Reading and Understanding DHS Tables

Statistical tables can look intimidating at first glance. These examples, from the 2005-06 Zimbabwe DHS report, highlight ways to read and understand tables.

Example 1: Knowledge of HIV Prevention Methods:
A question asked of all survey respondents

Step 1: Read the title and subtitle. They tell you the topic and the specific population group being described. In this case, the table is about women age 15-49 in Zimbabwe. This represents the entire female survey population in most DHS surveys.

Step 2: Scan the column headings - the top horizontal row. They describe how the information is categorized. In this case, each column represents one aspect of knowledge of prevention the women report to have. Note that the last column lists the (weighted) number of women in each group. (For more on weighting, see example 4.)

Step 3: Scan the row headings - the first vertical column. These show the different ways the data are divided up into categories based on population characteristics. In this case, the table presents knowledge of HIV prevention methods by age, marital status, urban/rural residence, province, educational level, and wealth. Most DHS tables will be divided up in the same categories.

Step 4: Look at the very last row at the bottom of the table. These percentages represent the totals of all women age 15-49 who know each method of HIV prevention. In this case, 75.7% of women age 15-49 know that using condoms reduces the risk of getting HIV.

Step 5: To find out what percentage of women in Harare know that using condoms and limiting sex to one uninfected partner reduces the risk of getting HIV, draw 2 imaginary lines, as shown on the table. 64.6% of women of women age 15-49 in Harare know that using condoms and limiting sex to one uninfected partner reduces the risk of getting HIV.

Practice: Use this table to answer the following questions (answers are upside down, below right):

a) What percentage of young women age 15-24 are aware that using condoms AND limiting sex to one partner reduces the risk of getting HIV?

b) In what province are women most aware that abstaining from sex reduces the risk of getting HIV?

c) Are urban or rural women more likely to know that using condoms reduces the risk of getting HIV?

Table 13.2 Knowledge of HIV prevention methods

Percentage of women age 15-49 who, in response to a prompted question, say that people can reduce the risk of getting HIV by using condoms every time they have sexual intercourse, by having one sex partner who is not infected and has no other partners, and by abstaining from sexual intercourse, by background characteristics, Zimbabwe 2005-2006

| Background characteristic | Women | | | | |
	Using condoms[1]	Limiting sexual intercourse to one uninfected partner[2]	Using condoms and limiting sexual intercourse to one uninfected partner[1,2]	Abstaining from sexual intercourse	Number of women
Age					
15-24	72.4	78.5	61.5	78.7	4,104
15-19	67.8	76.7	57.7	77.7	2,152
20-24	77.5	80.5	65.7	79.9	1,952
25-29	80.6	84.3	70.6	83.0	1,466
30-39	80.2	82.9	69.4	83.5	2,050
40-49	73.6	80.4	64.2	79.6	1,287
Marital status					
Never married	71.9	79.6	61.9	80.8	2,404
Married/living together	76.4	80.8	65.6	79.9	5,143
Divorced/separated/widowed	80.0	82.6	69.6	83.6	1,360
Residence					
Urban	79.6	85.6	70.1	85.7	3,502
Rural	73.2	77.7	62.1	77.4	5,405
Province					
Manicaland	76.4	84.8	68.6	78.1	1,043
Mashonaland Central	72.8	81.4	65.8	73.2	825
Mashonaland East	84.8	76.6	69.0	90.4	714
Mashonaland West	69.4	76.8	57.7	77.7	829
Matabeleland North	68.2	79.8	58.6	77.3	536
Matabeleland South	64.6	72.8	57.0	72.8	439
Midlands	82.8	77.7	67.7	83.2	1,193
Masvingo	71.7	77.3	59.2	73.1	1,137
Harare	74.6	85.1	64.6	85.7	1,492
Bulawayo	86.0	90.4	81.7	91.8	697
Education					
No education	62.8	68.9	51.4	70.2	380
Primary	69.9	74.7	58.1	74.2	2,902
Secondary	79.5	84.6	69.8	84.5	5,355
More than secondary	80.9	85.4	70.7	90.0	270
Wealth quintile					
Lowest	66.9	73.1	54.7	73.5	1,552
Second	73.5	77.2	61.9	76.6	1,500
Middle	77.0	80.2	67.0	79.4	1,546
Fourth	77.5	82.6	67.1	83.2	2,006
Highest	80.7	87.0	71.7	86.8	2,304
Total 15-49	75.7	80.8	65.2	80.7	8,907

na = Not applicable
[1] Every time they have sexual intercourse
[2] Who has no other partners

a) 61.5%; b) Bulawayo-91.8%; c) urban-79.6%, versus 73.2% rural

Example 2: Payment for Sex and Condom Use
A question asked of a subgroup of survey respondents

Step 1: Read the title and subtitle. In this case, the table is about two separate groups of men a) all men age 15-49 and b) only men in that age group who reported that they paid for sex in the past 12 months.

Step 2: Identify the two panels. First identify the columns that refer to all men (a), and then isolate the columns that refer only to the men who reported having paid for sex (b).

