

Reading and Understanding DHS Tables

Statistical tables can look intimidating at first glance. These examples, from recent DHS reports, 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 Tanzania. 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, zone of residence, educational level, and wealth. Most of the tables in DHS reports will be divided up in the same categories.

Step 4: Look at 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, 78.5% 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 the Eastern Zone know that using condoms and limiting sex to one uninfected partner reduces the risk of getting HIV, draw two imaginary lines, as shown on the table. This shows that 75.8% of women of women age 15–49 in the Eastern Zone know that using condoms and limiting sex to one uninfected partner reduces the risk of getting HIV.

Table 12.2 Knowledge of HIV prevention methods **1**
Percentage of women age 15–49 who, in response to prompted questions, say that people can reduce the risk of getting HIV by using condoms, by having sex with just one partner who is not infected and who has no other partners, and by abstaining, by background characteristics, Tanzania 2004–05 DHS

Background characteristic	Using condoms 2	Limiting sex to one uninfected partner	Using condoms and limiting sex to one uninfected partner	Abstaining from sex	Number of women
Age					
15–19	73.1	87.9	68.9	89.4	2,245
20–24	81.6	90.3	76.9	92.5	2,007
25–29	82.3	92.8	78.4	94.2	1,885
30–39	81.1	93.0	77.6	93.0	2,595
40–49	73.4	92.2	70.6	94.0	1,597
15–24	77.1	89.1	72.7	90.8	4,252
Marital status					
Never married	74.4	89.1	70.4	91.6	2,371
Ever had sex	85.6	92.6	81.5	94.9	1,022
Never had sex	65.9	86.4	62.1	89.1	1,350
Married/living together	79.2	91.8	75.4	92.4	6,950
Divorced/separated/widowed	82.9	92.3	79.1	95.0	1,007
Residence					
Urban	82.8	94.3	79.7	94.8	2,935
Rural	76.8	90.0	72.6	91.6	7,394
Zone					
Western	74.0	88.4	68.5	89.0	1,880
Northern	76.1	91.0	73.9	92.2	1,496
Central	65.4	79.0	59.5	86.0	799
Southern highlands	75.6	92.5	72.2	92.8	1,440
Lake	89.5	94.8	86.5	96.1	1,865
Eastern	79.6	93.9	75.8	95.3	1,670
Southern	88.3	94.9	85.2	95.2	866
Education					
No education	66.8	85.3	62.8	87.1	2,503
Primary incomplete	77.8	90.0	73.0	93.2	1,855
Primary complete	83.4	93.4	79.6	94.3	5,086
Secondary+	84.6	97.6	83.4	96.1	885
Wealth quintile					
Lowest	70.2	87.2	66.7	88.1	1,840
Second	75.7	88.9	71.3	90.7	1,944
Middle	78.9	91.7	75.2	92.8	1,943
Fourth	82.1	93.0	77.9	94.7	2,004
Highest	83.4	94.0	80.0	95.0	2,597
Total	4 78.5	91.2	74.6	92.5	10,329

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

- What percentage of youth age 15–24 are aware that using condoms AND limiting sex to one partner reduces the risk of getting HIV?
- In what zone are women most aware that abstaining from sex reduces the risk of getting HIV?
- Are urban or rural residents more likely to know that using condoms reduces the risk of getting HIV?

a) 72.7%; b) Urban—82.8%; Lake—96.1%; c) Urban—82.8%

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 10.6%.

Now look at the second panel. How many men are included in this group? Only 278, or 10.6% of 2,635 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 percentages to be reliable.

For example, look to see the percentage of men age 25–29 who used a condom at last paid intercourse: 57.2%. This percentage is in parentheses because there are fewer than 50 men (unweighted) in this category. Readers should use this number with caution—it may not be accurate. (For more information on weighted and unweighted numbers, see Example 4.)

Look also to see what percentage of men with secondary or higher education used a condom at last paid intercourse. There is no number in this cell—only an asterisk. This is because fewer than 25 men (unweighted) with secondary or higher education reported having paid for sex. Results for this group are not reported. The subgroup is too small, and therefore the data are not reliable.

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

- a) In what age group is payment for sex 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) 39.3%, but because this is based on fewer than 50 cases, you cannot use this number with any confidence.

