UNICEF East Asia and Pacific Regional Office
19 Phra Atit Road
Bangkok 10200 Thailand
Tel: (66 2) 356-9499 Fax: (66 2) 280-3563
Email: eapro@unicef.org
www.unicef.org/eapro

Understanding key terms and data related to HIV

Handbook for adolescents and young key populations at higher risk of HIV exposure in Asia and the Pacific

Understanding key terms and data related to HIV

Handbook for adolescents and young key populations at higher risk of HIV exposure in Asia and the Pacific

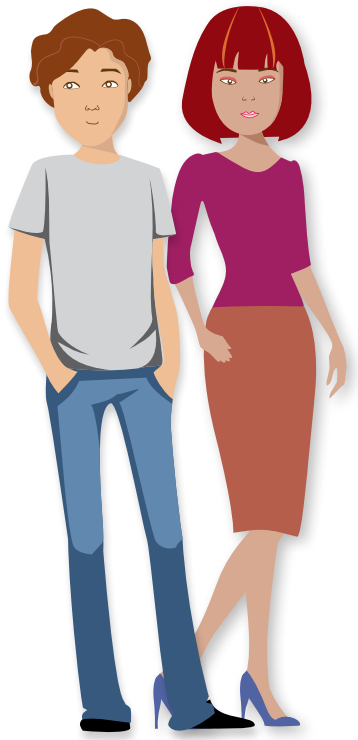
Published by: UNICEF East Asia and Pacific Regional Office
19 Phra Atit Road
Bangkok 10200 Thailand
Tel: (66 2) 356-9499 Fax: (66 2) 280-3563
Email: eapro@unicef.org
www.unicef.org/eapro

Suggested Citation: UNICEF EAPRO (2015), Understanding key terms and data related to HIV: Handbook for adolescents and young key populations at higher risk of HIV exposure in Asia and the Pacific. UNICEF EAPRO, Bangkok.

Illustration, design and layout: QUO Bangkok

ISBN: 978-974-685-149-7

CONTENTS



ACKNOWLEDGEMENTS

This handbook is a product of a collaborative effort of UNICEF East Asia and Pacific Regional Office and members of the Asia-Pacific interagency task team on Young Key Populations who responded to the need for a tool to equip young people who are interested in understanding key terms and data related to HIV. It is designed as a "comic book" and can be translated and used widely in both Asia-Pacific and other regions.

The main author of this handbook is Lisa G. Johnston, University of California, San Francisco, Global Health Sciences and Tulane University, School of International Public Health and Tropical medicine with support from Shirley Mark Prabhu, HIV/AIDS Specialist, Knowledge and Advocacy, UNICEF EAPRO. The core team who have advised on the content included Wing-Sie Cheng, Regional Adviser – HIV & AIDS; Devashish Dutta, Youth and Adolescent Specialist and Revati Chawla, Consultant, UNICEF EAPRO; Justine Sass, Regional HIV & AIDS Adviser for Asia and the Pacific, Chief, HIV Prevention and Health Promotion (HP2) Unit, UNESCO Bangkok; Josephine Sauvarin, Technical Advisor on HIV, Adolescent Sexual and Reproductive Health, UNFPA APRO; Dongbao Yu, Epidemiologist, HIV/AIDS and STIs, WHO SEARO; Pengfei Zhao, Scientist, HIV Prevention and Key Populations, HIV, HIV, Hepatitis and Sexually Transmitted Infections Unit, WHO WPRO; Aries Valeriano, Youth and Social Organization Officer, UNAIDS RST-AP; Khin Cho Win Htin, Data Specialist, and Ye Yu Shwe, Data Specialist, HIV/AIDS Data Hub, UNAIDS RST- AP; Thaw Zin Aye, Regional Coordinator; Gaj Gurung, Programme Coordinator; and Jeffrey P. Acaba, Education and Research Lead, Youth LEAD; Niluka Perera, Project Officer, Youth Voices Count; and Scott McGill, Senior Advisor at Save the Children, US.

Other members who provided comments included Priscilla Atwani Idele, Senior Advisor – Statistics and Monitoring Section, UNICEF New York; Lei Zhang, HIV/AIDS Specialist, UNICEF China; Lori Thorell, Senior Consultant, ICT for HIV and Programme, UNICEF EAPRO; Saba Moussavi, Consultant - Data and Research, UNICEF EAPRO and Nancy Zhang, Coordinator, China Youth Leaders' Resource Center, China.

ACRONYMS

CRC	Convention on the Rights of the Child
FSW	Female sex worker
GFATM	The Global Fund to Fight AIDS, Tuberculosis and Malaria
HTC	HIV testing and counselling
GARP	Global AIDS Response Progress Report
IBBS	Integrated Biological and Behavioural Survey
MSM	Males who have sex with males
NAC	National AIDS Council
NACP	National AIDS Control Programme
NCCWD	National Commission on Child Welfare and Development
NGO	Non-governmental organization
PWID	People who inject drugs
STI	Sexually transmitted infection
UN	United Nations
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNDP	United Nations Development Programme
UNFPA	United Nations Population Fund
UNGASS	United Nations General Assembly Special Session on AIDS
UNICEF	United Nations Children's Fund
WHO	World Health Organization

OVERVIEW

There are data and statistics all around us. We grow up seeing them in our game scores, talking about it whenever we buy something, and reading about them in newspaper articles or on the internet. Data can help us understand issues, and think about solutions to problems.

With the data we have on the HIV epidemic, we can learn many things. For example, we can understand how many people are infected, what behaviours put them at risk, what can protect them from HIV infection, how many people are being reached by our programmes, and what impacts are being made on people's lives.

Understanding these data can be complex for anyone. This handbook will help simplify some of the key questions you have about HIV terms, data and statistics. Let's get started.

This handbook is designed as a "comic book" in which you will follow Tran (an 18 year old who is an HIV peer outreach worker) and his friends, Lucia (a 17 year old who lives on the streets), Aran (a 15 year old who is a transgender person and lives with HIV), John (a 19 year old who works at the men who have sex with men action centre) and Bindu (a 22 year old who uses drugs)

With this group, you will go through the process of making sense of statistics. Their conversations will provide you with the minimal statistical skills necessary to understand HIV data and actively engage in discussions with young people and adults. In addition, Tran, Lucia, Aran, John and Bindu will show you how to ask important questions about where these statistics come from and whether they are valid or not.

Who this handbook is for

This handbook is for young people between the ages of 15 and 24 years old of age who are interested in HIV issues and have some basic math skills. Adolescent and young peer educators, young advocates and young people involved in HIV programming for young people, including those from key populations at higher risk of HIV exposure[1], will find it particularly useful.

What this handbook is for

There are a lot of data about HIV and AIDS. How much of it do you really understand? Have you ever read a document, watched a presentation or talked with your friends about HIV data and not understood the meaning of certain terms? For example, let's say you read that in 2008, 95% of all new HIV infections in young people were among young key populations.[2] Would you know if this statistic is describing HIV incidence or prevalence? Or what if you heard that "68.7% of males under the age of 25 who have sex with males in Cambodia reported using a condom at last sex with a male partner"[3]? Would you wonder how they were able to know about condom use among young MSM in ALL of Cambodia? What about graphs, charts or other graphics? Have you ever looked at a graphic and not understood what it was showing? Could you draw a trend chart to explain changes in HIV testing among young people in your country over time?

If you have ever asked any of these questions, this handbook is for you. It will help you to understand data and terms that are often found in reports, presentations, fact sheets and other materials about HIV. It will help you ask important questions about data such as where are they from? Who do they represent? When were they gathered? Finally, it will help you understand and use numbers and graphics to explain data to others.

What this handbook includes and how it can be used

This handbook can be used as a tool to equip you, as a young person, to be involved in advocacy work and to leverage funding from community and government leaders. It can help you influence the decision-making process and programming for you and your peers.

It can also help you to explain data to your peers so that they can be better advocates and leaders.

This handbook is designed as a "comic book", to take you through the explanations.

It is divided into four key sections:

A Defining and using key terms about data

B Reading tables and graphs

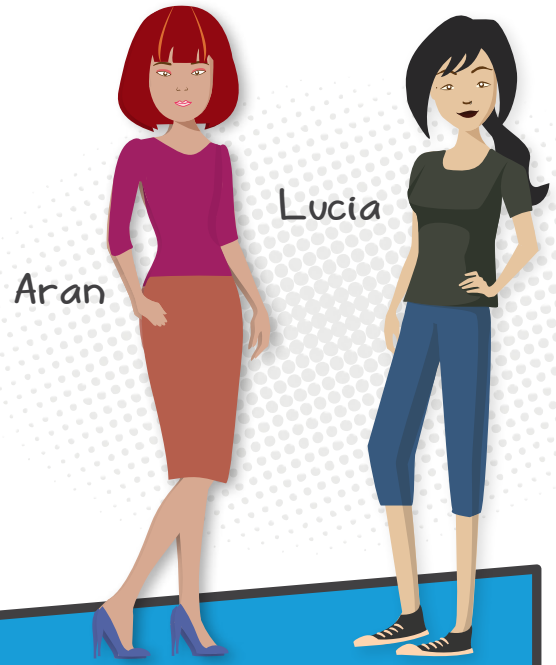
C Producing graphs for information sharing and advocacy

D Questioning data

At the end of the handbook you can find quizzes to help you test your learning, as well as definitions for terms used throughout the handbook.

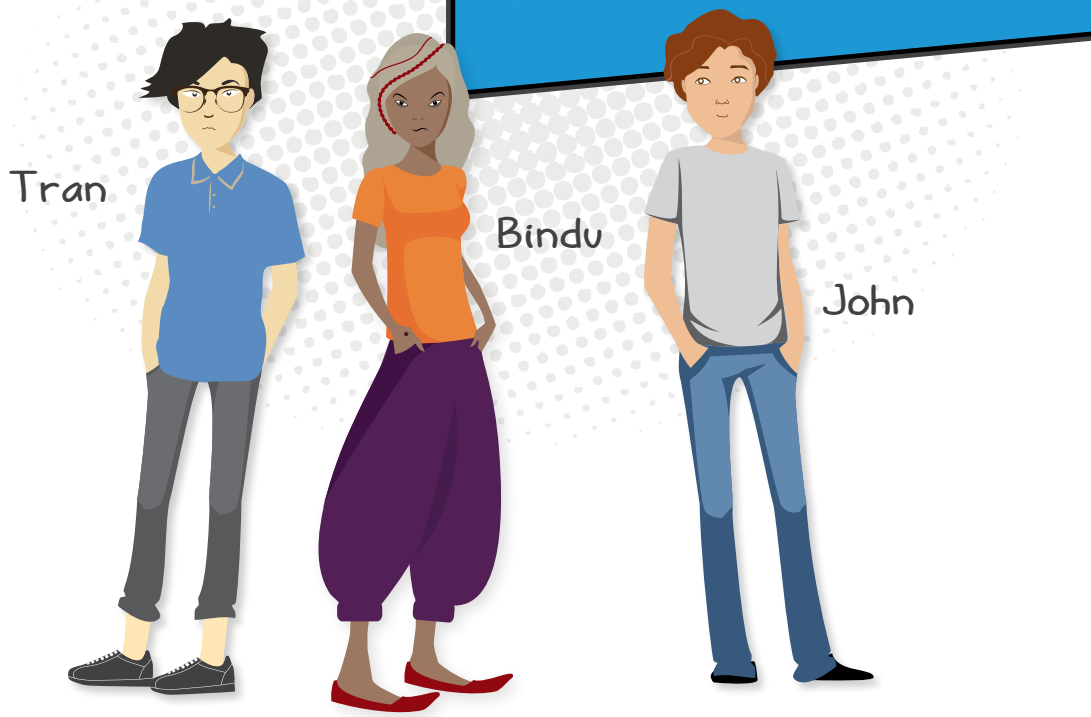
[1] See Appendix B for definitions of young key populations and age groups.
[2] Commission on AIDS in Asia (2008). Redefining AIDS in Asia: Crafting an Effective Response. Accessed at: http://www.unaids.org/en/resources/presscentre/featurestories"2008/march/20080326asiacommission/.
[3] Cambodia National AIDS Authority. (2012) Cambodia AIDS Response Progress Report 2012. Phnom Penh. National AIDS Authority. Accessed at: http://www.unaids.org/en/dataanalysis/knowyourresponse/countryprogressreports/2012countries/ce_KH_Narrative_Report[1].pdf

INTRODUCTION



Lucia and Aran have been chosen as youth representatives to speak at a community meeting about HIV. Both Lucia and Aran are excited to represent their peers but are uncomfortable since they are unfamiliar with statistics and data.

