 - b. Values of poverty lines are found on each country's webpage listed in a table.
 - c. In most PAT certification documents, you will find 2 sets of indicators and coefficients. The indicators will be the same in each but the coefficients will be different. The first set of coefficients should be used for the lowest poverty lines (\$1.25 or median national) and the 2nd set should be used for the higher poverty lines (\$2.50 or national). For PATs with more than 2 poverty lines, use the 1st set of coefficients for all of the lowest values and the 2nd set for the highest value lines.
 - d. In some countries, poverty lines have different values for different regions. These will be listed in the certification documents.
5. Repeat this procedure for each survey.

IF YOU PREFER MATHEMATICAL TERMS...

(hh = household)

Calculate...

$\text{Coef1*Var1} + \text{Coef2*Var2} + \text{Coef3*Var3} + \text{Coef4x*Var4x} + \text{Coef4*Var4} + \text{Coef5*Var5} + \text{Coef6*Var6} + \text{Coef7*Var7} \dots = \ln(\text{hh consumption})$



HH consumption = $e^{\ln(\text{hh consumption})}$ ← The resulting value is a household's per capita consumption (depending on the PAT, this will be a monthly or daily level)



After HH consumption is calculated for a given household, it is compared to the value of the relevant poverty line (in local currency)

If HH consumption > value of poverty line, then HH = above poverty line

If HH consumption < value of poverty line, then HH = below poverty line

STEP 3: Aggregate your results

6. Count the number of households above and below the poverty line.
 - a. In Excel you might use an equation like this:
 - i. $=\text{(COUNTIF(A1:A1,"below line"})$

7. Calculate your results as a percentage.
 - a. In Excel you might use an equation like this:
 - i. $=\text{COUNTIF}(\text{A1:A1},\text{"below line"})/\text{COUNTA}(\text{A1:A1})$

IF YOU PREFER MATHEMATICAL TERMS...

$n(\text{above poverty line})/n\text{HH} = \text{total \% above poverty line}$

$n(\text{below poverty line})/n\text{HH} = \text{total \% below poverty line}$

Reviewing Indicators

Review each indicator (variable) in the table in the certification document against its corresponding survey question to determine what transformations are necessary. Be sure to read any relevant notes as well. Things to look for:

- For questions that ask how many of an item are owned, does the indicator ask for the number of items owned or just *if* the item is owned (yes = 1, no = 0)?
- For indicators derived from the household roster, are there age, gender or relational restrictions?
 - Such indicators might specify “excluding household head,” “...female household members,” “age 6 and older” etc.
- “Share of” indicators involve dividing the # of household members that meet the criteria by the number of household members.
 - Example: “Share of household members age 16 and older excluding household head who can read and write”
 - Indicator = (# of HH members age 16 and older excluding household head who can read and write) / (total # of HH members)
 - Note that the numerator DOES NOT include HH members that are younger than 16, nor does it include the head of household
 - Also note that the denominator DOES include every member of the household regardless of age or status (head). This might seem counter-intuitive but it is a convention used by PAT developers unless otherwise indicated.
- “Dependency ratio” is the sum of the # of HH members 14 or younger and the # of HH members 65 or older divided by the number of HH members between the ages of 15 and 64.

Use of household consumption data

PATs are designed to be accurate at an aggregate level which is why the instructions in this manual lead you through the process of combining results from all households to calculate the percent of households above and below poverty lines. When you analyze the data outside the provided templates, you will end up with household-level consumption calculations. Be very cautious with using this level of data. The PAT model uses an error balancing technique which maximizes overall accuracy by balancing misclassification errors. That is, for every household incorrectly calculated as being below a poverty line (when they should be above) there is another household incorrectly calculated as being above a poverty line (when they should be below it). This doesn't mean household consumption levels are entirely inaccurate, it just points to the fact that there are expected errors at the household level. Full details of accuracy and poverty balancing criteria are available in each PAT's certification document.

If you do want to use household-level consumption data, the following are suggested:

- Compare HH consumption information with other HH data. Does it tell a consistent story? When possible, triangulate the information, comparing PAT consumption calculations with at least 2 other pieces of information that point to household economic poverty. For example: participatory wealth ranking data, income estimates, asset indices, food security or dietary data, housing indices, etc.
- Do not make decisions that affect HH member lives based solely on PAT data (i.e. do not use PAT-calculated consumption information as the only criteria for household targeting).

Quality Control

Regardless of the manner in which you analyze your data you should always implement several layers of quality control to ensure the validity of your data. The USAID PAT Implementation Manual (available at povertytools.org/implementation.html) explains quality control measures that can be taken in the field while collecting and storing data. Here are several tips for maintaining quality control during data entry and analysis:

QC measure	What it does	How to do it
Double data entry & database comparison	Catches human error in the data entry process (ex: mis-typing or mis-reading data).	Have 2 data processors enter <i>each</i> completed survey into your database; compare their results.
Checking response frequencies	Finds unreasonable responses in the data due to interviewer mistakes or illegible writing (ex: a household head that is 3 years old; a dwelling with 12 rooms when every other dwelling in the region only has 1 or 2; a question response of option 5 when only values 1-4 are available).	For each data point, identify the full range of values entered across surveys and how often they occur; look for values that are out of place. Many software include frequency commands; histogram plots and max/min commands may also be useful.