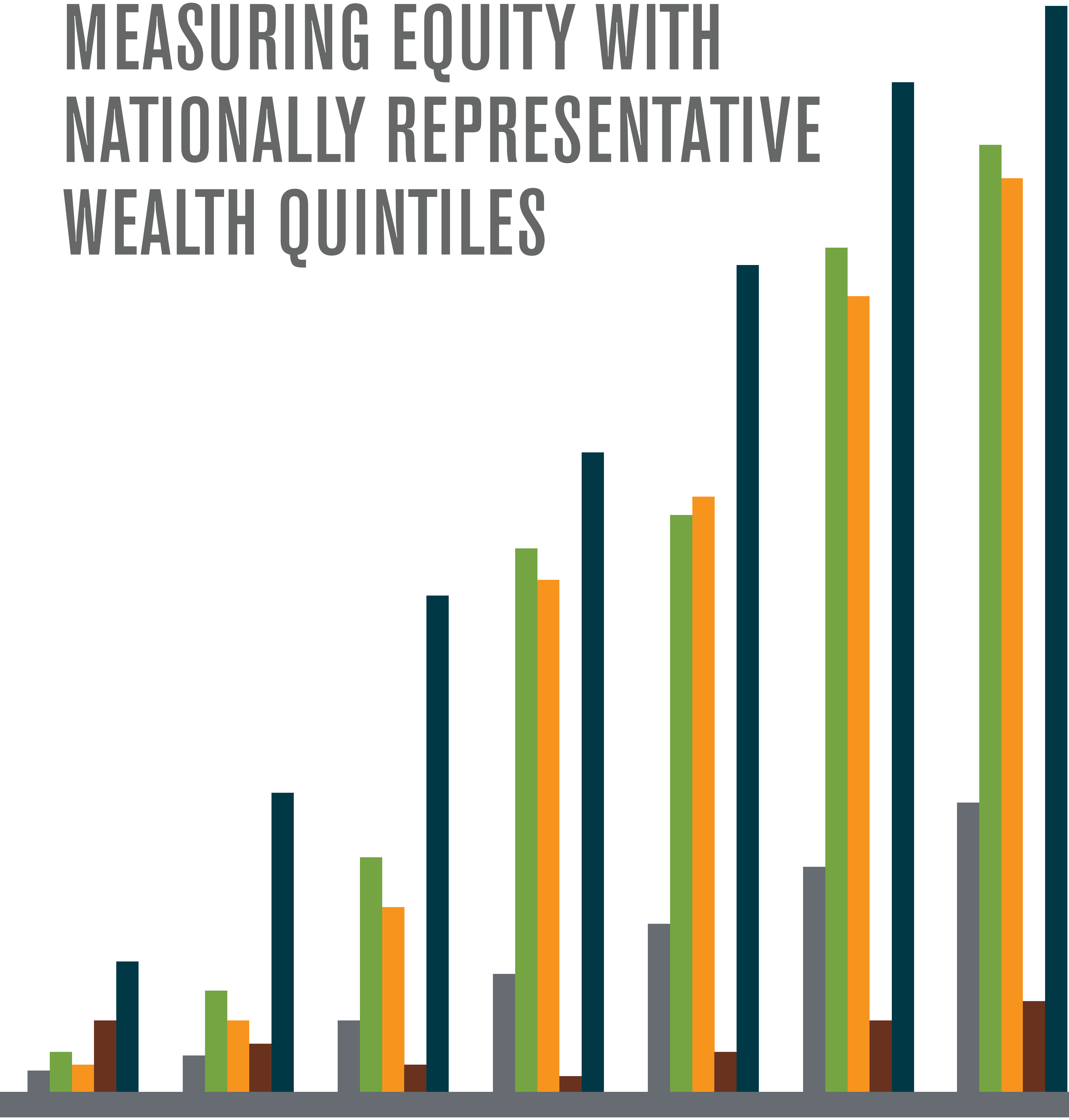
MEASURING EQUITY WITH NATIONALLY REPRESENTATIVE WEALTH QUINTILES



psi

Healthy lives. Measurable results.

This document may be freely reviewed, quoted, reproduced or translated, in part or in full, provided the source is acknowledged.

RECOMMENDED CITATION

Fry K., Firestone R., Chakraborty N.M. (2014).Measuring Equity with Nationally Representative Wealth Quintiles. Washington, DC: PSI.

ORDERING INFORMATION

This publication is available for electronic download at: http://www.psi.org/equity-wealth-quintile-guide.

PSI shares its technical guides with all interested individuals or organizations. Please note that technical guides are updated periodi¬cally based on the latest available epidemiological, demographic, intervention effectiveness and utilization data. As a result, numbers used in this document should be considered as illustrative only.

For more information about this technical guide contact:

Rebecca Firestone
1120 19th St., NW, Suite 600
Washington, DC 20036
rfirestone@psi.org

ACKNOWLEDGEMENTS

This technical guide was developed by Kenzo Fry, independent consultant, Rebecca Firestone, and Nirali Chakraborty. Special thanks to Sethson Kassenge for sharing data from Cote d'Ivoire and for reviewing the document. Megan Kays and Gary Mundy provided comments on earlier drafts as well.

POPULATION SERVICES INTERNATIONAL (PSI)
1120 19th Street, NW, Suite 600
Washington, DC 20036

TABLE OF CONTENTS

LEARNING OBJECTIVES

This guide covers the process of determining which national wealth quintiles each survey respondent falls into.

After going through this guide, the reader will understand:

- wealth indices and wealth quintiles
- how to create a wealth index for a national population using an existing national survey with Principal Components Analysis
- how to apply the national wealth index to survey respondents and determine which national wealth quintiles they fall into
- how to interpret and use wealth quintile data coming from the survey for equity monitoring and other purposes

The first section provides a conceptual overview of the wealth index and wealth quintiles. The second section is a detailed step-by-step guide for determining which wealth quintiles your survey respondents are in using either SPSS or STATA.

SECTION ONE: OVERVIEW

The type of poverty measured by the wealth index

Poverty is a state where one's material living standards are deemed to be inadequate. There are various ways of defining inadequate living standards which can be separated into two broad categories: absolute definitions of poverty and relative definitions of poverty.

Absolute poverty refers to living standards that do not meet specific minimum requirements, usually defined by a poverty line. For example, the World Bank definition of poverty is living on less than $1.25 per day. Absolute poverty lines are often based on having enough resources to meet a person's basic human needs, such as having enough food and shelter to live.

> **Why income is hard to measure**
>
> One of the most common ways to measure and compare living standards is through monetary income. However, income is deceptively hard to measure. Many people may be unwilling to truthfully report their incomes in survey settings, since they may be embarrassed to report how much money they have or have not received in a given time period. Respondents may also struggle to remember all of their possible sources of income, and in many settings, it may be challenging to quantify or turn into a monetary value some types of economic activity, such as agricultural labor.
>
> For these reasons, social scientists have investigated other ways of measuring living standards, especially in low and middle-income countries (Filmer and Pritchett 2001).

