Local Health Department Academy of Science

Survey School

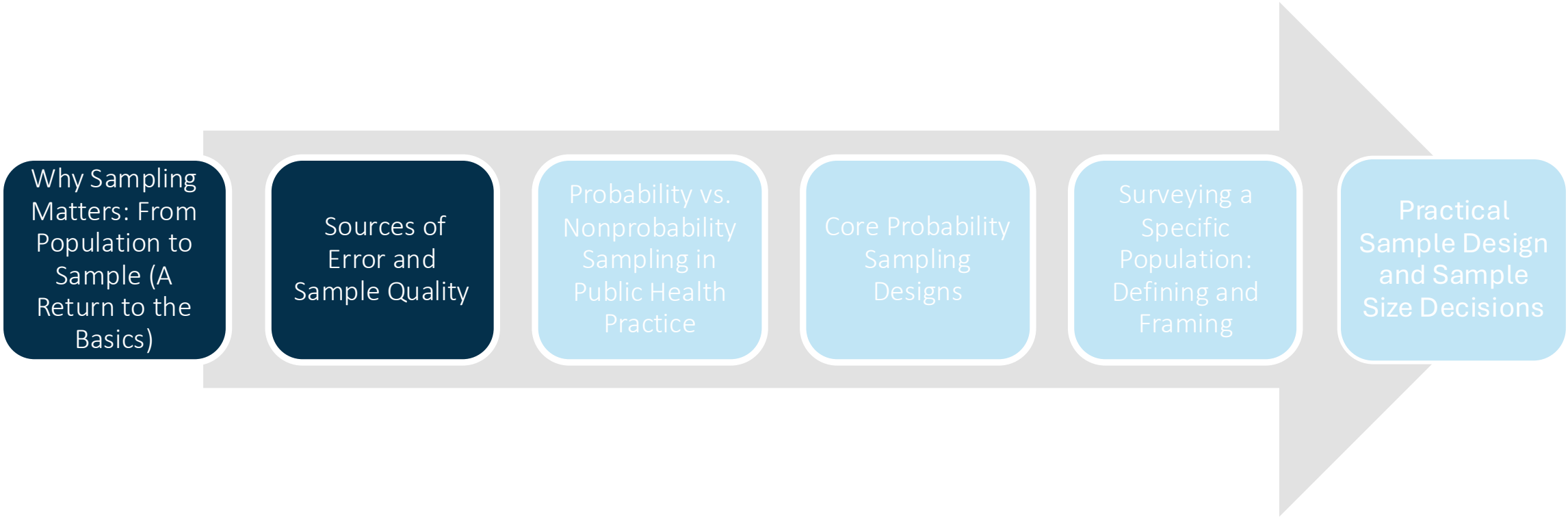
Lesson 02:

Sources of Error & Sample Quality

“Ensuring Survey Results Reflect the Communities we Serve”