They have asked Tran, an outreach worker, to help them. Two other friends, Bindu and John, decide to join them as well.

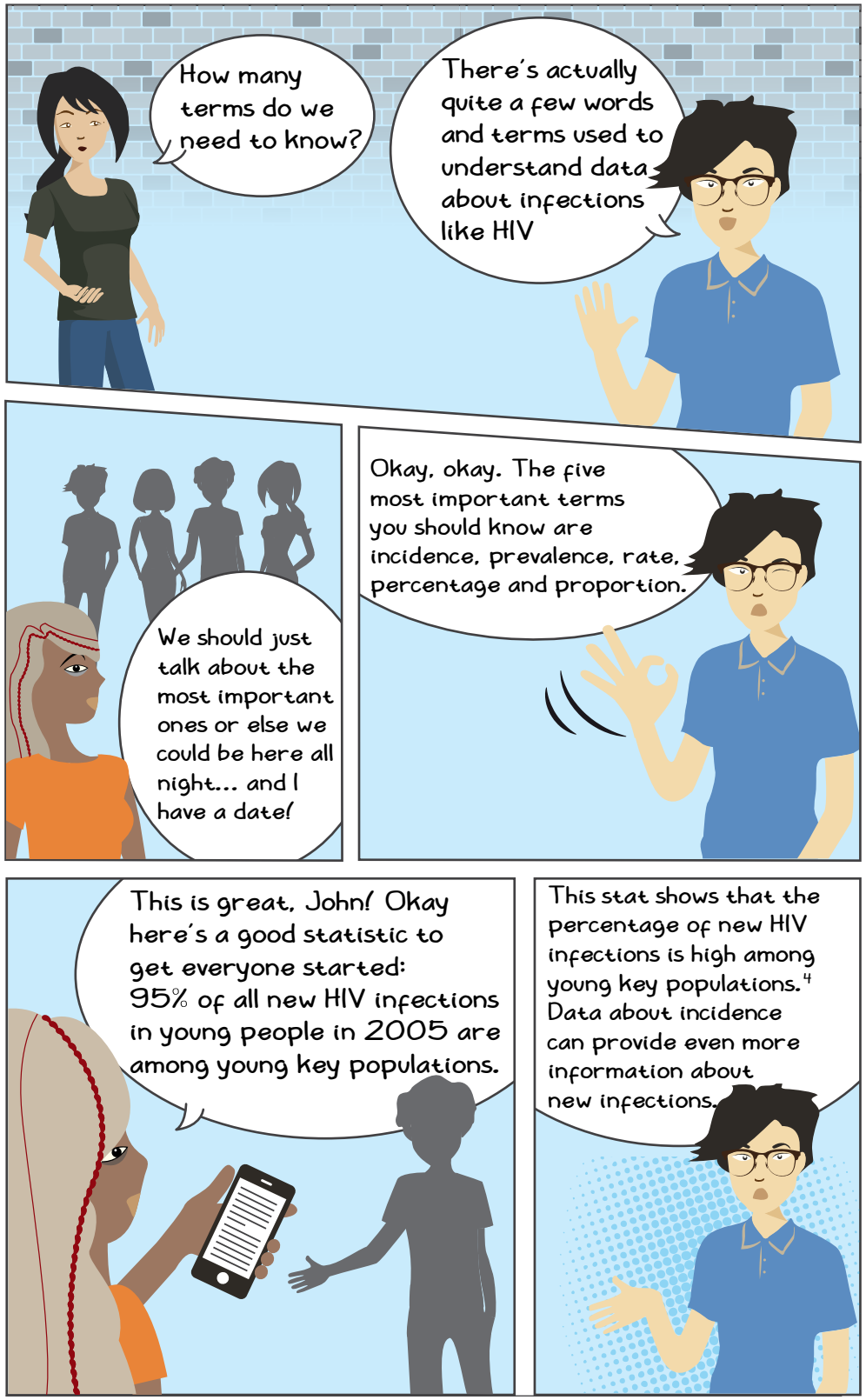



Thanks for helping us. We've been so lost on our own.
Yeah, we really appreciate all the help we can get. We didn't realize how important statistics and understanding data were when we agreed to be youth representatives.
No problem, we're happy to help.
I'd be even happier if you offered to buy us dinner after we're done.
Don't listen to Bindu. The three of us are excited to work with you.
Where would you like to start?


We're not really sure.
Well I know that when I learned about data for my job, I learned the key terms and definitions first, and that was a big help. So we should start there.
That sounds like a great idea, let's get started!

DEFINING AND USING KEY TERMS ABOUT DATA

In this section Tran meets with Lucia, Aran, John and Bindu at a café to discuss the meaning of incidence, prevalence, rates, percentages and proportions.



[4]Background Commission Paper Asia Epidemic Model (AEM); Unified Budget and Work plans (UBW, 2004-2005). Found in: Commission on AIDS in Asia (2008). Redefining AIDS in Asia: Crafting an Effective Response, pp. 145-146.

In 2013, there were 38,000 new HIV infections in South-East Asia and the Pacific among adolescents ages 15-19 years. In this example, incidence is the number of new HIV infections occurring in a specific group, adolescents, within a specific period of time, 2013.

This means that a lot of young people were newly infected with HIV in 2013.

So whenever you see the words ''new infections'' and a specific time period, like one year, then you can think about incidence.

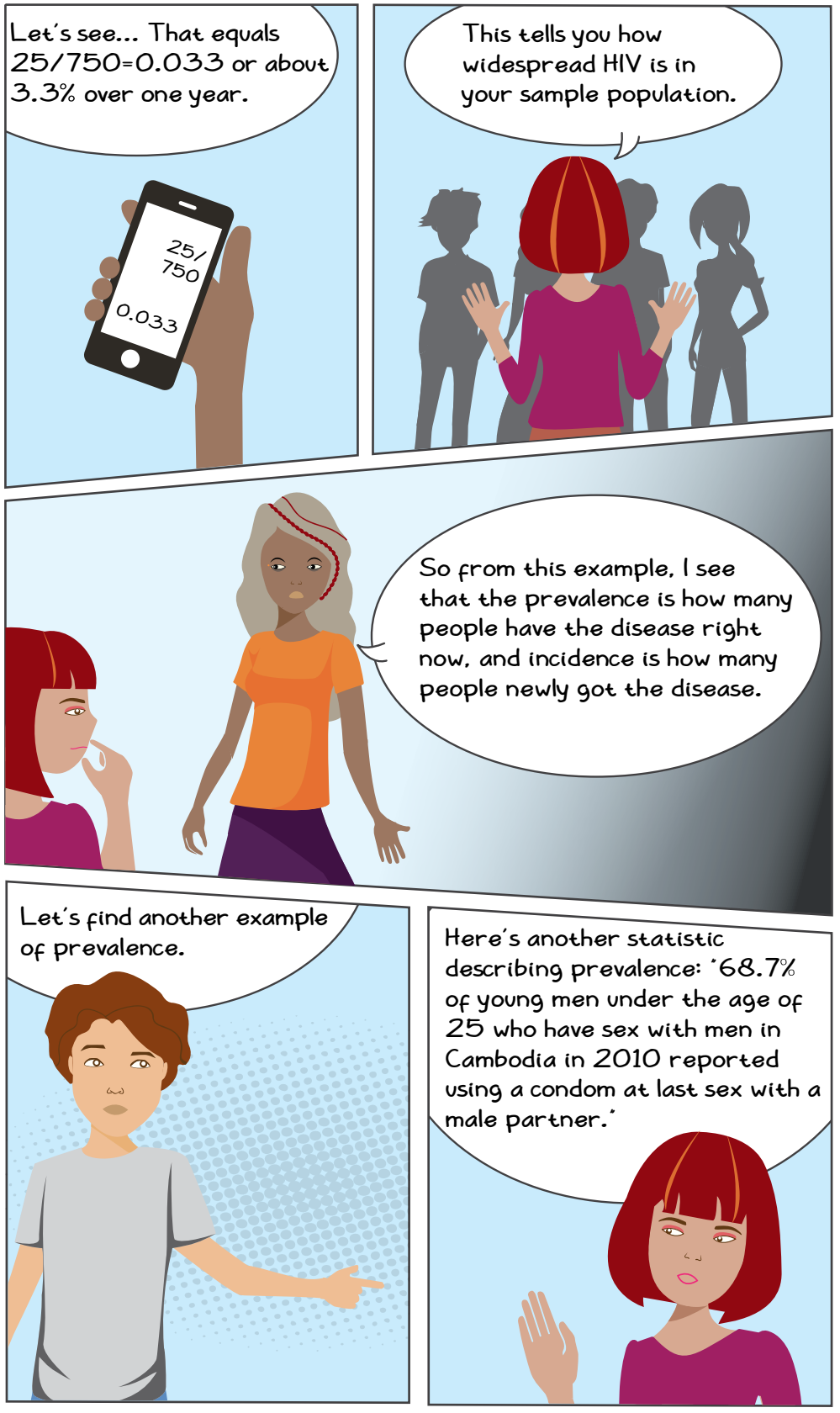
I have another example. I recently read about a study among males who have sex with males in Bangkok. At the beginning of the study no males were infected with HIV. Over 4 years, 6% of all males in the study became infected with HIV and among males aged 18 to 21 years, 30% became infected with HIV.

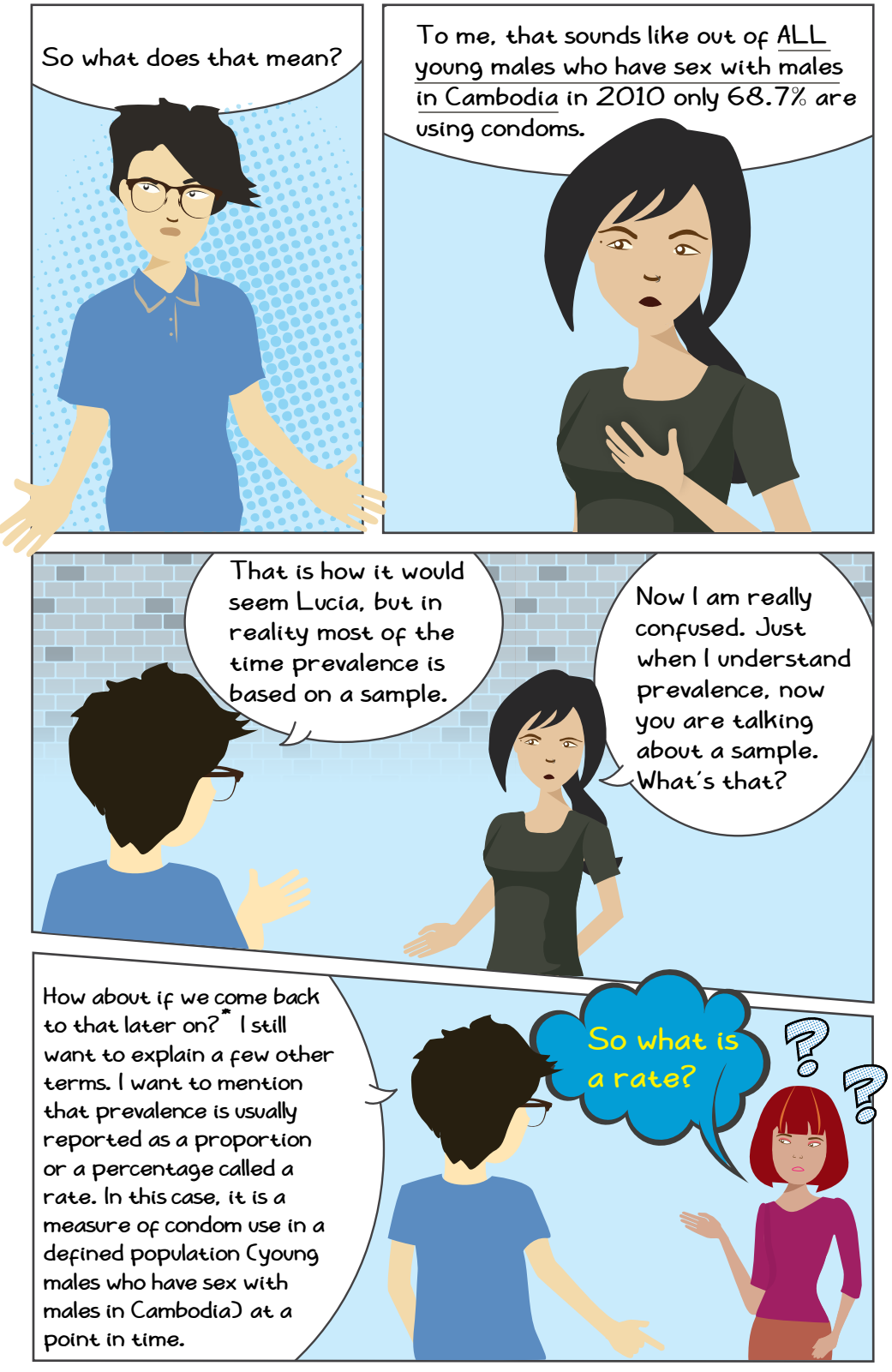
That's right Aran. You're getting it!

