

# What is sampling bias?

---

*Bias* is a systematic error that can prejudice your evaluation findings in some way. So, *sampling bias* is consistent error that arises due to the sample selection. For example, a survey of high school students to measure teenage use of illegal drugs will be a biased sample because it does not include home schooled students or dropouts. A sample is also biased if certain members are underrepresented or overrepresented relative to others in the population. For example, distributing a questionnaire at the end of a 3-day conference is likely to include more people who are committed to the conference so their views would be overrepresented. Interviews with people who walk by a certain location is going to over-represent healthy individuals or those who live near the location. Selecting a sample using a telephone book will underrepresented people who cannot afford a telephone, do not have a telephone, or do not list their telephone numbers.

Sampling bias can occur any time your sample is not a random sample. If it is not random, some individuals are more likely than others to be chosen. Always think very carefully about which individuals are being favored and how they differ.

## **Why does it matter?**

Sampling bias means that the data you collect may not be accurate or represent the group.

**How can we know if the sample is biased?** Sometimes you can identify sampling bias just by being very thoughtful and comparing the characteristics of respondents in your sample to what you know about the population in general. Think about the demographic characteristics that might have an important relationship to their answers. For example, if you know that gender is an important variable, and you know that the population includes 50% males and 50% females, then the sample needs to include the same proportions. If the sample includes 20% males, your results are likely to be biased because you don't have enough responses from men.

**What can we do about it?** If you do identify differences between respondents in your sample and in the target population, be careful about how you report the results. Report the results in terms of who did respond. Don't imply that the results apply to anyone other than those who responded.