Relative poverty refers to living standards that are lower than those of other people in the population. For example, poverty could be defined as being one of the poorest 20% of people in the population, or in other words having material living standards that are lower than 80% of the people in the population. Thus relative poverty is dependent on which population the person lives in. The wealth index is one way of determining relative poverty.

How the wealth index measures living standards

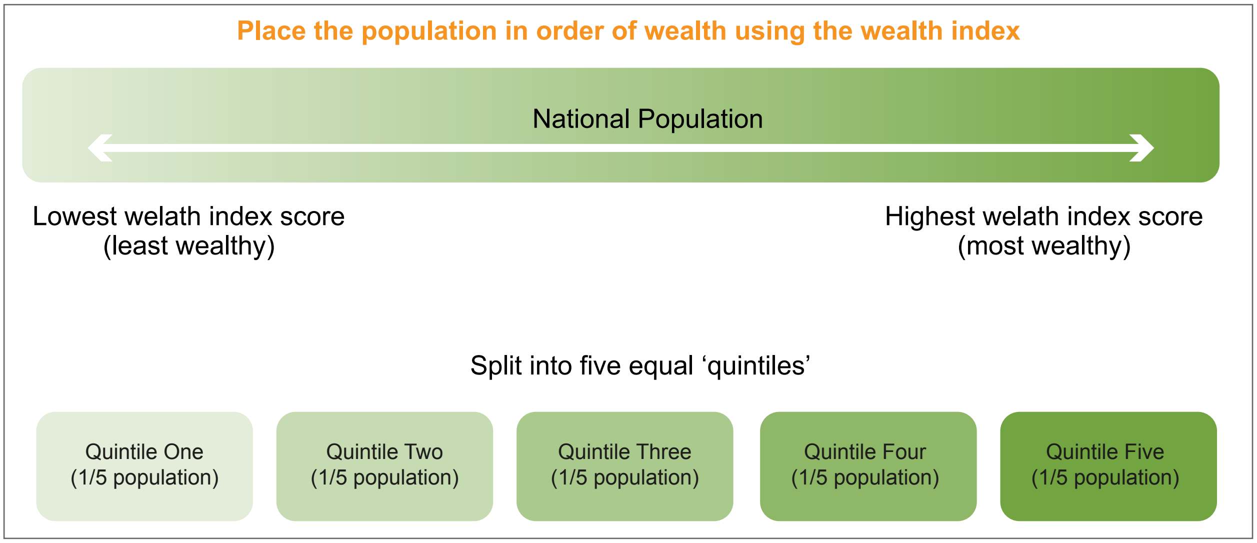
How we measure living standards is also variable. In some cases, poverty refers purely to monetary income or expenditure. Most poverty lines, like the $1.25 World Bank poverty line, are defined in these terms. However, household income is very difficult to capture properly. It can be volatile and some forms of wealth are not well represented by monetary income (see Box 1).

Other definitions of poverty are based on ownership of assets and the characteristics of the person's household. For example, someone with a piped water supply and a car would be wealthier than someone who has to retrieve water from a river and does not own a car. Household characteristics in many instances may be considered to be a better or more valid reflection of living standards than monetary income, since they capture long-term wealth and cover both monetary and non-monetary wealth (Rutstein & Johnston 2004). In some cases, particularly for household surveys, it may be both easier and more accurate to collect data on household assets than monetary income or expenditure. The wealth index is based on a set of household characteristics and asset ownership and was explicitly designed to overcome acknowledged challenges in measuring income (Filmer & Pritchett 2001).

There is no single household characteristic or asset that gives us enough information to determine whether someone is poor or not. Thus the wealth index is based on a variety of household characteristics and assets that are relevant for that country. For example, the wealth index in the 2011 Cote d'Ivoire Demographic and Health Survey included 26 variables covering items from household drinking water supply to ownership of a television. Individuals in the sample are then assigned a score based on how their households rank on ownership of assets and other household characteristics.

Wealth quintiles

One simple way of looking at relative poverty is to divide the population into equal quintiles. A quintile is a fifth (20%) of the population. For example, in a population of 500,000 people, each quintile would have 100,000 people in it. If we had wealth index scores for each of those 500,000 people, we could categorize the 100,000 with the lowest scores as the ‘poorest quintile’. The 100,000 with the highest scores would be categorized as the ‘wealthiest quintile’. A key reason for creating wealth quintiles is to look at how equitably other indicators are distributed by wealth status, for example male condom use, under-five child mortality, etc (Chakraborty, Firestone & Bellows 2013). A population-level estimate of an indicator such as skilled birth attendance may look positive, but if the majority of women had skilled attendants at birth were in higher wealth quintiles, it suggests inequity in access to skilled birth attendants (WHO 2014).



Principal Components Analysis

It can be difficult to summarize household wealth represented by each of the assets and household characteristics reported in a survey. One simple way could be to simply add up the number of assets each household owns. However, household characteristics like the type of water supply a household uses cannot be ‘added up’. Also, not all assets are a sign of wealth. For example in urban Rwanda, a person that owns a bicycle may be more likely to be relatively poor than relatively wealthy. A household in rural India may be considered relatively wealthy because the household keeps livestock, but livestock ownership would not be a good measure of wealth in urban India. Further complicating the matter is the fact that some household characteristics and assets are stronger indicators of relative wealth than others. Adding up each item assumes that each asset contributes equally to measuring the household’s wealth.

The solution to these challenges is Principal Components Analysis (PCA). PCA is a statistical method which determines the relative importance of each variable when seeking to summarize a set of variables (DeVellis 2012). When applied to asset and household characteristic data from a nationally representative survey, PCA can be used to create one summary measure of household wealth (Vyas & Kumaranayake 2006).

In a survey dataset with a set of variables that are correlated in complex and unknown ways along multiple dimensions, PCA is generally used to reduce those variables by assessing which variables behave in a similar manner. Based on the variables and their relationships to each other, PCA creates a new set of variables, each called a ‘principal component’. The first principal component accounts for the largest possible variance across the specified variables. The second principal component is not linearly correlated to the first principal component and accounts for as much of the remaining variance as possible. Each succeeding component accounts for as much of the remaining variance as possible and are not linearly correlated to any of the preceding variables.

For the wealth index, the first principal component from the PCA is assumed to represent relative wealth (Filmer and Pritchett). In other words, it is assumed that wealth is the factor that accounts for the largest amount of variance between households' assets and characteristics. Based on this first principal component, each variable is given a 'factor weight.' The factor weight represents the relative importance of each variable to the constitution of the first principal component.

Once the PCA has been done on the national survey dataset, we can extract the factor weights for each variable. We can also calculate wealth index scores based on these factor weights for each respondent in the national survey dataset. We can then separate the population into wealth quintiles based on the wealth index scores, and see the range of wealth index scores for each of the five quintiles.

Applying the national wealth index to a survey dataset

The wealth index and wealth quintiles described above are usually based on the national population and is created using a national survey dataset. Once we have a wealth index from the national population, we can then use it to figure out which national wealth quintiles our survey respondents fall into.