Step 3: Look at the first panel. What percentage of all men age 15-49 have paid for sex in the past year? It's 3.9%.

Now look at the second panel. How many men are included in this group? Only 267, or about 4% of the 6,863 men who were asked about paying for sex. The second panel is a subgroup of the first.

Step 4: There are very few men who report having paid for sex in the past year. Once these men are further divided into the background characteristics categories, there may be too few cases for the results to be reliable.

For example, look to see the percentage of men in the lowest wealth quintile who used a condom at last paid intercourse: 46.5%. This percentage is in parentheses because it is based on fewer than 50 men (unweighted) in this category. This means that you should use this number with caution—it may not be accurate. (For more information on weighted and unweighted numbers, see example 4.)

If a number is replaced by an asterisk (as for men with more than secondary education who used a condom at last paid sex), it is because there are fewer than 25 unweighted people in that category, and results for this group are not reported. The subgroup is too small, and therefore those data are not reliable.

Table 13.9 Payment for sexual intercourse and condom use at last paid sexual intercourse: men

Percentage of men age 15-49 reporting payment for sexual intercourse in the past 12 months, and among them, the percentage reporting that a condom was used the last time they paid for sexual intercourse, by background characteristics, Zimbabwe 2005-2006

Background characteristic	Percentage who paid for sexual intercourse in the past 12 months	Number of men	Percentage reporting condom use at last paid sexual intercourse	Number of men who paid for sexual intercourse in the past 12 months
Age				
15-24	3.3	3,358	76.7	111
15-19	1.1	1,899	(76.6)	21
20-24	6.2	1,459	76.7	91
25-29	5.6	1,082	78.1	60
30-39	4.0	1,545	76.5	61
40-49	3.9	878	(50.3)	34
Marital status				
Never married	3.8	3,404	76.4	128
Married or living together	3.0	3,132	66.7	95
Divorced/separated/ widowed	13.2	327	(80.5)	43
Residence				
Urban	4.1	2,767	85.8	114
Rural	3.7	4,096	64.5	153
Education				
No education	8.8	88	(64.2)	8
Primary	4.4	1,782	57.2	78
Secondary	3.6	4,588	81.3	167
More than secondary	3.6	405	*	14
Wealth quintile				
Lowest	3.9	1,042	(46.5)	41
Second	3.3	1,137	(74.0)	37
Middle	4.0	1,194	(69.1)	47
Fourth	4.9	1,892	79.4	93
Highest	3.0	1,599	(89.3)	49
Total 15-49	3.9	6,863	73.6	267

Note: Figures in parentheses are based on 25-49 unweighted cases. An asterisk indicates a figure is based on fewer than 25 unweighted cases and has been suppressed.

Note: When parentheses or asterisks are used in a table, the explanation will be noted under the table. So if there are no parentheses or asterisks on a table, you can proceed with confidence that enough cases were included in all categories.

Practice: Use this table to answer the following questions (answers are upside down, below):

a) In what age group is payment for sex the most common?

b) Among the men with no education who paid for sex, what percentage used a condom the last time they had paid sex? Can you use this answer with confidence? Why or why not?

a) 20-24; b) 64.2%. There are parentheses around this figure, indicating that this result is based on only 25-50 unweighted cases. We cannot use this result with confidence.

Example 3: HIV Prevalence by Background Characteristics
Comparing Data and Understanding Patterns

Step 1: Read the title and subtitle. In this case, the table is about HIV prevalence among women age 15-49 in Zimbabwe.

Step 2: Scan the column headings - the top horizontal row. In this case there is only one variable, the percent of women who are HIV-positive.

The third column, "number," indicates how many women in each category were tested for HIV in the survey. In this case, 2,670 women living in urban areas were tested for HIV in the survey. In total, 6,947 women were tested.

Step 3: Scan the row headings - the first vertical column. These show the different ways the data are divided up into categories based on population characteristics. In this table, HIV prevalence is presented by urban/rural residence, province of residence, educational level, and wealth quintile. The data in these categories will help you understand how HIV prevalence varies throughout the country.

Step 4: Answer the following questions to understand how HIV prevalence is spread throughout the population:

-In which provinces is HIV prevalence the lowest and highest (range)? Prevalence ranges from 17.3% among women in Masvingo to 24.6% among women in Matabeleland South.

-Look for patterns: do results go up or down? Does prevalence vary within specific populations? For example, is there a clear pattern of HIV prevalence by wealth? by education?

-Compare different groups: do urban residents have a different HIV prevalence than rural residents?