So are you saying that new infections, or incidence, of HIV in this group of males who have sex with males over 4 years was highest for those aged 18-21 years?

So how is incidence different from prevalence?

Let's say we want to know about HIV infection in a group of 800 high risk youth. We begin by taking blood tests and find that 50 are already infected. This would be prevalence or 50/800 which is 0.0625 or better is 6.25% . Then, if we want to estimate incidence of HIV in the remaining uninfected population of 750 young people over 12 months, we need to do more blood tests. Let's say that at the end of one year we find that 25 newly diagnosed with HIV (from those tested HIV_ negative one year ago). Who can tell me the incidence?





*See section on Questioning Data page 43.

That is a good question. I will use the example of people who die from AIDS-related deaths in a given city. To get a rate we need a denominator.

So the denominator is all the people at risk for a disease or in this case, people living with HIV/AIDS

Yes, So let's say that we know that the population of persons who are at risk of dying from AIDS related deaths is 200,000 and that there have been five deaths over a specific calendar year. If we do the math, this is 200000/5 or 0.000025. This is known as a crude death rate.

The rate is so small. I cannot understand it.

This is a good point. This is why we often use a constant such as 100,000 so we can make the rate into a whole number. Simply multiply the rate by 100,000 and see what you get.

That would be 2.5 deaths per 100,000 population.

That sounds like the same thing as a percentage or even a proportion. What is the difference between a percentage, a proportion and a rate?

The difference between a percentage and a proportion is that a percentage is multiplied by 100. So this means in the Cambodian males who have sex with males example, the proportion of condom users at last sex would be .687 and the percentage would be 68.7%.

A rate is different from a proportion or a percentage. All are calculated using a numerator and a denominator. But, the relationship between the two makes the rate different from the proportion or percentage. A rate is the measure of the number of times an event occurs in a defined population over a specified time period. Sometimes a rate is very small so it has to be divided by a larger number, such as 100,000, to make it into a number that makes sense.

READING TABLES AND GRAPHS

In this section Tran meets with Lucia, Aran, John and Bindu to discuss how to read and understand tables and different types of graphs including, tables and pie, line, and bar charts.

A. Reading Tables

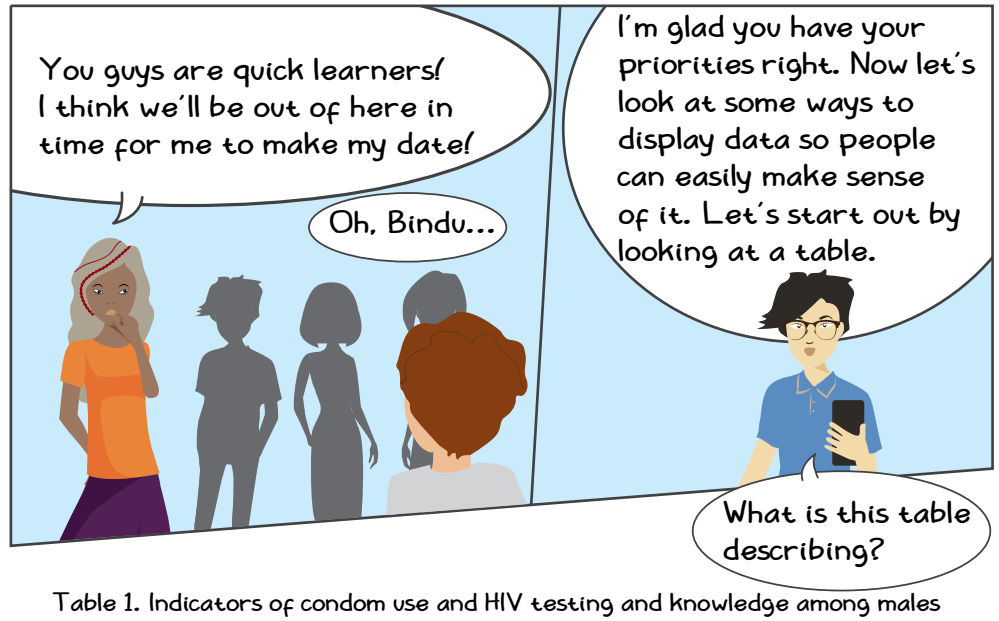
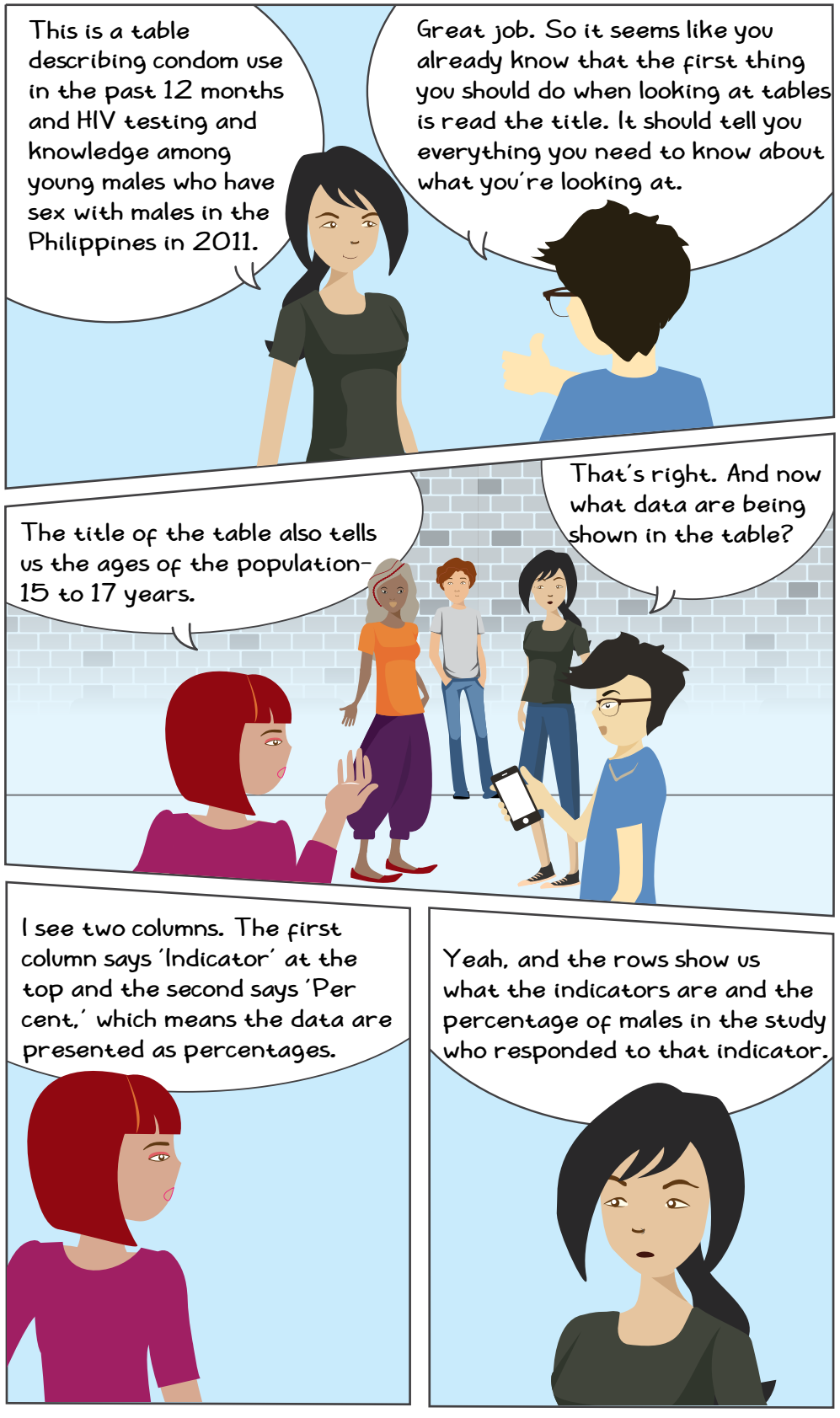


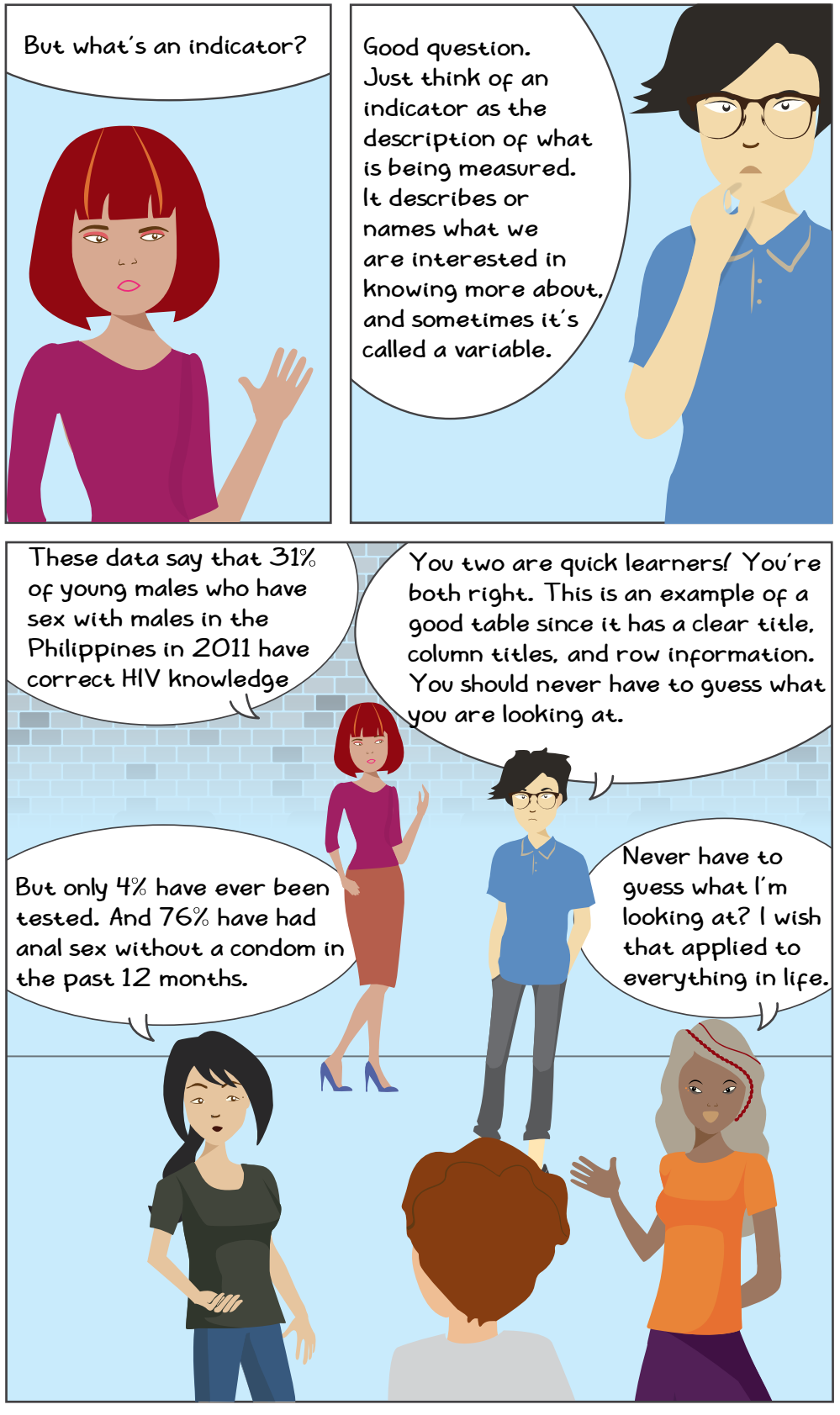
Table 1. Indicators of condom use and HIV testing and knowledge among males who have sex with males ages 15 to 17 years in the Philippines, 2011[5]

Indicator	Per cent*
Percentage who had anal sex without a condom in the past 12 months	76
Percentage who have ever been tested for HIV	4
Percentage who have correct HIV knowledge	31

* When we say 'per cent' we are really saying 'per 100'. One per cent (1%) means 1 per 100.