It is possible to calculate a wealth index and wealth quintiles from any quantitative survey. However, this method has several pitfalls if one is not careful about interpreting these quintiles in the population represented by the survey. Wealth quintiles are always a relative measure of how wealth is distributed within the population from the quintiles were calculated. Thus, wealth quintiles calculated for a representative sample of sex workers will represent how wealth was distributed across that population of sex workers. The same holds true for wealth quintiles calculated from a survey representative of one specific region of a country. The wealth quintiles will only represent the distribution of wealth in that geographic region.

However, if the survey dataset contains the same asset and household characteristic variables as the ones that the national wealth index is based on, we can calculate the national wealth index score for each survey respondent. The benefit of this is that we can quickly and easily calculate a nationally representative measure of wealth for a research study that is not necessarily nationally representative. The primary requirements are:

1. Asking the same questions in our survey as were asked in the nationally representative survey, and
2. Following the analysis steps laid out in this guide

For the analysis, we simply need to apply the factor weights extracted from the national survey dataset to our dataset. Then we can put each of the respondents into a national wealth quintile based on their wealth index score, because we know the range of wealth index scores for each national wealth quintile. The benefit this approach is that we are now measuring completely nationally representative wealth quintiles, and not just wealth quintiles that come from our original survey.

Summary

- The wealth index tells us about relative poverty
- An easy way to understand relative poverty is by dividing the population into wealth quintiles based on the wealth index
- The wealth index is based on asset ownership and household characteristics rather than monetary income
- We need a variety of asset and household characteristics to create a meaningful wealth index
- Principal Components Analysis turns the various asset and household characteristics into a wealth index
- Through PCA, each asset and household characteristic is given a factor weight and based on these each respondent in our survey can be given a wealth index score

SECTION TWO: THE ANALYSIS PROCESS STEP BY STEP

In this section, we will go through the steps needed to determine which national wealth quintiles survey respondents fall into. The steps can be grouped into four phases: preparation, survey, analysis of the national dataset, analysis of the survey dataset and interpreting the results.

Step 1: Preparation

1. Understand what data you need and why
2. Obtain a national survey dataset with relevant variables
3. Select variables

Step 2: Conduct the Survey

1. Create the questionnaire
2. Conduct survey

Step 3: Analysis of the national survey dataset

1. Prepare variables for Principal Components Analysis
2. Run the Principal Components Analysis
3. Assign wealth index scores to each respondent
4. Determine the range of wealth index scores for each quintile

Step 4: Analysis of the survey dataset

1. Standardize survey variables by national survey mean and standard deviation
2. Apply factor weights to survey dataset
3. Apply wealth quintile cut-offs

Step 5: Interpreting the results

1. Understanding and using the wealth quintiles

There are some circumstances in which you can skip some of these steps because someone else has done them for you:

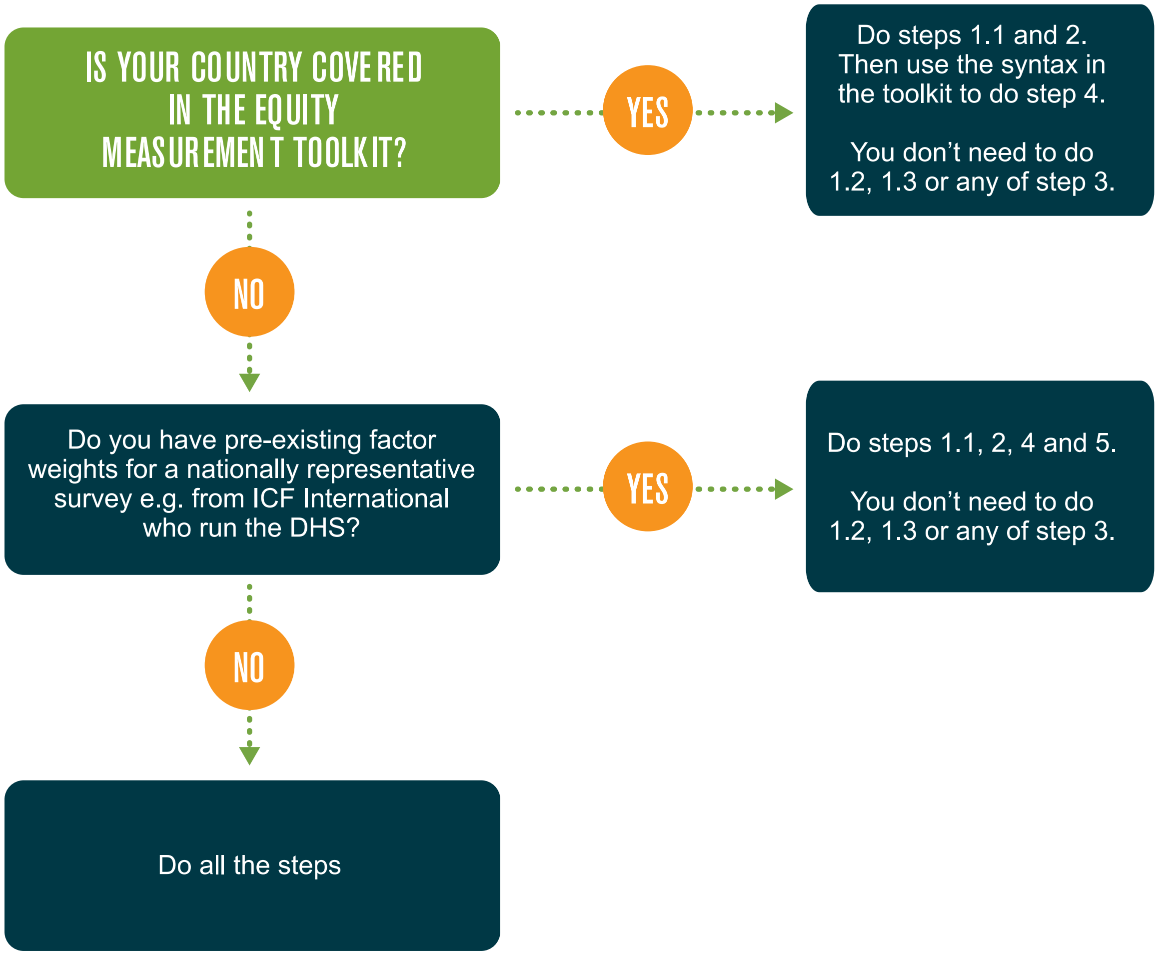
If you have access to pre-existing factor weights

For some countries, you may be able to obtain a pre-existing set of factor weights. If you already have a set of factor weights from a national survey, you will be able to skip Steps 1.2, 1.3 and Step 3. You should read Appendix 3: Using an existing set of factor weights.

If your country is covered by the Equity Measurement Toolkit

The Equity Measurement Toolkit developed by the Social Franchising Metrics Working Group contains pre-prepared questionnaires, data entry tools and analysis syntax for some countries. If your country is one of those covered by the toolkit and you decide to use it, you will be able to skip Steps 1.2 and 1.3 and Steps 3, because these steps have already been done for you. Also, the syntax needed to complete Step 4 is included with the toolkit. However, you should read Appendix 4: Using the equity measurement toolkit.