Table 14.4 HIV prevalence by socioeconomic characteristics 1

Percentage HIV positive among interviewed women age 15-49 who were tested, by background characteristics, Zimbabwe 2005-2006

Background characteristic	Women	
	Percentage HIV positive	Number
Residence		
Urban	21.6	2,670
Rural	20.8	4,277
Province		
Manicaland	22.3	823
Mashonaland Central	22.9	665
Mashonaland East	21.3	560
Mashonaland West	22.5	666
Matabeleland North	22.8	421
Matabeleland South	24.6	345
Midlands	20.1	935
Masvingo	17.3	898
Harare	21.1	1,169
Bulawayo	19.6	466
Education		
No education	20.0	301
Primary	22.4	2,263
Secondary	20.7	4,194
More than secondary	15.8	189
Wealth quintile		
Lowest	17.7	1,223
Second	21.1	1,183
Middle	22.7	1,240
Fourth	26.8	1,579
Highest	17.1	1,722
Total	21.1	6,947

Step 5: What does all this mean? Once you see that HIV prevalence is spread pretty evenly throughout the country, how can this knowledge affect your work? Program planners and policymakers can use this detailed information to inform their activities. Perhaps earlier prevention programs had focused on urban populations, but these data show that women in rural areas have approximately the same HIV prevalence as those in urban areas. Furthermore, it is clear that health centers in all provinces need to be prepared to deal with large numbers of HIV-positive women, as HIV prevalence among women is over 17% in all provinces.

What other patterns exist? In Zimbabwe, HIV prevalence is highest among fairly wealthy women- those in the fourth wealth quintile. What does this mean for the program planner? It means that outreach should not be targeted only at the poorest women in the population. Also, women of all educational levels have high HIV prevalence, although prevalence does drop slightly among those with more than secondary education. Programs should be addressed to the full range of women—those who have received no education, as well as the most educated.

Example 4: Understanding Samples and Weighting in DHS Tables

A sample is a group of people that has been selected for a survey. The sample is supposed to represent the entire population that you would like to learn about. Most countries want to collect data and report information for the entire country, as well as for a country's regions or provinces.

DHS surveys are designed to provide these national and regional statistics. We want the sample surveyed in each region to resemble the actual population of that region, just as we want the national sample to resemble the actual population of the country. If the regions in a particular country vary in size and especially if some regions have very small populations, then a randomly-drawn sample may not include enough people from each region to allow for analysis.

For example, let's say that you have enough money to interview 14,000 women for a survey that should be representative of both the regions and the entire country (as in the Ethiopia table below*). In Ethiopia, the regions are not evenly distributed - some regions are very heavily populated, while others have very small populations.

A sampling statistician can determine how many women should be interviewed in each region in order to get reliable statistics for the specific indicators the country is interested in. In the case of Ethiopia, the blue column (1) shows the actual number of women selected and interviewed in each region, ranging from 729 in Gambela to 2, 230 in Oromiya. Now there are enough interviews to get reliable results in each region.

But now there is a new challenge. With this distribution of interviews throughout the regions, some regions are overrepresented and some regions are underrepresented. For example, Gambela's population in 2005 was about 250,000 people, which only represents 0.3% of the entire Ethiopian population. On the other hand, Oromiya's population in 2005 was 27,000,000, or approximately 36% of the Ethiopian population. But based on the blue column, our survey has interviewed only about 3 times as many women in Oromiya as in Gambela. Does this represent the population of the country?

In order to get statistics that are representative of the entire country, the distribution of the women in our sample needs to resemble the distribution of the women in the country. Women from a very small region, like

Table 3.1 Background characteristics of respondents			
Percent distribution of women by selected background characteristics, Ethiopia 2005			
	Women		
Background characteristic	Weighted percent 3	Weighted 2	Unweighted 1
Region			
Tigray	6.5	919	1,257
Affar	1.0	146	789
Amhara	24.7	3,482	1,943
Oromiya	35.6	5,010	2,230
Somali	3.5	486	669
Benishangul-Gumuz	0.9	124	846
SNNP	21.3	2,995	2,087
Gambela	0.3	44	729
Harari	0.3	39	844
Addis Ababa	5.4	756	1,869
Dire Dawa	0.5	69	807
Total	100.0	14,070	14,070

Gambela, should only contribute a very small amount to the national total. Likewise, women from a large region, like Oromiya, should contribute more. So, DHS statisticians adjust or "weight" the numbers of women from each region so that each region's contribution to the total is proportionate to the actual population of the region. The numbers in the purple column (2) represent the "weighted" values. The total sample size of 14,070 women has not changed, but the distribution of the women in the regions has been changed to represent their contribution to the total population size.

How do statisticians weight each category? They recalculate the categories to reflect the real population of the country. If you were to compare the light red column (3) to the actual population distribution of Ethiopia, you would see that women in each region are contributing to the total sample with the same weight that they contribute to the population of the country. The weighted number of women in the survey now accurately represents how many women live in Oromiya- 36 percent of the Ethiopian population - and how few women live in Gambela - less than 1 percent of the population.

With sampling and weighting, it is possible to interview enough women to provide reliable statistics at both the national and regional level, without distorting the overall distribution of population within the country. In general, only the weighted numbers are shown in each of the DHS tables, so don't be distressed if these numbers seem low- they may actually represent a larger number of women interviewed. And remember, the table will use parentheses and asterisks to warn you if there are too few unweighted cases in any category.

* Because the population in Zimbabwe is relatively evenly spread across provinces, weighted/unweighted differences by provice are very small. To illustrate the differences in weighted and unweighted samples, the Ethiopia example, which is more dramatic, is used. The same general sampling principles shown in the Ethiopia example also apply to Zimbabwe.