[5]Samonte GM. Capturing younger cohorts through surveillance systems: Philippines. Methodologies for obtaining strategic information on young people at higher risk of HIV exposure, Thailand, September 3-5, 2012.


This is a table describing condom use in the past 12 months and HIV testing and knowledge among young males who have sex with males in the Philippines in 2011.
Great job. So it seems like you already know that the first thing you should do when looking at tables is read the title. It should tell you everything you need to know about what you're looking at.
The title of the table also tells us the ages of the population– 15 to 17 years.
That's right. And now what data are being shown in the table?
I see two columns. The first column says 'Indicator' at the top and the second says 'Per cent,' which means the data are presented as percentages.
Yeah, and the rows show us what the indicators are and the percentage of males in the study who responded to that indicator.


But what's an indicator?
Good question. Just think of an indicator as the description of what is being measured. It describes or names what we are interested in knowing more about, and sometimes it's called a variable.
These data say that 31% of young males who have sex with males in the Philippines in 2011 have correct HIV knowledge
You two are quick learners! You're both right. This is an example of a good table since it has a clear title, column titles, and row information. You should never have to guess what you are looking at.
But only 4% have ever been tested. And 76% have had anal sex without a condom in the past 12 months.
Never have to guess what I'm looking at? I wish that applied to everything in life.

B. Reading a Bar Chart

Now let's take a look at other ways data can be displayed. Some common graphs are bar charts, pie charts and line charts. Does anyone know the difference?

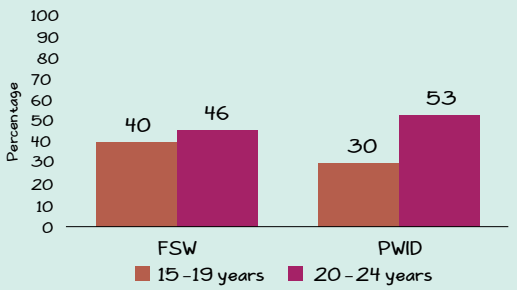
It seems like the name of the chart tells you everything. A bar chart is made up of bars, a pie chart looks like a pie and a line chart has lines in it. Right?

What's a pie?

Let's talk about the bar chart first, and then we can talk about pie charts after.

Great idea, Bindu. Let's start with a bar chart. A bar chart usually uses vertical (up and down) bars to display and compare data.

Figure 1. Percentage of Young Female Sex Workers (FSW) and People who Inject Drugs (PWID) who were reached by an outreach worker in the past three months, Indonesia, 2009



Source: UNICEF, Age Group Disaggregation of Survey and Research Data: Indonesia, 2009

Well, from the title I know that it is describing young female sex workers and people who inject drugs in Indonesia in 2009.

And it shows the percentages that have been reached by an outreach worker in the past three months.

Can you tell if there is more than one age group shown here?

At the bottom of the graph there are two small brown and pink boxes. Next to the brown is written '15 - 19 years' and next to the pink box is '20 - 24 years.'

So that means the colour represents specific age groups.

But besides two different colours, there are also two groups of bars. Can you explain that?

Well beneath the first set of two bars is FSW, which, from the title, are female sex workers, and beneath the second set is PWID who are people who inject drugs.

So that means the first set of bars describes female sex workers and the second set describes people who inject drugs.

That's good. It shows that this chart displays all the information you need to quickly understand its point. The part describing the vertical side, the percentages, is called the y-axis and the part describing the horizontal side, the group (FSW or PWID), is called the x-axis.

At the bottom the two coloured boxes the legend. A legend gives any extra information needed to read the graph, and is displayed as small coloured or patterned boxes matching the bars in the graphs with descriptions next to them. In this case, it gives us information about the age groups.

What else can you say about this bar chart?

The chart says that 53% of 20 - 24 year olds who inject drugs in Indonesia have been reached by an outreach worker in the past three months. That's the highest percentage out of all four bars.

And 40% of female sex workers aged 15 - 19 years have been reached by an outreach worker in the past three months.

That's right.

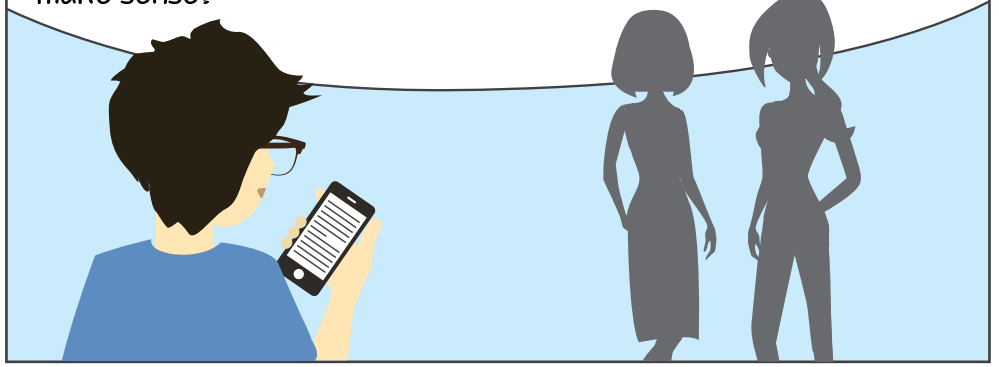
Don't forget to ask about where the data come from so you know whether it might be describing a sample or a population.

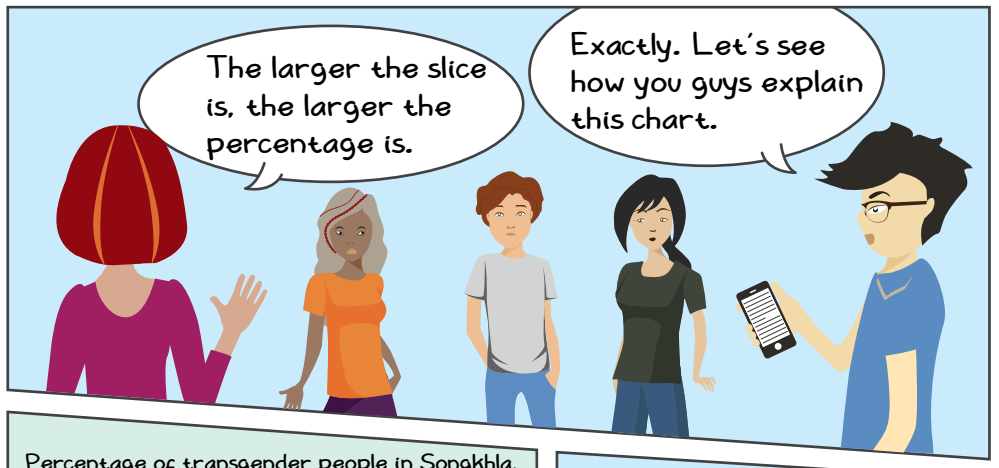
Right, I would want to know who and how many people this information describes and how they got the data.

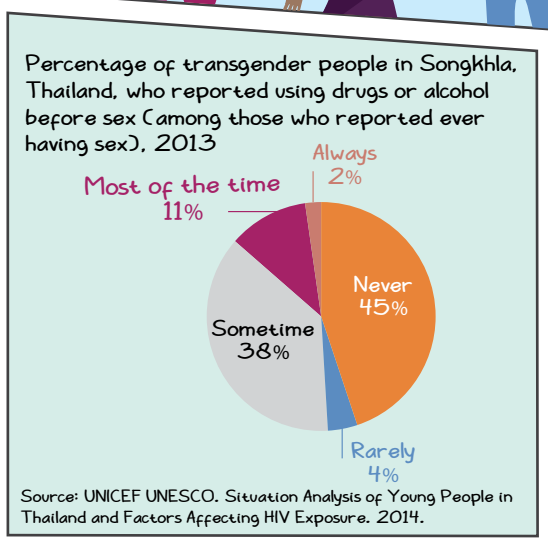
At the bottom of the chart I see the source, which is where I would look to find out more about the data.

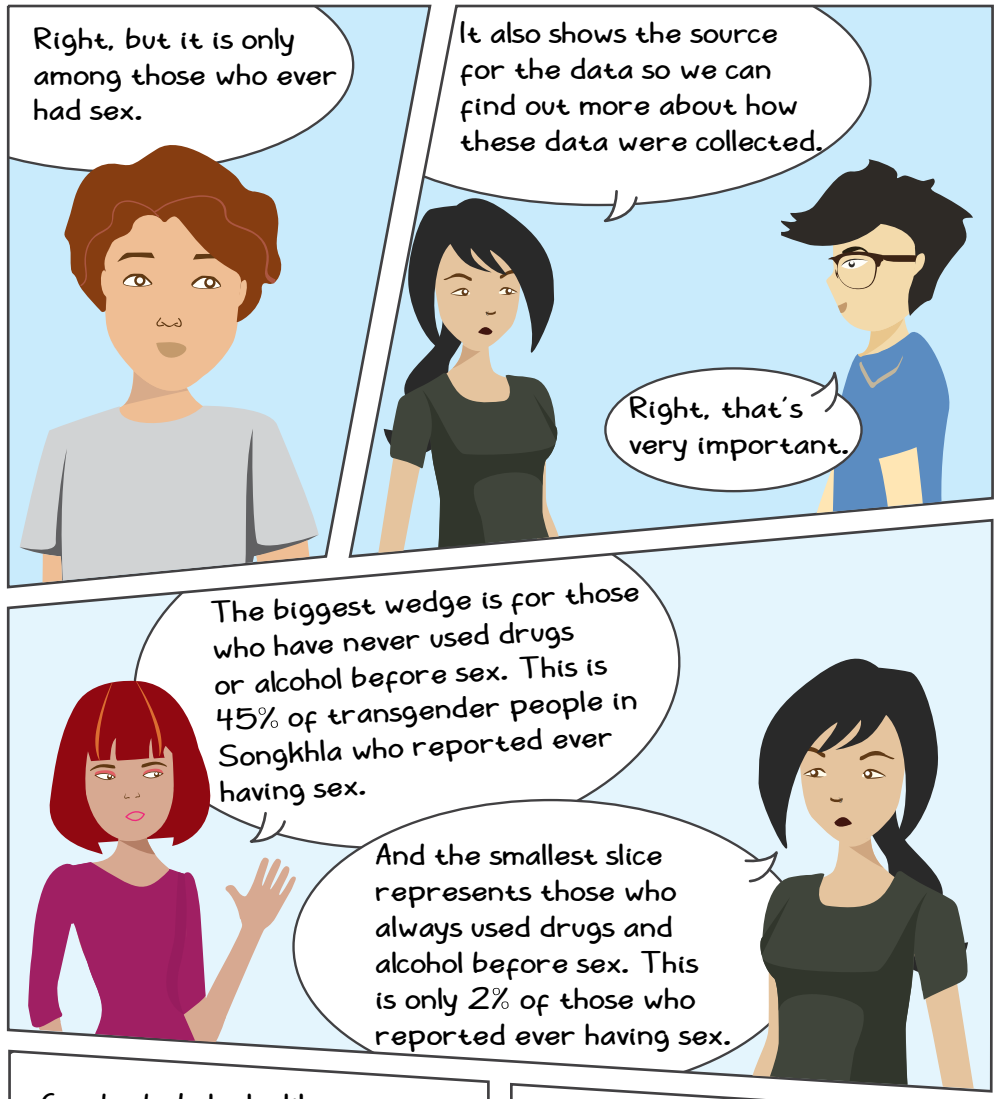
C. Reading a Pie Chart

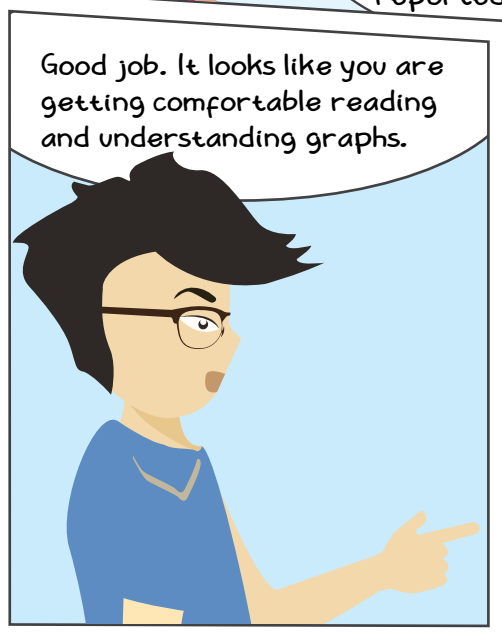
Great job you two! Now let's discuss pie and your average pie chart. A pie is a dessert that comes in the shape of a circle, like a pizza. Just like a pizza, a pie can be divided into slices that can be different sizes. In a pie chart, the different size slices represent different percentages. So if there were a pie with two equally sized slices, each would represent 50/. Does that make sense?