Steps in the Analysis Process



Step 1 – Preparation

1.1: Understand what data we need and why

Before starting our data collection and analysis, we will need to have a clear idea of what data we need and why. The following questions should be considered at this stage:

- Is the wealth index the appropriate measure of wealth for our needs?
  The wealth index is appropriate if you are interested in the relatively poor in your country but less relevant if you are interested in people living below specific minimum living standards.
- How will we use the wealth index in the final analysis?
  For example, you might want to see whether your target population is relatively poor, or you may like to compare wealth quintiles to each other to see how relative poverty affects your target population.

Clearly defining how you use will use the wealth index data after you have finished the survey early in the process will result in a successful and useful survey.

1.2: Obtain a national survey dataset with relevant variables

To create a national wealth index and national wealth quintiles, you will need a national survey dataset. This dataset should include the types of asset and household characteristics that are used in wealth indices (see 'Step 1.3: Selection of variables').

The dataset should not be too old, because the relationship between asset and household characteristics and relative wealth changes over time. For example, the relatively poor in most countries are much more likely to own a mobile phone now than 10 years ago. Ideally, the survey should have been done in the last 5 years. A survey more than 10 years old is almost certainly too old.

- Demographic and Health Surveys (DHS) are a good source of nationally-representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, health, and nutrition.
- The DHS Program also conducts other types of national surveys such as the Malaria Indicator Surveys and AIDS Indicator Surveys which usually include the required household characteristics and asset variables needed for creating wealth indices.
- Multiple Indicator Cluster Surveys (MICS) are similar to DHS, though they are not as standardised. These surveys are also a good source of data for creating a national wealth index. They contain standard questions about health, household characteristics and demographics.
- If there is no recent DHS or MICS, then you may be able to find another national survey with the required variables. For example, in Myanmar, a national survey focused on tuberculosis contained appropriate variables for constructing a wealth index.

Tip: Check with your national bureau of statistics to locate recent nationally representative surveys.

1.3: Select variables

The types of variables normally used to create a wealth index are ownership of various assets such as:

Vehicles	Electronics	Animals and equipment	Furniture
Bicycle	Television	Water pump	Table
Car	Mobile phone	Cattle	Chair
Motorcycle	Radio	Chickens	Clock

And household characteristics, such as:

Materials	Water and sanitation	Household members
Materials walls are made of	Water supply	Number of household members per bedroom
Material floor is made of	Type of toilet	Owns a bank account
Material roof is made of	Whether the toilet is shared	Owns agricultural land

Once we have identified appropriate variables, it is important to check that there is some variation within the population. If there are variables with no variance, for example 0% of people own a tractor or 100% of people do own a bicycle, then those variables will not be useful for the wealth index. They do not tell us anything about relative wealth because they are the same for everyone. Most large, nationally representative surveys will take as many possible indicators of wealth as are in the survey, as long as there is some variation in how the indicator is distributed in the sample.

For each variable, check the frequency to ensure there is some variation within the population using the tab command in STATA or the frequencies command in SPSS. We will use a DHS dataset from Cote D'Ivoire that was conducted in 2011 to illustrate the analysis process in Step 3. The survey questionnaire can be found here.

In this dataset, we can identify that the following list of variables will be useful for the wealth index (variable name is in parentheses):

Assets Owned By Household Or A Member Of The Household:

- Electricity (Hv206)
- Radio (Hv207)
- Television (Hv208)
- Mobile Telephone (Hv243a)
- Non-Mobile Telephone (Hv221)
- Refrigerator (Hv209)
- Vcd / Dvd Player (Sh110l)
- Washing Machine (Sh110i)
- Cooker (Sh110j)
- Computer (Sh110n)
- Internet At Home (Sh110)
- Watch/ Clock (Hv243b)
- Tractor (Sh118g)
- Bicycle (Hv210)
- Motorcycle/Scooter (Hv211)
- Animal-Drawn Cart (Hv243c)
- Car/Truck (Hv212)
- Boat With Motor (Hv243d)
- Agricultural Land (Hv244)

Household Characteristics:

- Water Source (Hv201)
- Toilet Type (Hv205)
- Cooking Fuel (Hv226)
- Material Of Floor (Hv213)
- Material Of Walls (Hv214)
- Material Of Roof (Hv215)
- Number Of Rooms Used For Sleeping (Hv216)

The asset ownership variables are all yes/no questions and should be binary. The household characteristic variables are all categorical variables where there are multiple response options, except for the last variable on rooms used for sleeping, which is a continuous variable.

Step 2 - Conduct The Survey

2.1: Create the questionnaire for the survey

In step 1.3 we identified the variables that would be used in the wealth index. We can now go back to the national survey questionnaire and identify the questions that correspond to those variables. We will need to add these questions to the survey. The wording and sequencing used for both the questions and answers in a questionnaire has an important influence on the data collected. Since the wealth index we are using is based on the national survey, the questions in the survey questionnaire must be the same as in the national survey questionnaire. We do not need to use all of the questions from the original questionnaire, just those that give us the variables that the wealth index is based on.

Fortunately, the questionnaires are provided with the report for each DHS and MICS. Here is an example of a question extracted from the Cote D'Ivoire DHS questionnaire and inserted to a PSI Cote D'Ivoire survey.

Q201	What is the main source of drinking water for members of your household? (*Select one answer)*	Piped Water Piped Into Dwelling 1 Piped To Yard/Plot 2 Public Tap/Standpipe 3 Tube Well Or Borehole 4 Dug Well Protected Well 5 Unprotected Well 6 Water From Spring Protected Spring 7 Unprotected Spring 8 Rainwater 9 Tanker Truck 10 Cart With Small Tank 11 Surface Water (River/Dam/Lake/ Pond/Stream/Canal/ Irrigation Channel) 12 Bottled Water 13 Other (Specify) 96

In this case, the response codes were simplified a little and numbered from 1 to 13. This is fine because the same question has been asked and the same response options are present.

2.2: Conduct the survey

The plans and sampling procedure for the survey will be as usual, based on your organization's standard guidance. We should make sure that the sample size and design are adequate to achieve what you need with the wealth index. Remember that if we want to analyze wealth quintiles separately or compare quintiles to each other, we are likely to require a larger sample than if we simply wanted to check which quintiles our target population falls into.

At this point, we can collect the data, enter it to a database, clean it and open it for analysis with either STATA or SPSS.

Step 3 - Analysis Of The National Survey Dataset

The STATA do-file and SPSS syntax for analyzing the Cote D'Ivoire 2011 DHS dataset can be accessed from PSI upon request.

3.1: Recode or restructure variables in preparation for the PCA

Binary variables

For binary variables, ensure that responses are coded as either 0 or 1. The convention from ICF International when creating wealth indices for the DHS is to recode 'don't know' answers, often coded as 9, 98, 99, 998 or 999, to zero.

An example from the Cote D'Ivoire DHS is variable hv210 about ownership of a bicycle:

Answer	Code	Frequency	Percent
No	0	5,541	57.21
Yes	1	4,129	42.63
Unknown	9	16	0.17
Total		9,686	100

The '9' answers should be recoded to '0'. We can use the following commands to first display the frequencies for the variable, then understand the codes that are used and then create new variable which is the same as the old one but with the '9' responses recoded to '0'. It is a good idea to retain the old variable hv210 without changing it in case we need to go back to it later.

