



Ensuring Data Accuracy

When sharing data as part of a communications plan, take steps to ensure you are sharing the data accurately. This handout provides tips on how to avoid common mistakes. To quote The United Nations Children's Fund, Multiple-Indicator Survey Handbook (2006): "These numbers, whether they are correct or not, will take on a life of their own once released...and may cause you a lot of trouble if you have not checked your arithmetic carefully."

When we report data, we often share "how many" (percentage, number) of "who" (youth, parents, community members) said "what" (reasons for using marijuana, sources of alcohol). Such statements provide multiple possibilities for error.

How many?

- Avoid adding percentages inappropriately. For example, 14% of boys reporting substance use and 12% of girls reporting substance use does not mean that 26% of youth reported use. Similarly, 20% of youth reporting alcohol use, 10% reporting marijuana use, and 5% reporting other drug use does not mean that 35% of youth have used a substance. It may be the case that the 20% of used alcohol also used marijuana and other drugs. You need to determine the correct size of the total pie from which you are showing a slice.
- Similarly, avoid calculating an average using two percentages. Let's say one school in the community had 17% of youth reporting vaping, while another school in the community had a rate of 23%. Simply adding 17% and 23%, and then dividing by two, would only be valid if each school had the same number of students participating in the survey and answering that specific question about alcohol use. If one school has a much larger survey sample size, the average will be closer to that school's rate.
- Both numbers and percentages are important to share. Twenty youth out of 25 reporting greater educational engagement compared to 20 youth out of 200 makes a big difference. Conversely, 5% of youth reporting homelessness may seem like a low percentage, but that's still a large number if your statewide survey sample is 150,000.

Who?

- Make sure you describe who the respondents are. For example, rather than "women reported..." a clearer description would be "women, the majority of whom are White and between the ages of 30 and 50, report..." Similarly, clarify which grade levels are represented rather than simply saying 10% of students use marijuana. It makes a big difference if those "students" represent elementary and middle school students as well as high school students.
- Indicate who might be missing from the data. For example, saying "100% of parents who completed surveys..." may not be representative of all parents if the survey was only conducted among those who were able to attend the parent-teacher conferences during which time the survey was conducted. As another example, if school enrollment data show that 36% of students are African American, but only 12% of survey respondents identified as African American, it would be important to note that as a



limitation of the data. In this case, African American students would be under-represented in the data.

- “Missing” is especially important in a case in which people are systematically excluded.

What?

- Provide important context. For example, youth reporting “any alcohol use in the past 30 days” provides more clarification than youth reporting “drinking.” Further, reporting that “5% of youth bought alcohol at a bar or restaurant” would be missing important context if that is 5% out of only those youth who reported past 30-day alcohol use.
- Avoid overgeneralizing the findings. Ten percent of young adults reporting “five or more drinks in one evening” does not necessary equate to “10% of young adults have a drinking problem.”
- Avoid cherry-picking the data. It is important to share both the positive and areas needing improvement. For example: “While past 30-day alcohol use among high school seniors has decreased over the past 10 years, rates of past-year suicidal ideation have increased during that same period of time.”