Table 12.9 Payment for sexual intercourse by men and condom use at last sexual intercourse

Percentage of men age 15–49 who reported 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, Tanzania 2004-05 DHS

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				
15-19	9.5	637	46.4	60
20-24	16.9	493	65.2	83
25-29	13.4	405	(57.2)	54
30-39	9.3	665	58.5	62
40-49	4.2	435	*	18
15-24	12.7	1,130	57.3	144
Marital status				
Never married	12.8	1,100	62.4	140
Married or living together	7.6	1,401	54.8	106
Divorced/separated/ widowed	23.7	135	(58.0)	32
Residence				
Urban	11.1	716	79.1	80
Rural	10.4	1,919	50.9	199
Education				
No education	9.2	312	(39.3)	29
Primary incomplete	11.7	646	42.8	76
Primary complete	11.1	1,381	66.2	153
Secondary+	7.1	296	*	21
Wealth quintile				
Lowest	9.3	484	(38.2)	45
Second	11.2	504	45.2	56
Middle	9.0	516	(62.2)	46
Fourth	12.3	517	58.9	63
Highest	10.9	615	(82.2)	67
Total	10.6	2,635	59.0	278

Note: Figures in parentheses are based on 25-49 unweighted cases. An asterisk indicates that 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. If there are no parentheses or asterisks on a table, you can proceed with confidence that enough cases were included in all categories that the data are reliable.

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 Uganda.

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 tested) indicates how many women in each category were tested for HIV as part of the survey. In this case, 1,435 women living in urban areas were tested for HIV.

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. This table presents HIV prevalence by urban-rural residence, region of residence, educational level, employment status, and wealth. 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:

- What is the highest prevalence and the lowest prevalence (range)? Within the regions, prevalence ranges from 2.7% in West Nile to a high of 11.8% in Kampala.
- 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?

Step 5: What does all this mean? Once you see that urban residents and residents of Kampala and Central regions have the highest HIV prevalence, how can this knowledge affect your work? Program planners and policymakers can use this kind of detailed information to inform their activities. Perhaps they should focus prevention campaigns in those areas, and make sure that there are health centers prepared to deal with larger numbers of HIV-positive women in those areas.

What other trends exist? In Uganda, HIV prevalence increases as a woman's wealth increases. What does this mean for the program planner? It means that outreach should not be targeted only at the poorest women in the population. Wealthy women are more likely to carry HIV infection than poorer women. Programs should be designed to address them. There is no clear pattern among educational levels,

Table 8.4 HIV prevalence among women age 15–49 by background characteristics, Uganda 2004-05 AIS

Background characteristic	Women 15-49	
	Percentage HIV positive	Number tested
Residence		
Urban	12.8	1,435
Rural	6.5	7,956
Region		
Central	10.2	1,565
Kampala	11.8	634
East Central	7.5	1,467
Eastern	6.2	813
Northeast	3.6	779
North Central	9.0	918
West Nile	2.7	906
Western	7.8	1,076
Southwest	7.1	1,232
Education		
No education	5.8	2,129
Primary incomplete	7.7	4,355
Primary complete	9.8	1,064
Secondary+	7.6	1,826
Employment		
Currently working	8.4	5,758
Not working	6.1	3,560
Wealth quintile		
Lowest	4.8	1,532
Second	6.6	1,911
Middle	6.7	1,760
Fourth	7.0	1,895
Highest	11.0	2,294
Total 15-49	7.5	9,391

but there is no educational group that is safe from HIV. Programs should be addressed to the full range of women—those who have received no education through those who have higher levels of education.

Example 4: Understanding Sampling Weights in DHS Tables

A sample is a group of people that have 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 both for the entire country and also 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 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.

This creates a new challenge. With this distribution of interviews, 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. In contrast, 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 three times as many women in Oromiya as in Gambela. Does this represent the population of the country?

Percent distribution of women by selected background characteristics, Ethiopia 2005 DHS

Background characteristic	Women		
	Weighted percent	Weighted	Unweighted
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

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 Gambela, should only contribute a very small amount to the national total. Likewise, women from a large region, like Oromiya, should contribute more. Therefore, DHS statisticians adjust or "weight" the number of women from each region so that each region's contribution to the total is proportionate to the actual population of the country. 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% of the Ethiopian population) and how few women live in Gambela (less than 1% 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.