STATA Commands

```
tab hv210
label list hv210

gen bicycle = hv210
recode bicycle (9=0)
```

SPSS Commands

```
frequencies hv210.

compute bicycle = hv210.
recode bicycle (9=0).
Execute.
```

Continuous variables

For continuous variables, recode any missing values and 'don't know' answers (often coded as 98 or 998 or something similar) to zero.

An example of a continuous variable from the Cote D'Ivoire dataset is hv216 – the number of rooms used for sleeping in the household. The responses were as follows (on next page):

	Frequency	Percent
0	5	0.05
1	3,681	38
2	2,809	29
3	1,710	17.65
4	726	7.5
5	299	3.09
6	120	1.24
7	72	0.74
8	38	0.39
9	25	0.26
10	15	0.15
11	3	0.03
12	5	0.05
13	2	0.02
22	1	0.01
99	175	1.81
Total	9,686	100

Once again, we need to recode the '99' answers, which are not labeled in the dataset but clearly mean 'don't know' rather than '99 rooms', to zero.

This is an imperfect solution for this variable, since usually households have at least 1 room in which people sleep. This value therefore indicates error either in data collection or data entry. The DHS generally recodes 'don't know' values into 'no' responses, so for consistency and simplicity it is the procedure we will use here.

STATA Commands
tab hv216 gen rooms = hv216 recode rooms (99=0)

SPSS Commands
frequencies hv216. compute rooms = hv216. recode rooms (99=0). Execute.

Categorical variables

We will need to convert categorical variables into multiple binary variables. The number value of the codes used in a categorical variable do not mean anything themselves. In the example below, we can see the response codes given for the question 'What is the main source of drinking water for members of your household?'

	Code	Frequency	Percent
Piped Water			
Piped Into Dwelling	11	923	9.53
Piped To Yard/Plot	12	1,957	20.2
Public Tap/Standpipe	13	1,616	16.68
Tube Well Water			
Tube Well Or Borehole	21	1,443	14.9
Dug Well (Open/Protected)			
Protected Well	31	1,661	17.15
Unprotected Well	32	1,202	12.41
Surface Water			
Protected Spring	41	47	0.49
Unprotected Spring	42	256	2.64
River/Dam/Lake/Ponds/Stream/Canal/ Irrigation Channel	43	461	4.76
Rainwater	51	0	0
Tanker Truck	61	2	0.02
Cart With Small Tank	62	3	0.03
Bottled Water	71	29	0.3
Other	96	79	0.82

There is no particular reason that the code for 'unprotected spring' is a much higher number (42) than for 'piped into dwelling' (11). For this reason, categorical variables will not work in a PCA. The solution to this problem is to turn this variable into a set of binary variables. We can make a binary variable that represents each of the options in the categorical variable. Instead of one water supply variable, we will now have 14 variables – one for each type of water supply.

If we have categories with very few respondents, we should merge that category in with another category that it is similar to or to an 'other category'. For example, in the Cote D'Ivoire dataset we can see that less than 1% of respondents gave the answers 'rainwater', 'tanker truck', 'cart with small tank' and 'bottled water', so we could recode all of these into the 'other' category.

In some cases, we could group more responses together. In this example, it may make sense to group together the options into the categories 'Piped Water', 'Tubewell or borehole', 'Dug well' and 'Surface water', since people within these groups have largely similar types of water supply and each group contains a reasonable proportion of the population. For water supplies and sanitation specifically, you can also categorize into whether the supply is 'improved' or 'not improved'. This also applies to sanitation types.

STATA Commands

```
tab hv201
label list HV201

gen wpiped = 0
replace wpiped = 1 if hv201 >= 11 & hv201 <= 13
label var wpiped 'Piped water supply'

gen wtube = 0
replace wtube = 1 if hv201 == 21
label var wtube 'Tubewell water supply'

gen wwell = 0
replace wwell = 1 if hv201 >= 31 & hv201<=32
label var wwell 'Dug well water supply'

gen wsurf = 0
replace wsurf = 1 if hv201 >= 41 & hv201<=43
label var wsurf 'Surface water supply'

gen woth = 0
replace woth = 1 if hv201 >50
label var woth 'Other water supply'
```

SPSS Commands

```
frequencies hv201.

compute wpiped = 0.
recode hv201 (11 thru 13=1) into wpiped.

compute wtube = 0.
recode hv201 (21=1) into wtube.

compute wwell = 0.
recode hv201 (31 thru 32=1) into wwell.

compute wsurf = 0.
recode hv201 (41 thru 43=1) into wsurf.

compute woth = 0.
recode hv201 (50 thru 96=1) into woth.
```

We will need to check every variable and prepare them in these ways before proceeding to the next step.

3.2: Run the Principal Components Analysis

Fortunately, we do not have to have an in-depth understanding of how to perform a PCA because both STATA and SPSS are able to run PCA for us. Now that we have recoded 'don't know' answers and converted categorical variables to multiple binary variables, we can run a PCA on them using the following commands:

STATA Commands
Here should we should run the pca command with a list of all the required variables. If the dataset requires a weight to be applied, we can do so by adding [aweight = name_of_weight_variable] to the end of the command.* Add the 'means' option at the end, so Stata will also give us the mean and standard deviation of each variable. pca electricity radio television mobile landline fridge dvd washer /// cooker computer internet watch tractor bicycle motorcycle cart car /// boat land wpiped wtube wwell wsurf woth tflush tlatimp tlatopen /// tnoneoth clpg cchar cwood cnone fnat ffinish fcement wallnat /// wallcem wallfinish rnat rmetal rfinished rooms [aweight = weight], means *Note that if working with a DHS dataset, you will first need to create a weight variable which is the hv005 variable divided by 100000. In Stata, use the command gen weight=hv005/100000. In SPSS use COMPUTE weight=HV005/100000. EXECUTE. WEIGHT BY weight.

SPSS Commands

FACTOR*
  /VARIABLES electricity radio television mobile landline fridge dvd washer cooker computer internet watch tractor bicycle motorcycle cart car
boat land wpiped wtube wwell wsurf woth tflush tlatimp tlatopen tnoneoth clpg cchar cwood cnone fnat ffinish fcement wallnat wallcem wallfinish rnat rmetal rfinished rooms
  /MISSING PAIRWISE
  /ANALYSIS electricity radio television mobile landline fridge dvd washer cooker computer internet watch tractor bicycle motorcycle cart car
boat land wpiped wtube wwell wsurf woth tflush tlatimp tlatopen tnoneoth clpg cchar cwood cnone fnat ffinish fcement wallnat wallcem wallfinish rnat rmetal rfinished rooms
/PRINT FSCORE
/FORMAT SORT
/PLOT EIGEN
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25)
  /ROTATION NOROTATENOR
  /SAVE REG(ALL)
/METHOD=CORRELATION.
 	
*Note that if working with a DHS dataset, you will first need to create a weight variable which is the hv005 variable divided by 100000 and then apply the weight in the subsequent analysis.
In SPSS use COMPUTE weight=HV005/100000. EXECUTE. WEIGHT BY weight.