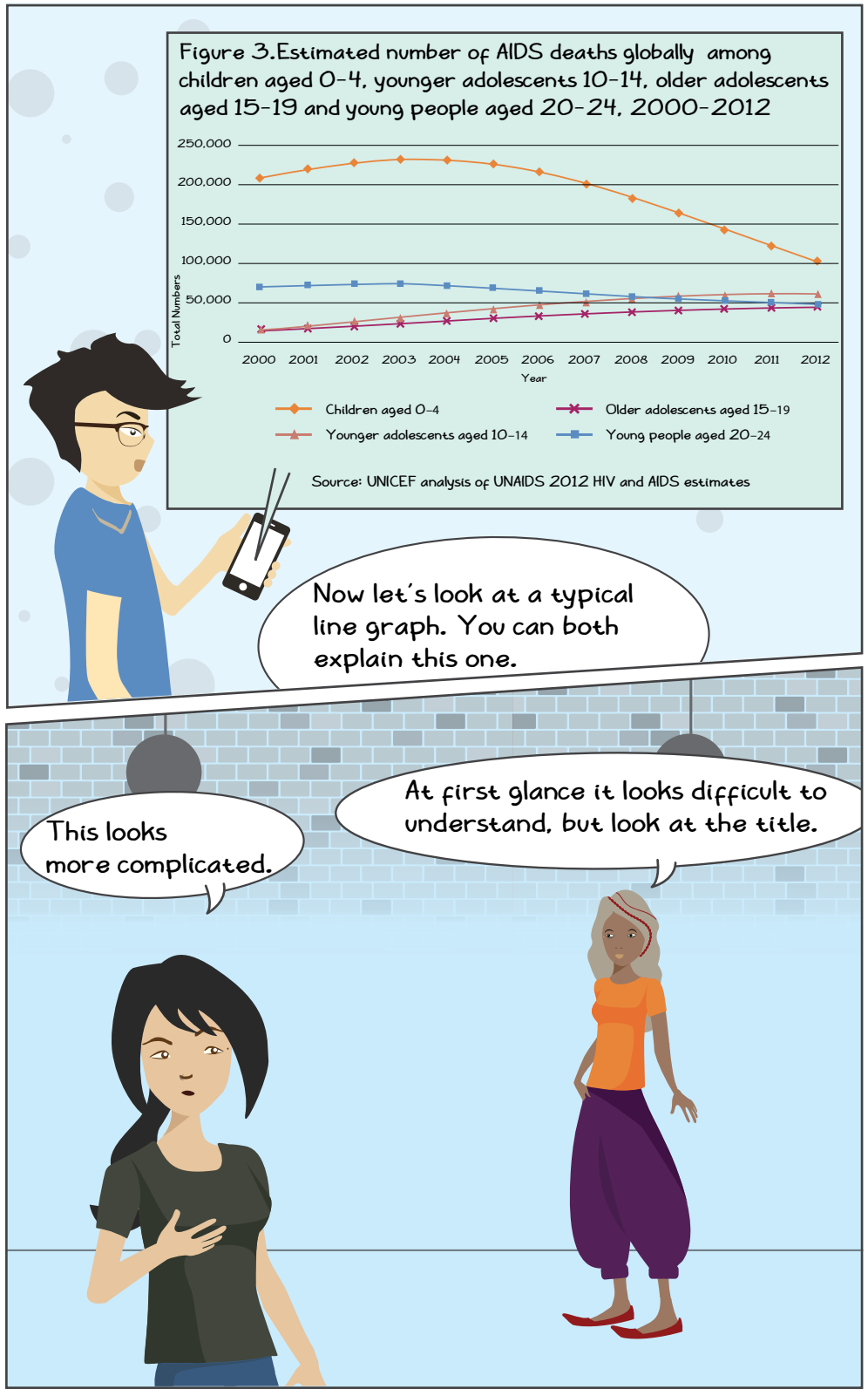


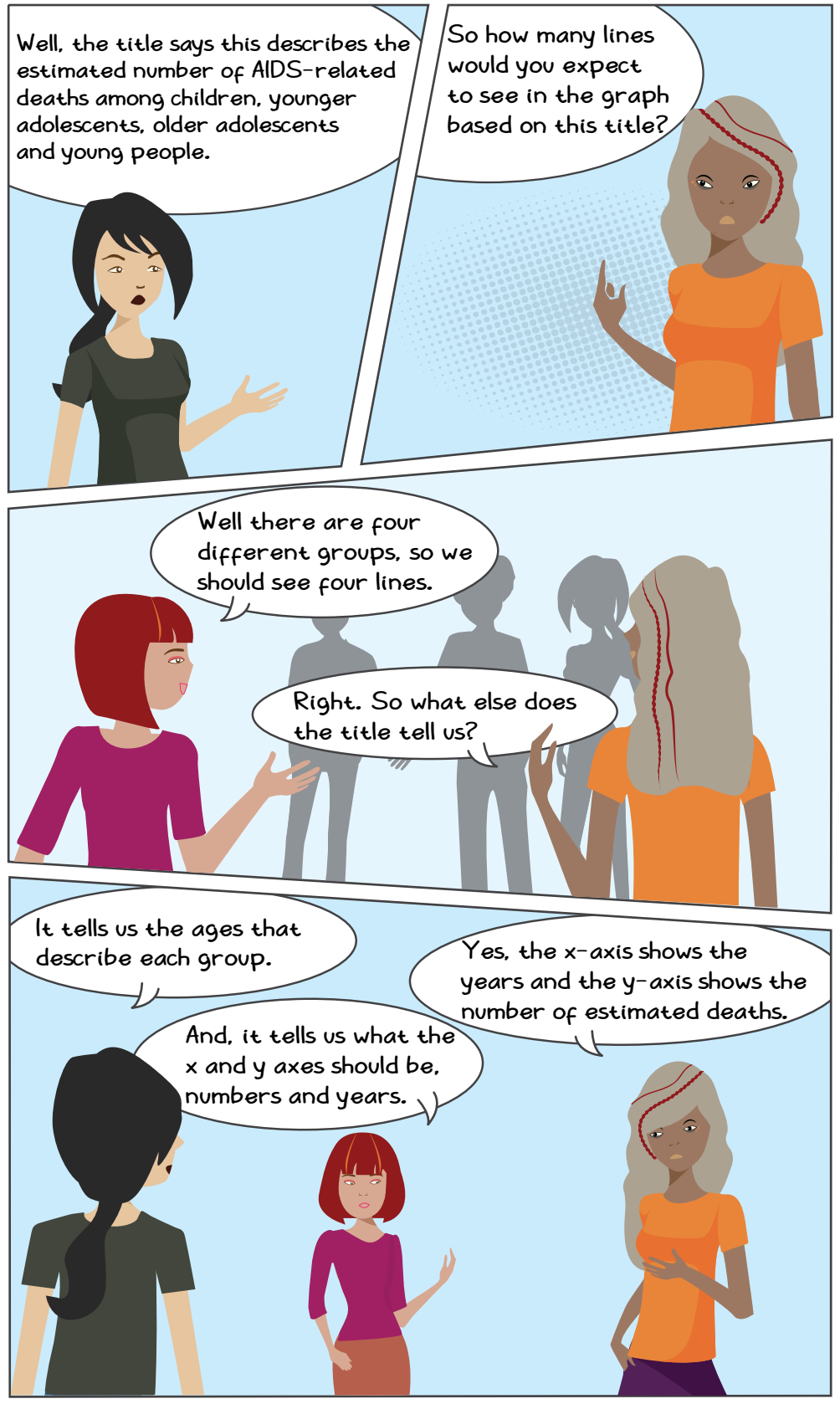






D. Reading a Line Graph



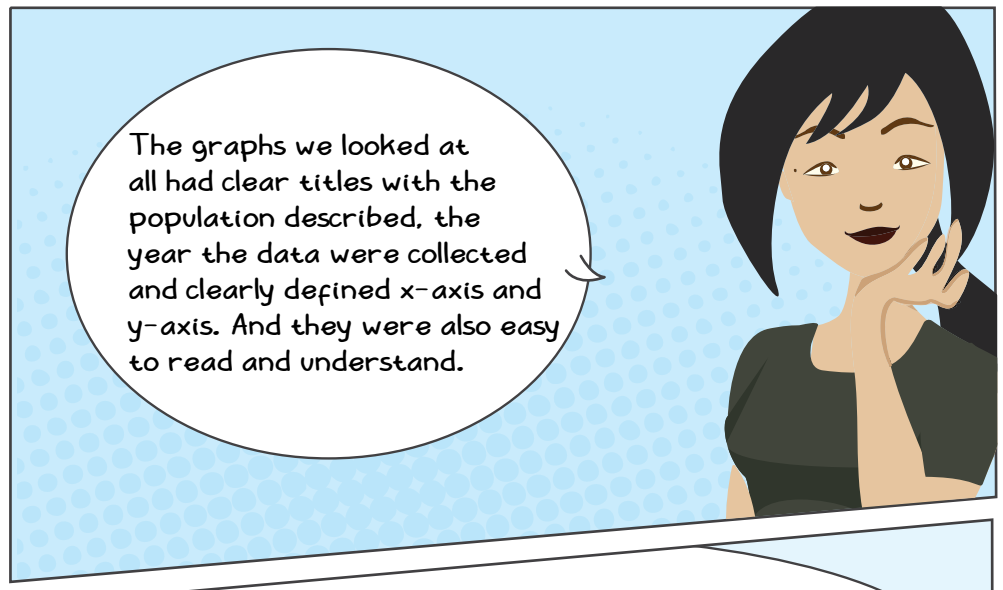


PRODUCING GRAPHS FOR INFORMATION-SHARING AND ADVOCACY

In this section, Tran, Lucia, Aran, John and Bindu discuss what types of graphs are useful for which types of data, how to present data to convey information and how to use data to advocate for adolescents and young key population at risk.

Now we're on one of our final topics that we need to discuss, in order for you two to be informed and knowledgeable youth representatives! Let's discuss how we can present data so we can participate in HIV and AIDS policy and decision making. Designing good tables and graphs is essential to conveying information and making a point. What are some of the key things you saw in all the charts we discussed already?

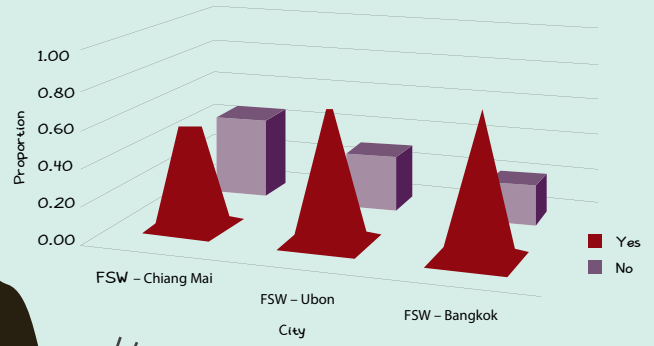
Great so far!