In STATA, we will be presented with two tables. The first contains eigenvalues as well as showing us the proportion of variance accounted for by each component. We will not be using these numbers. The second table entitled Principal components (eigenvectors) presents us with the components for each variable. We will be using the very first component only (Comp1). The value given under Comp1 for each variable is the factor weight for that variable. STATA will also give us a table with the mean and standard deviation for each variable, because we added the option 'means'.

In SPSS, we will be presented with various tables. The one entitled Component Score Coefficient Matrix contains the factor weights under the Component1 column.

Here are the factor weights, means and standard deviations from the Cote D'Ivoire DHS dataset:

	Comp1	Mean	SD		Comp1	Mean	SD
electricity	0.2679	0.558208	0.496626	wwell	-0.1118	0.309354	0.462251
radio	0.0867	0.546097	0.497896	wsurf	-0.1012	0.082347	0.274906
television	0.2573	0.431628	0.495329	woth	0.0244	0.01197	0.108754
mobile	0.1427	0.808064	0.393844	tflush	0.2351	0.211495	0.40839
landline	0.1006	0.02045	0.141542	tlatimp	0.055	0.255834	0.436351
fridge	0.1907	0.11344	0.317146	tlatopen	-0.031	0.193881	0.395357
dvd	0.2185	0.258382	0.437768	tnoneoth	-0.2277	0.338791	0.473323
washer	0.0342	0.002918	0.053943	clpg	0.2216	0.149667	0.356763
cooker	0.212	0.132877	0.339459	cchar	0.1203	0.178077	0.382597
computer	0.1408	0.037403	0.189757	cwood	-0.2545	0.601801	0.489552
internet	0.1021	0.01518	0.122273	cnone	-0.0015	0.069361	0.25408
watch	0.1507	0.374273	0.48396	fnat	-0.2075	0.196407	0.397301
tractor	0.0018	0.000667	0.025817	ffinish	0.1767	0.10118	0.301582
bicycle	-0.1482	0.391902	0.4882	fcement	0.0622	0.69919	0.458634
motorcycle	0.0001	0.193441	0.395016	wallnat	-0.2473	0.319178	0.466182
cart	-0.0386	0.016149	0.126055	wallcem	0.1968	0.495465	0.500005
car	0.1058	0.0334	0.179689	wallfinish	-0.006	0.060278	0.238014
boat	0.0113	0.000404	0.020085	rnat	-0.1973	0.192132	0.393997
land	-0.1914	0.538652	0.49853	rmetal	0.1619	0.746273	0.435166
wpiped	0.225	0.476377	0.499467	rfinished	0.0303	0.061595	0.24043
wtube	-0.1094	0.119953	0.324923	rooms	-0.0032	2.07635	1.382244

These are the numbers we will need to use later when analyzing the survey. One easy way to keep track of these is select the cells of the table we need (the variable name column and the Component 1 column) in the STATA output, right click and select 'Copy Table'. We can then paste this into an Excel file. In SPSS, you can do the same by double clicking the Component Score Coefficient Matrix table and selecting the first two columns (variable name and Component 1) and copying and pasting to Excel.

3.3: Assign wealth index scores to each respondent

Each variable we included in the PCA now has a factor weight associated with it. At this point, we can assign a wealth index score to each respondent in the survey based on their asset ownership and household characteristics.

STATA Commands
predict wealthscore

This will create a variable called wealthscore which contains each respondent's wealth index score.

In SPSS, running the PCA automatically creates this variable and names it fac1_1.

3.4: Determine the range of wealth index scores for each quintile

Once each respondent's household has been given a wealth index score, we put all of them in order of wealth and separate them into quintiles. We can then see the range of wealth index scores that correspond to each quintile.

STATA Commands

```
xtile quintile=wealthscore [pweight=weight], nq(5)
```

This creates a variable called 'quintile' where each respondent is given a value from 1 to 5, where 1 is the poorest quintile and 5 is the wealthiest. We can then run the following commands to check the range of wealth index scores in each quintile.

```
summ wealthscore if quintile ==1
summ wealthscore if quintile ==2
summ wealthscore if quintile ==3
summ wealthscore if quintile ==4
summ wealthscore if quintile ==5
```

SPSS Commands

```
FREQUENCIES
 VARIABLES=fac1_1
 /NTILES= 5
 /ORDER= ANALYSIS /FORMAT NOTABLE.
```

In the Cote D'Ivoire dataset, the range of wealth index values corresponding to each wealth quintile was as follows:

Quintile	Wealth index score minimum	Wealth Index score maximum
Poorest	-5.09077	-2.74422
Poor	-2.7442	-1.12234
Medium	-1.12055	0.625245
Wealthy	0.62759	2.645443
Wealthiest	2.646179	8.686915

Step 4 - Analysis Of The Survey Dataset

The STATA do-file and SPSS syntax for analyzing a survey can be accessed from PSI upon request.

4.1: Prepare the survey variables for analysis

We now need to perform the same procedure for each variable as we did for the national survey in step 3.1. The binary variables must be coded to only 0 or 1. Missing or 'don't know' values in continuous variables should be recoded to 0. Categorical variables will need to be converted into binary variables in the same way we converted them for the national survey dataset in step 3.1.

STATA Commands

```
tab q201
label list HV201

gen wpiped = 0
replace wpipd = 1 if hv201 >= 1 & q201 <= 3
label var wpiped 'Piped water supply'

gen wtube = 0
replace wtube = 1 if q201 == 4
label var wtube 'Tubewell water supply'

gen wwell = 0
replace wwell = 1 if q201 >= 5 & q201<=6
label var wwell 'Dug well water supply'

gen wsurf = 0
replace wsurf = 1 if q201==7 & q201==8 | q201==12
label var wsurf 'Surface water supply'

gen woth = 0
replace woth = 1 if q201 ==9 | q201 ==10 |  q201 ==11 |   q201 ==96
label var woth 'Other water supply'
```

SPSS Commands

```
frequencies q201.

compute wpiped = 0.
recode wpipd ( 1 thru 3 = 1).

compute wtube = 0.
recode wtube(4=1).

compute wwell = 0.
recode wwell (5 thru 6 = 1).

compute wsurf = 0.
recode wsurf (7=1) (8=1) (12=1).

compute woth = 0.
recode woth (9=1) (10=1) (11=1) (96=1).
Execute.
```

4.2: Standardize survey variables by national survey means and standard deviations

It is important to standardize each variable used in wealth index calculations, so that they are on the same scale and can be compared. Some variables, such as owning a television will be either 0 (does not have) or 1 (does have). Other variables, like number of chickens the household owns, can range from 0 to over 100. Thus the variance in the latter will be much greater even though it is not necessarily a better predictor of wealth. Standardization of a variable here involves subtracting the mean of that variable from each value and then dividing it by the standard deviation. Once standardized, the variable has a mean of zero and a standard deviation of 1.

When standardizing the survey variables, we will use the mean and standard deviation from the national survey for that variable. This way, the variables are standardized against the original national survey dataset, and the factor weights from the PCA conducted on the national survey can be applied to the dataset.

The mean for the electricity variable was 0.05582208 and the standard deviation was 0.496626. The following command will generate a standardized version of electricity called electricity2. You should repeat this procedure for each variable.

STATA Commands
gen electricity2 = (electricity - 0.558208)/0.496626

SPSS Commands
compute electricity2 = (electricity - 0.558208)/0.496626.

4.3: Apply factor weights

To calculate wealth index scores for each survey respondent, we simply need to multiply the standardized variable by the factor weight, and then add all the values together. The factor weight for electricity was 0.2679. The following commands show how to create a variable called electricity3 which is the standardized variable electricity2 multiplied by the factor weight. You should repeat this for each variable, and then create a variable called wealthscore which adds them all together.

STATA Commands

```
gen electricity3 = electricity2* 0.2679

gen wealthscore = electricity3+radio3+television3+mobile3+landline3+fridge3+dvd3+ ///
washer3+cooker3+computer3+internet3+watch3+tractor3+bicycle3+ ///
motorcycle3+cart3+car3+boat3+land3+wpiped3+wtube3+wwell3+wsurf3+ ///
woth3+tflush3+tlatimp3+tlatopen3+tnoneoth3+clpg3+cchar3+cwood3+cnone3+ ///
fnat3+ffinish3+fcement3+wallnat3+wallcem3+wallfinish3+rnat3+rmetal3+ ///
rfinished3+rooms3
```

SPSS Commands

```
compute electricity3 = electricity2* 0.2679.

compute wealthscore = electricity3+radio3+television3+mobile3+landline3+fridge3+dvd3+
washer3+cooker3+computer3+internet3+watch3+tractor3+bicycle3+
motorcycle3+cart3+car3+boat3+land3+wpiped3+wtube3+wwell3+wsurf3+
woth3+tflush3+tlatimp3+tlatopen3+tnoneoth3+clpg3+cchar3+cwood3+cnone3+
fnat3+ffinish3+fcement3+wallnat3+wallcem3+wallfinish3+rnat3+rmetal3+
rfinished3+rooms3.
```

4.4: Apply wealth quintile cut-offs

Now each respondent has a wealth index score, we can determine which wealth quintile they are in. We will use the range of wealth index scores corresponding to each wealth quintile that we identified in step 3.4. The following commands will create a variable called quintile with each respondent assigned to a quintile.

STATA Commands
generate QUINTILE = . replace QUINTILE = 1 if wealthscore<-2.74422 replace QUINTILE = 2 if wealthscore<-1.12234 & wealthscore>=-2.74422 replace QUINTILE = 3 if wealthscore<0.625245 & wealthscore>=-1.12234 replace QUINTILE = 4 if wealthscore<2.645443 & wealthscore>=0.625245 replace QUINTILE = 5 if wealthscore>=2.645443

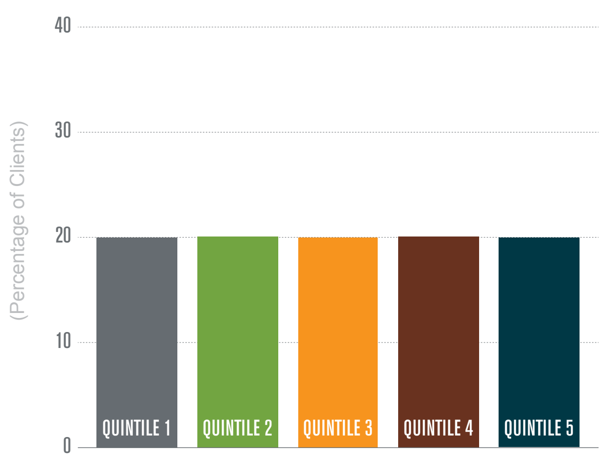
SPSS Commands
RECODE wealthscore (LOWEST THRU -2.74422 = 1) (-2.74422 THRU -1.12234 = 2) (-1.12234 THRU 0.625245 = 3) (0.625245 THRU 2.645443 = 4) (2.645443 THRU HIGHEST = 5) INTO QUINTILE. Execute.

Step 5 - Interpreting The Results

5.1: Using and understanding the wealth quintiles

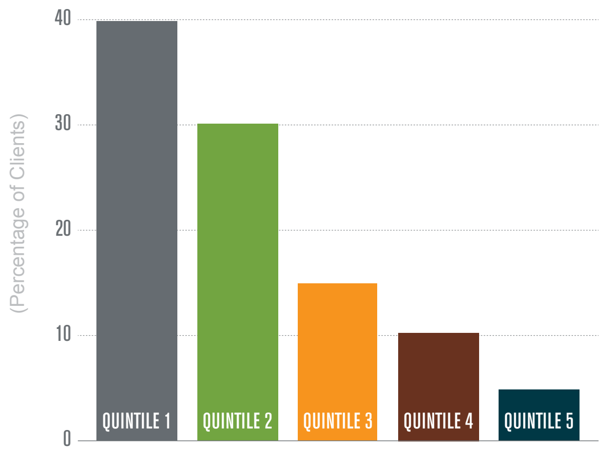
At this point we have successfully assigned a national wealth quintile to each of our survey respondents. We can now use the tab command in STATA or frequencies command in SPSS to see the percentage of the target population that falls into each national wealth quintile.

If the target population is no different to the national population, then 20% would fall into each national wealth quintile as in this graph:



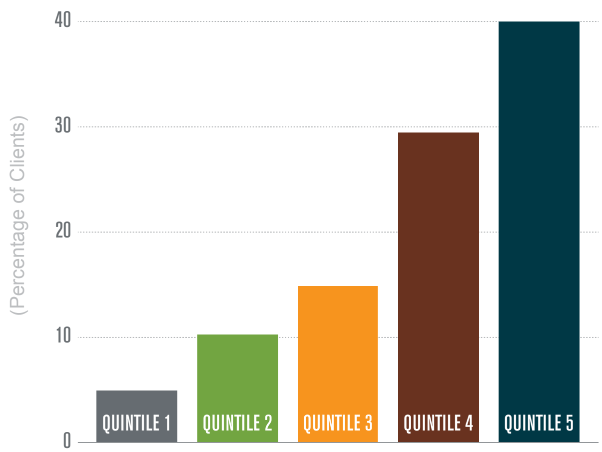
This target population is similar to the general population

If there is a disproportionately large percentage (>20%) in the lower quintiles and low percentages (<20%) in the higher quintiles, then the target population is poor relative to the rest of the population, as in this graph:



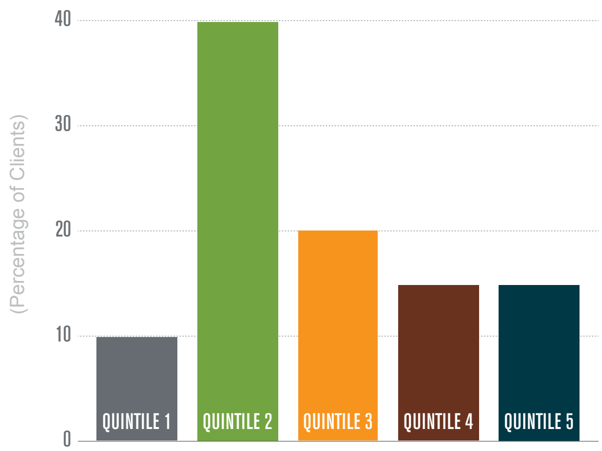
This target population is relatively poor compared to the general population