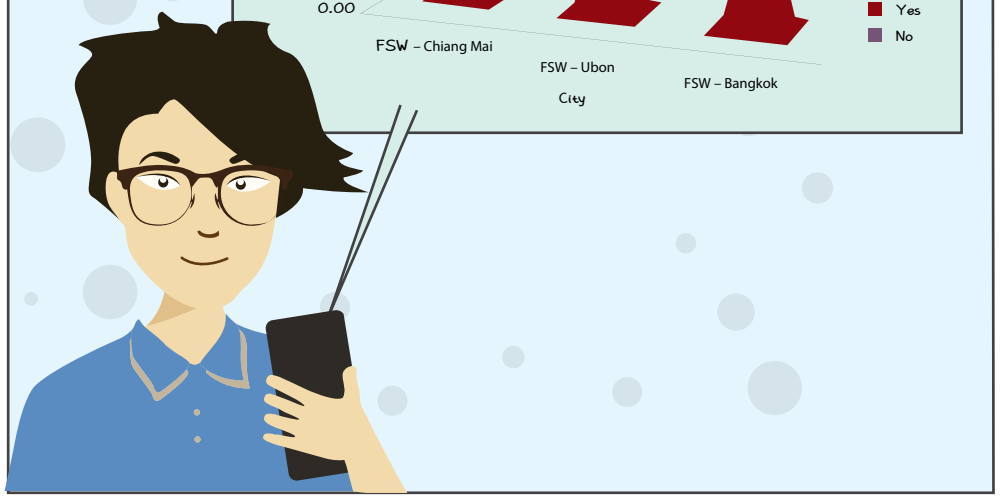


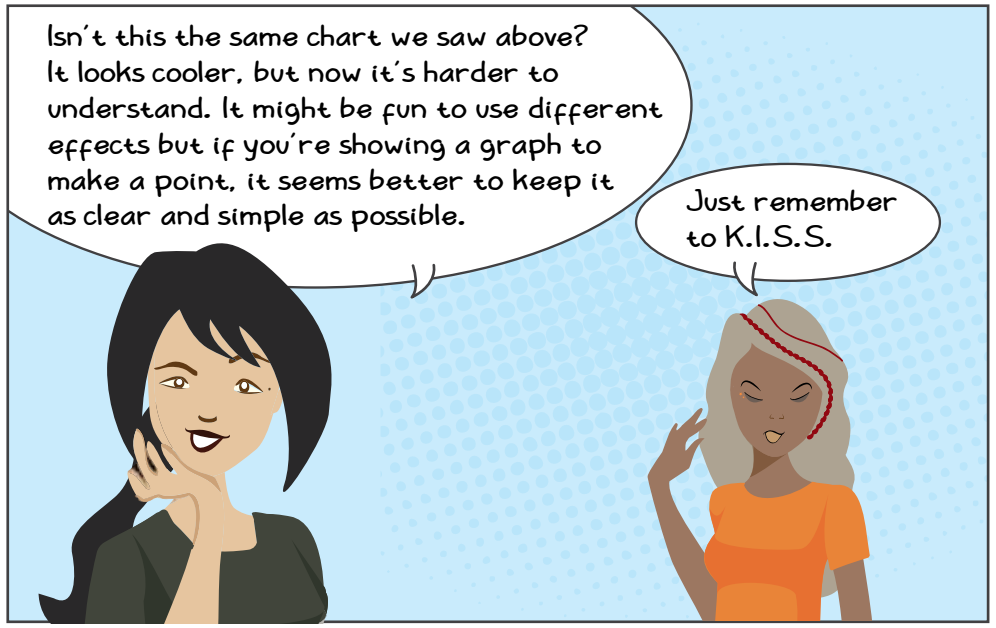
The graphs we looked at all had clear titles with the population described, the year the data were collected and clearly defined x-axis and y-axis. And they were also easy to read and understand.

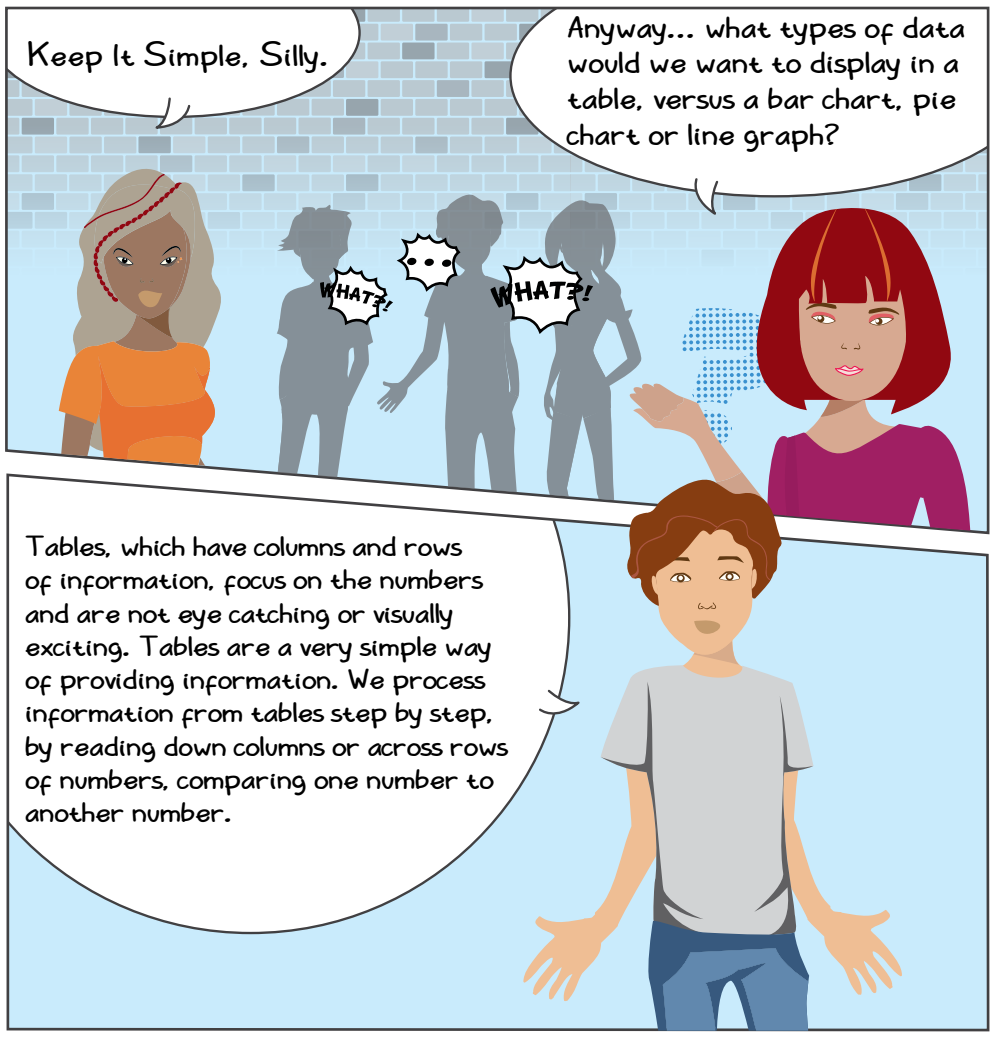
Good points Lucia. Many software programs will let you use special features like 3D effects, but these can make understanding the graph difficult. Look at this chart.

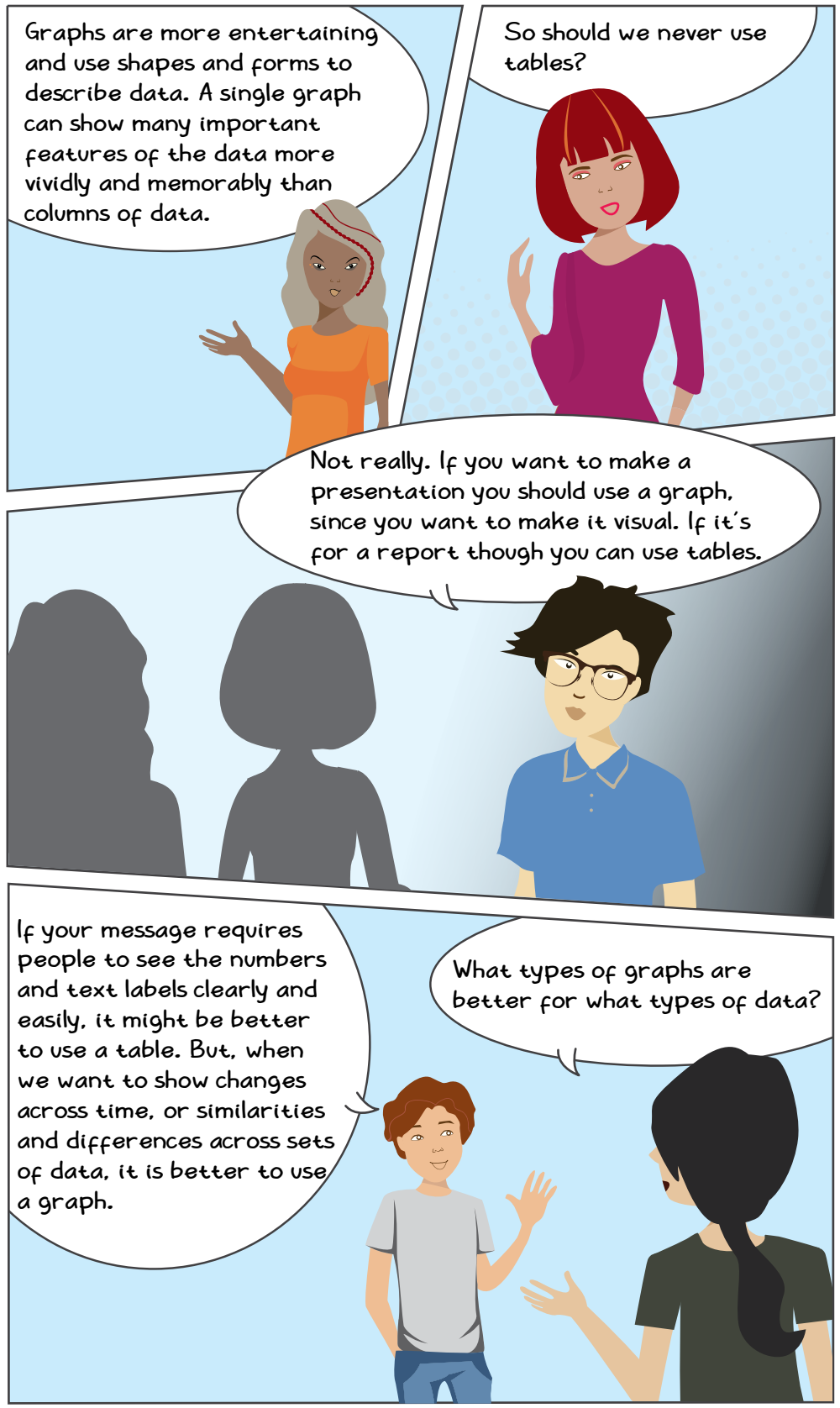
Figure 4. Proportion of Young Female Sex workers who talked to anyone about HIV/AIDS in the past year in Chiang Mai, Ubon and Bangkok, Thailand (2013)






Isn't this the same chart we saw above? It looks cooler, but now it's harder to understand. It might be fun to use different effects but if you're showing a graph to make a point, it seems better to keep it as clear and simple as possible.
Just remember to K.I.S.S.


Keep It Simple, Silly.
Anyway... what types of data would we want to display in a table, versus a bar chart, pie chart or line graph?
WHAT?!
WHAT?!
Tables, which have columns and rows of information, focus on the numbers and are not eye catching or visually exciting. Tables are a very simple way of providing information. We process information from tables step by step, by reading down columns or across rows of numbers, comparing one number to another number.

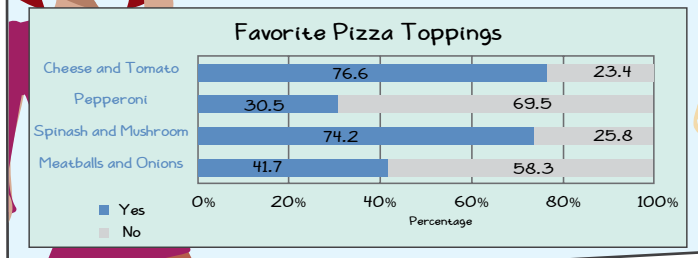

Graphs are more entertaining and use shapes and forms to describe data. A single graph can show many important features of the data more vividly and memorably than columns of data.
So should we never use tables?
Not really. If you want to make a presentation you should use a graph, since you want to make it visual. If it's for a report though you can use tables.
If your message requires people to see the numbers and text labels clearly and easily, it might be better to use a table. But, when we want to show changes across time, or similarities and differences across sets of data, it is better to use a graph.
What types of graphs are better for what types of data?

A bar chart is the most common type of chart. It helps to show information about separate categories or groups. This means that each person in the study or population can belong to one and only one group, like age groups. Separate categories could be 0-5 years, 6-10 years, and so on.

Most of them have vertical bars which means the taller the bar the larger the category.

But I have seen horizontal bar charts which show the longer the bar the larger the category.

That's right. Vertical bars are useful when the different categories have long titles or when there are many different categories that would not fit into a horizontal bar chart. They are basically the same but a vertical bar chart is flipped onto its side.



What about pie charts?

Pie charts are good for showing proportional data that when added up is equal to 100%. It is helpful to shade in the slices with different colours or patterns so that you can easily tell one slice from another. To make sure the chart is easy to read and understand, you should have no more than 6 or 7 slices per pie.

Stop making me so hungry by talking about pies.

And what about a line graph?

Line graphs are usually used to show changes over time, for example in a behaviour or characteristic, like knowledge about HIV transmission, during a certain period of time. Usually the x-axis represents a variable, such as year or month, and the y-axis represents the data for that time period.

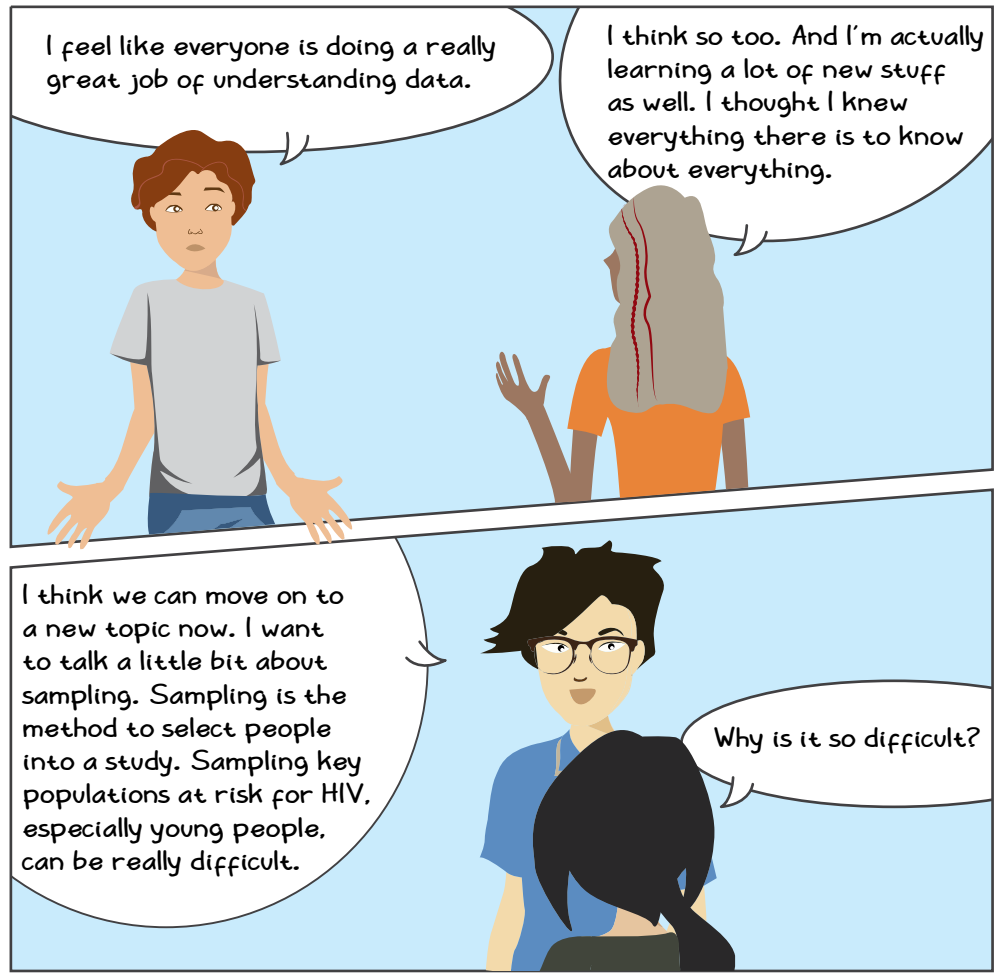
So I guess it's good to have at least three points in order to show change.

Excellent! Otherwise you are just comparing two points and this is not really a trend.

The final point I want to make about this topic is that some data are not displayed well. Sometimes the titles are vague or the graphs do not have the x- or y-axis properly labelled. You should never feel afraid about asking questions about a graph or data that you do not understand.

QUESTIONING DATA

In this section, Tran, Lucia, Aran, John and Bindu are discussing how data are gathered and are coming up with questions to ask in order to understand data. The discussion points include sampling methods (representative or not representative), sample measurement (who do the data really represent) and questions to ask yourself or a presenter about data on HIV and behaviours.



Think about it first and then tell me why.

Well, key populations are part of groups that other people judge.

I know a lot of people who use drugs and they could be put in jail if someone found out, so they keep it a secret.

If some of my family knew I liked boys, they would be very upset, so I keep it a secret from them. I understand why people from key populations feel like they need to hide from others.

These are all very good reasons why sampling adolescents and young people at higher risk of HIV exposure is so hard.

In sampling, you want to collect a small portion of people from the population that will actually represent that entire population.

Explain that, please.

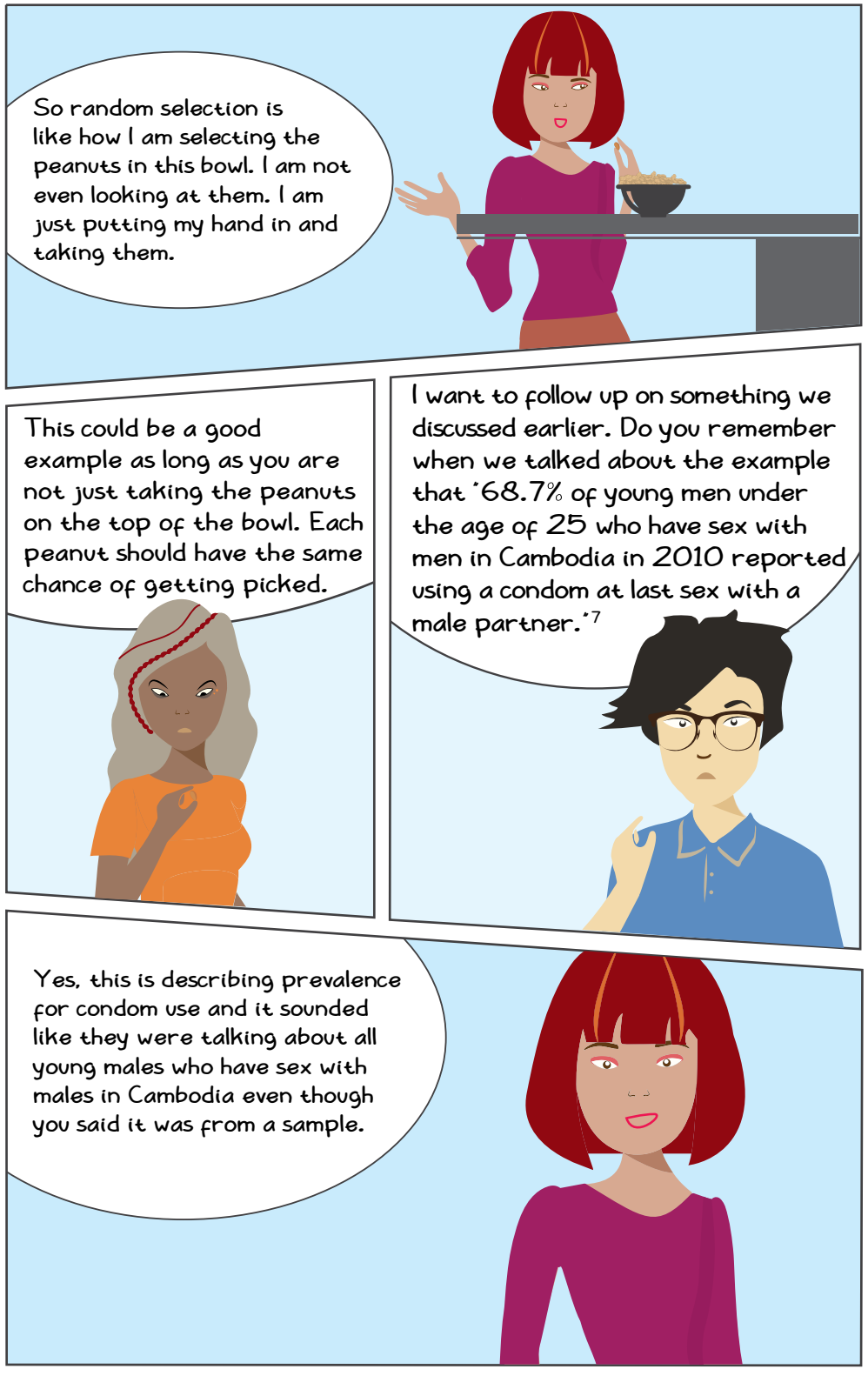
I think I get it. If you want to get a sample of people who inject drugs you should try to select them so that they are like the people who inject drugs who you did not sample. Both the people who inject drugs you sample and the people who inject drugs you do not sample are part of the whole population of people who inject drugs.

So then you do not have to ask every single person who injects drugs about their behaviours to be able to describe the behaviours of all people who inject drugs, right?

So how do you do that?

The best way to do this is to pick people randomly. There are many types of sampling methods for doing this, but what you really want to know is whether a study used random sampling[6].

[6]Also known as probability sampling.



[7]Cambodia National AIDS Authority. (2012) Cambodia AIDS Response Progress Report 2012. Phnom Pehn. National AIDS Authority. Accessed at: http://www.unaids.org/en/dataanalysis/knowyourresponse/countryprogressreports/2012countries/ce_KH_Narrative_Report[1].pdf

Exactly. This is why it's always good to question statistics. A sample is one small part of the larger population that you study. For instance, the population described in Cambodia was young males who have sex with males. It is important to know how the sample is formed in order to know if it can really say something truthful about the larger population.

So it is describing just a part of the population of young males who have sex with males and not all young males who have sex with males in the whole country?

Do you think they just gathered the information from young males who have sex with males in just one city?

And, what if they only asked this question of a specific group of young males who have sex with males, like those who only sell sex for money, rather than from young males who have sex with males in general?

Could they have only asked a dozen young males who have sex with males?

Both of you are asking the right questions! I happen to know that this survey was conducted in eight urban areas, that the prevalence for young males who have sex with males was estimated on a subset of a larger sample of 1,168 males who have sex with males, and that the people in the sample were not selected in any random way[8].

With that extra info we should understand the prevalence of condom use among young males who have sex with males in Cambodia in a very different way.

Yeah, you're right. It might be better to say something like 68.7% of a sample of young men under the age of 25 who have sex with men in eight urban areas in Cambodia in 2010 reported using a condom at last sex with a male partner.

[8]Bros Khmer 2010. Behavioral Risks On-Site Serosurvey among At-Risk Urban Men in Cambodia. Accessed at: http://www.fhi360.org/sites/default/files/media/documents/Bros%20Khmer-2012%20Behavioral%20Risks%20On-Site%20Serosurvey%20Among%20At-Risk%20Urban%20Men%20in%20Cambodia.pdf

That's good since we do not know if young urban males who have sex with males are more likely to use condoms than rural males who have sex with males.

Also, since they were not selected in a random way they may not be like the population of urban males who have sex with males but may be a specific type of male who have sex with males, maybe those who use condoms more.

So the questions you want to ask when presented with statistics from surveys are 'who do they represent?', 'where was the study conducted?' and 'how was the population described?'

You also want to ask, 'how many people does the study represent?'. If it is only a few people, then it may be hard to understand the data.

And then you want to also know how the people in the sample were selected, because if they selected themselves or volunteered to participate, they might not be like other young males who have sex with males.

These are all excellent points. Sampling is a very important part of statistics and it is great that you guys now understand it so well.