If the opposite is true, then the target population is relatively wealthy, as in this graph:



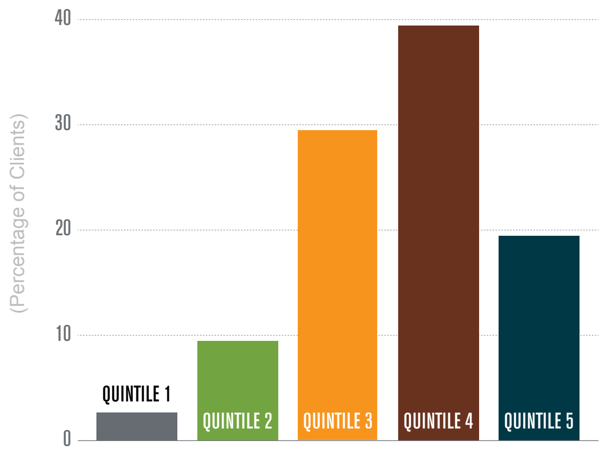
This target population is relatively wealthy compared to the general population

In some cases the distribution will be more mixed, as in this graph:



This target population is poor relative to the general population, but not the poorest of the poor

Here is the result from our Cote D'Ivoire analysis:



The Cote D'Ivoire survey respondents tended to be relatively wealthy. The bottom two quintiles were underrepresented and the middle and fourth (wealthy but not wealthiest) quintiles were overrepresented.

We can also look at how wealth affects other variables. For example, we could assess if contraceptive use differs between wealth quintiles in the target population. How to do this is left up to the data analyst and the specific research objectives of the study. Wealth quintiles can be used as explanatory variables in a bivariate or multivariate analysis. These can also be used to adjust for socio-economic status in other analyses where wealth is not specifically of interest.

Nationally representative wealth quintiles, calculated following the steps provided in this guide, ultimately give researchers and program practitioners a powerful but simple to conduct tool to gain more insight into the relative wealth or poverty of the populations they are trying to serve. An important application is in equity monitoring for health and other social development programs (Chakraborty, Firestone & Bellows 2013).

APPENDICES

Appendix A: STATA – advanced do-file writing

If you are an advanced user of STATA, you may be able to write more concise do-files using more sophisticated commands than those shown in the guide above.

STATA is able to store values in its memory from one dataset and use them when analyzing another dataset. This is useful here because we can store the factor weights from the national survey dataset and use them on the survey dataset without having to copy them to Excel or write them down. We can also store the means and standard deviations of the wealth index variables and the range of wealth index scores for each wealth quintile. Familiarity with STATA's language for writing macros is required.

Appendix B: Resources and further reading

For more on types of poverty measurement and the World Bank poverty lines:

Haughton, J. & Shandker, S.R. (2009). Handbook on Poverty and Inequality. Washington, D.C.: The World Bank. Retrieved from: http://siteresources.worldbank.org.

For information on why the wealth index is often preferred over expenditures and incomes:

Rutstein, S.O. & Johnson, K. (2004). The DHS Wealth Index. DHS Comparative Reports No. 6. Calverton, MD: ORC Macro. Retrieved from: http://dhsprogram.com/pubs/pdf/CR6/CR6.pdf.

For more on the origins of the wealth index:

Filmer, D. & Pritchett, L.H. Estimating wealth effects without expenditure data--or tears: An Application to education enrollment systems in States of India. Demography (pre-2011), 38(1), 115-32. Retrieved from: http://search.proquest.com/docview/222949375?accountid=11243.

For more guidance on the application of PCA for creating wealth indices:

Vyas, S. & Kumaranayake, L. (2006). Constructing socio-economic status indices: how to use principal components analysis. Health Policy and Planning21(6), 459-468. Retrieved from: http://heapol.oxfordjournals.org/content/21/6/459.full.

Appendix C: Using an existing set of factor weights

You may be able to obtain a set of factor weights that someone has already extracted from a national survey dataset. In this case, you can skip steps 1.2, 1.3 and 3. It is still a good idea to read through and understand these steps so that you understand the numbers that have been given to you. You will need not only the factor weights, but also the means and standard deviations of each variable so that you can standardize your variables in step 4.1.

You will also need a clear understanding of which questions from the national survey are required to get to the wealth index variables and make sure you have all those questions in your questionnaire (step 2.1).

Note: If you miss out a question, you will not be able to use the factor weights you received, because the numbers only work with all of the variables.

Appendix D: Using the equity measurement toolkit

The Equity Measurement Toolkit is a tool designed to measure the wealth level of a program's clients. However, the questionnaires and analysis files in the toolkit can be used for population based surveys. For each country covered by the toolkit there is a STATA do-file and SPSS syntax file which will calculate the wealth index score for each survey respondent based on the most recent DHS for that country.

Note: These analysis files only work if you retain the question coding and response coding from the questionnaire provided with the toolkit exactly.

REFERENCES

Chakraborty, N., Firestone, R., and Bellows, N. (2013). Equity monitoring for social marketing: use of wealth quintiles and the concentration index for decision making in HIV prevention, family planning, and malaria programs. BMC Public Health, 13(Suppl.2), S6.

DeVellis R.F. (2012). Scale Development: Theory and Applications, 3rd edition. Thousand Oaks, CA: Sage Publications.

Filmer, D. & Pritchett, L.H. (2001). Estimating Wealth Effect Without Expenditure Data or Tears: An Application to Educational Enrollments in States of India. Demography (pre-2011), 38, 115-32.

Haughton, J. & Shandker, S.R. (2009). Handbook on Poverty and Inequality. Washington, D.C.: The World Bank. Retrieved from: http://siteresources.worldbank.org.

Rutstein, S.O. & Johnson, K. (2004). The DHS Wealth Index. DHS Comparative Reports No. 6. Calverton, MD: ORC Macro. Retrieved from: http://dhsprogram.com/pubs/pdf/CR6/CR6.pdf.

Social Franchising for Health (2014). Equity. San Francisco: University of California, San Francisco. Retrieved from: http://sf4health.org/measuring-performance/equity.

The DHS Program (2014). DHS Overview. Rockville, MD: ICF International. Retrieved from: http://dhsprogram.com/What-We-Do/Survey-Types/DHS.cfm.

UNICEF (2014). Multiple Indicator Cluster Surveys. New York: UNICEF. Retrieved from: http://www.childinfo.org/mics_introduction.html.

Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: how to use principal components analysis. Health Policy and Planning, 21(6), 459-468.

WHO. (2014). Monitoring health inequality: an essential step for achieving health equity. Geneva: WHO. Retrieved from: http://www.who.int/gho/health_equity/en/.

WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation. Improved and unimproved water and sanitation facilities. Geneva: WHO; New York: UNICEF. Retrieved from: http://www.wssinfo.org/definitions-methods/watsan-categories/.