Right, Tran. Well I think we covered just about everything. Lucia and Aran, I had a lot of fun going over all these topics with you two! We should all do this again sometime.

Sure, how about in ten years?

?!

Well thank you, John, Tran, and Bindu! All the information you gave us was all very helpful and I'm so glad that you guys were here to walk us through everything.

Yes, thank you all so much. Now we're prepared for the meeting and maybe even to be the next leaders to eliminate HIV!

THANK YOU

Questions to ask yourself or someone presenting statistics on HIV

Who does this represent

If you are given a statistic about the opinions of urban young males who have sex with males, it's probably not going to be useful for understanding rural young males who have sex with males. This could be understood by asking about the rules or guidelines used to select people into the study (i.e., what was the eligibility criteria you used for enrolment in the study?). Also ask when the survey took place if this information is not already provided.

How many people does it represent

Most studies should rely on some type of calculation to ensure that the sample has enough people in it to conduct meaningful analysis. This could be understood by asking whether any calculation was used to determine the size of the sample.

How were the people selected

Let's say that there was the statistic that 90 per cent of young people who inject drugs share needles but then you ask about how the data were gathered. You are told that the young people who inject drugs were not randomly selected but from clients of an NGO that has been providing information about safe injection practices for five years. Would you understand this in the same way if they had collected the data using a random sampling method?

APPENDIX

A. Test your knowledge

1. Put the number of the correct definition with the terms below (answers provided in Appendix D)

Incidence	_____
Rate	_____
Prevalence	_____
Percentage	_____
Proportion	_____

a. A quantity, amount or measure considered as a portion or part of another quantity, amount or measure (find example from literature).
b. The part of a population that has a characteristic or disease at a specific time such as (find example from literature).
c. The number of new HIV infections and during a specific time period.
d. A portion or part in its relation to the whole such as (find example from literature).
e. A part in relation to a whole (which is usually the amount per hundred) find example from literature).

2. In 2007, 0.18% of Indian young adults (aged 20-24) were living with HIV[9]. Is this:
   a. Incidence
   b. Prevalence

3. In 2012, 12% of young Thai MSM aged 15 to 21 became newly infected with HIV[10]. Is this:
   a. Incidence
   b. Prevalence

4. Someone tells you that in a study of 1000 transgender persons that 550 of them had correct knowledge of HIV transmission. You want to know the percentage. How do you calculate this?
   a. 1000 multiplied by 550?
   b. 550 divided by 1000?
   c. 550 divided by 1000 multiplied by 100?
   d. 1000 minus 550?

[9] International Institute for Population Sciences (IIPS) and Macro International (2007) 'National Family Health Survey (NFHS-3) 2005-06', India: Volume 1 - See more at: http://www.avert.org/india-hiv-aids-statistics.htm#footnote7_dep7gzs
[10] http://www.irinnews.org/report/98439/thailand-s-msm-face-alarming-hiv-rates

5. You read that in a study of 700 young PWID that 12 of them are sharing needles and syringes. What is the percentage?
    a. 0.017%
    b. 58.3%
    c. 1.7%
    d. Not enough information to get a percentage

6. If, over the course of one year, ten young male sex workers are diagnosed with HIV, out of a total study population of 400 (who do not have HIV at the beginning of the study period), what would be the incidence of HIV in this population?
    a. 0.025.
    b. 2,500 per 100,000 male sex workers (per one year of study)
    c. 25 per 1000 male sex workers (per one year of study)
    d. 2.5%
    e. All of the above

7. The ideal sampling characteristics for a study is:
    a. Representative
    b. Random
    c. Large enough for the researchers to be able to draw meaningful conclusions
    d. All of the above

8. Among the 160 eligible PWID participants recruited in Shanghai methadone clinics in 2009-2010, the prevalence of Hepatitis C infection was 51.3%[11]. What can we tell from this study?
    a. That Hepatitis C is affecting mostly PWID in methadone clinics
    b. That PWID in methadone clinics in Shanghai in 2009-2010 have a Hepatitis C infection prevalence above 50%
    c. That HIV infection is a problem among PWID in Shanghai
    d. All of the above

9. What is NOT a key characteristic of a good graph?
    a. Clear and complete title
    b. Fancy designs
    c. Chart type that represents the data well
    d. Clear labelling of the x and y axis
    e. Clear labelling of the indicators

10. What type of graph would you use to compare condom use among young male sex workers of different cities?
    a. Pie chart
    b. Bar chart
    c. Line graph

11. What type of graph would you use for representing the proportion of types of places where MSM between the ages of 15 and 19 get condoms in your country?
    a. Pie chart
    b. Bar chart
    c. Line graph

12. What type of graph would you use to show the trend of HIV infection among young PWID in your country?
    a. Pie chart
    b. Bar chart
    c. Line graph

[11] Hser Y et al. Hepatitis C among methadone maintenance treatment patients in Shanghai and Kunming, China. Journal of Public Health. 2012. 34(1):24–31. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3285118/

B. Population terms

Adolescents: Anyone between the ages of 10 and 19 years.

Child/Children: Anyone under the age of 18 years, unless the laws of a particular country set a younger legal age for adulthood

Key populations at higher risk of HIV exposure: Anyone who is most likely to be exposed to HIV or to transmit it. These populations normally include people who inject drugs, men who have sex with men, sex workers and their partners, and transgender persons.

Youth: Anyone between the ages of 15 and 24 years.

Young people: Anyone between the ages of 10 and 24 years.

Young Key populations at higher risk of HIV exposure: Anyone between the ages of 10 and 24 years who is most likely to be exposed to HIV or to transmit it.

C. Terms about statistics

Denominator: The lower portion of a fraction used to calculate a rate or ratio. In a rate, the denominator is usually the population at risk, such as young persons who inject drugs. See numerator.

Epidemic: When more cases of a disease than expected occur in a given area or among a specific group of people over a particular period of time.

Epidemiology: The study of the measurement of health-related characteristics or behaviours (e.g., condom use, HIV status, etc.) in specified populations. Epidemiology is used to understand and respond to health related issues.

Graph: A way to show quantitative data (measurements of quantity such as number or percentage of people who use condoms) visually, often using two or more points such as percentage by characteristic or behaviour.

Incidence: Number of new cases in a fixed time period divided by the number of people at risk (A measure of the frequency with which an event, such as a new case of illness, occurs in a population over a period of time. The denominator is the population at risk; the numerator is the number of new cases occurring during a given time period.)

Formula for incidence rate = Number of onsets/Number at risk during a given time period x 100

Indicator: A measure that reflects, or indicates, the characteristic or behaviour of persons in a defined population (e.g., HIV status, HIV testing in the past 12 months, condom use).

Mortality rate: Mortality is another term for death. The mortality rate is the number of deaths occurring in a population in a specific period (usually a year) divided by the number of persons at risk of dying during that period. The mortality rate is typically expressed in number of deaths per 1,000 or 100,000 individuals per year.

Numerator: The upper portion of a fraction used to calculate a rate or ratio. In a rate, the numerator is usually the subgroup of population at risk that has a characteristic or behaviour (e.g., condom use, HIV status, etc.), such as the number of young persons who inject drugs who are female (versus male). See numerator.

Percentage: proportion or rate per hundred parts. Basically, this could be the number of people in a sample with a characteristic or behaviour (e.g., condom use, HIV status, etc.) over the total number of people sampled multiplied by 100. See prevalence.

Population: A group of people with a common characteristic; target population is a population for which you would like to make some conclusions.

Proportion: A part considered in relation to the whole. Basically, this could be the number of people in a sample with a characteristic or behaviour (e.g., condom use, HIV status, etc.) over the total number of people sampled, but not multiplied by 100. See prevalence.

Prevalence: Number of people with the characteristic or behaviour (e.g., condom use, HIV status, etc.) at a given time divided by the number of people could possibly have this characteristic, attribute or behaviour. Usually prevalence is measured in a sample, for instance the number of people in a sample who have HIV infection over the total number of people in the sample.

Formula for prevalence proportion = Number of cases/Number of individuals in the study
Formula for prevalence percentage = Number of cases/Number of individuals in the study x 100

Random sample: A sample collected by selecting participants such that each person of a given population has the same chance of selection.

Rate: a measure of the number of times a characteristic, attribute, behaviour or event occurs in a defined population in a defined time

Representative sample: A sample collected in such a way to be able to understand or something meaningful about the population that was sampled.

Statistics: Branch of mathematics dealing with collecting, organizing, and understanding data.

Sample: A selected subset of a population. A sample may be similar to the population it wants to say something meaningful about or not. If a sample is representative it can be used to understand and say something about the population. If the sample is non-representative it leaves out certain groups from the population and cannot be used to understand and say something the population.

Trend: A long-term movement or change in a statistic, usually over time. Trends can move upwards, downwards or remain flat over time.

Variable: Any characteristic, attribute or behaviour (e.g., condom use, HIV knowledge, etc.) that can be measured.

D. Test your knowledge: Answers

1. Put the number of the correct definition with the terms below

Incidence	___c___
Rate	___a___
Prevalence	___b___
Percentage	___e___
Proportion	___d___

a. A quantity, amount or measure considered as a portion or part of another quantity, amount or measure (find example from literature).
b. The part of a population that has a characteristic or disease at a specific time such as (find example from literature).
c. The number of new HIV infections and during a specific time period.
d. A portion or part in its relation to the whole such as (find example from literature).
e. A part in relation to a whole (which is usually the amount per hundred) find example from literature).

2. In 2007, 0.18% of Indian young adults (aged 20-24) were living with HIV[12]. Is this:
   a. Incidence
   b. Prevalence

3. In 2012, 12% of young Thai MSM aged 15 to 21 became newly infected with HIV[13]. Is this:
   a. Incidence
   b. Prevalence

[12] International Institute for Population Sciences (IIPS) and Macro International (2007) 'National Family Health Survey (NFHS-3) 2005-06', India: Volume 1 - See more at: http://www.avert.org/india-hiv-aids-statistics.htm#footnote7_dep7gzs
[13] http://www.irinnews.org/report/98439/thailand-s-msm-face-alarming-hiv-rates

4. Someone tells you that in a study of 1000 transgender persons that 550 of them had correct knowledge of HIV transmission. You want to know the percentage. How do you calculate this?
   a. 1000 multiplied by 550?
   b. 550 divided by 1000?
   c. 550 divided by 1000 multiplied by 100?
   d. 1000 minus 550?

5. You read that in a study of 700 young PWID that 12 of them are sharing needles and syringes. What is the percentage?
   a. 0.017%
   b. 58.3%
   c. 1.7%
   d. Not enough information to get a percentage

6. If, over the course of one year, ten young male sex workers are diagnosed with HIV, out of a total study population of 400 (who do not have HIV at the beginning of the study period), what would be the incidence of HIV in this population?
   a. 0.025.
   b. 2,500 per 100,000 male sex workers (per one year of study)
   c. 25 per 1000 male sex workers (per one year of study)
   d. 2.5%
   e. All of the above

7. The ideal sampling characteristics for a study is:
   a. Representative
   b. Random
   c. Large enough for the researchers to be able to draw meaningful conclusions
   d. All of the above

8. Among the 160 eligible PWID participants recruited in Shanghai methadone clinics in 2009-2010, the prevalence of Hepatitis C infection was 51.3%[14]. What can we tell from this study?
   a. That Hepatitis C is affecting mostly PWID in methadone clinics
   b. That PWID in methadone clinics in Shanghai in 2009-2010 have a Hepatitis C infection prevalence above 50%
   c. That HIV infection is a problem among PWID in Shanghai
   d. All of the above

9. What is NOT a key characteristic of a good graph?
   a. Clear and complete title
   b. Fancy designs
   c. Chart type that represents the data well
   d. Clear labelling of the x and y axis
   e. Clear labelling of the indicators

10. What type of graph would you use to compare condom use among young male sex workers of different cities?
    a. Pie chart
    b. Bar chart
    c. Line graph

11. What type of graph would you use for representing the proportion of types of places where MSM between the ages of 15 and 19 get condoms in your country?
    a. Pie chart
    b. Bar chart
    c. Line graph

12. What type of graph would you use to show the trend of HIV infection among young PWID in your country?
    a. Pie chart
    b. Bar chart
    c. Line graph

[14] Hser Y et al. Hepatitis C among methadone maintenance treatment patients in Shanghai and Kunming, China. Journal of Public Health. 2012. 34(1):24–31. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3285118/