Kevin Brooks, MSc, PhD

Course Overview

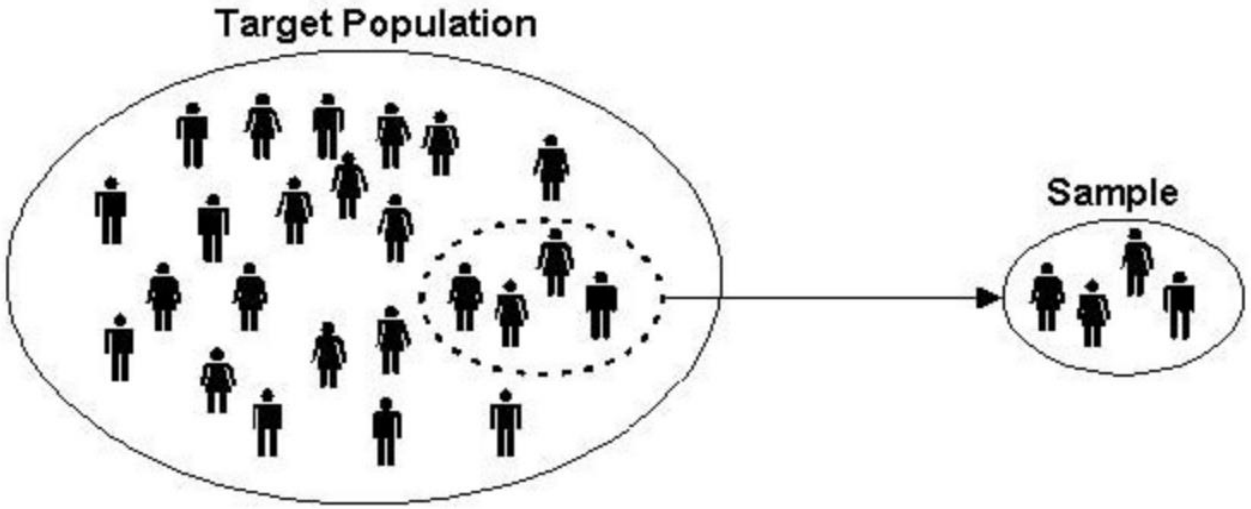


Objectives

- Differentiate among nonsampling error, sampling error, and sample bias.
- Identify common sources of bias in survey research.
- Evaluate how sampling decisions affect study quality and representation.

Special attention is given to strategies for representing hard-to-reach populations and ensuring that survey findings tell an accurate story about the communities being studied.

Population ➔ ➔ Sample



Vaccination

If a public health agency surveys community members and finds that 64% support school-based vaccination clinics, you want that percentage to be a reliable estimate of the views held by the broader population the agency serves.

Health services

If a survey of Medicaid beneficiaries shows that 72% would like more access to telehealth services, you want that result to accurately reflect the preferences of all beneficiaries, not just those who completed the survey.

There are three reasons

- **Sampling Error**
  - ✓ Random variation due to studying a sample
- **Nonsampling Error**
  - ✓ Interviewer, response, or coding errors
- **Sample Bias**
  - ✓ Systematic differences between the sample and population

Important: Bias is usually a larger threat than sample size.

Sampling Error

Occurs because a sample *does not perfectly* represent the population due to random chance, even when the sample is selected properly

- Every sample is slightly different
- Groups may be overrepresented/underrepresented by chance alone
- As a result, survey estimates vary from sample to sample
- This variation is normal and expected

Sampling error is NOT a mistake

Sampling Error

Example: A county health department wants to estimate influenza vaccination coverage among adults. Survey 100 adults and find that 48% report receiving a flu vaccine during the past year.

If select a different sample of 100 adults, might find 43% or 51% vaccinated instead.

These differences don't necessarily mean vaccination rates are changing. More often, they reflect random variation in who happened to be selected.

Note: different samples from the same population can produce slightly different results simply **due to chance**

Sampling Error in the context of Random Variation

If you flip a fair coin:

- You expect 50% heads and 50% tails.
- But in 10 flips you may get:
  - 4 heads, 6 heads, 7 heads
- Random chance affects the outcome

Similarly, survey estimates vary because different individuals are selected into the sample.

Different samples of residents may produce different estimates of:

Vaccination rates, Tobacco use, Access to healthcare, Mental health needs

Sampling Error The Role of Sample Size

Larger Samples = Less Sampling Error

As sample size increases:

- Estimates become more stable
- Sample results tend to be closer to the true population value
- Random variation decreases
- Confidence in survey estimates increases

Sampling Error

The goal of survey sampling is NOT TO eliminate sampling error, but to measure and minimize it.

Nonsampling Error

- ✓ Consists of all error sources unrelated to the sampling of respondents.
- ✓ General sources include interviewer error related to the administration of the survey, response error related
- ✓ to the accuracy of responses as given, and coding error related to the accuracy of responses as recorded

Note, these errors can sometimes have a larger impact on survey quality than sampling error.

Nonsampling Error

Three common types

- ✓ **Interviewer Error:** Occurs when surveys are not administered consistently
  - ✓ Standardized training and clear survey protocols are so important
- ✓ **Response Error:** occurs when respondents provide inaccurate answers
  - ✓ Especially important when measuring health behaviors
- ✓ **Coding Error:** Occurs after data collection when responses are entered, categorized, or processed incorrectly.

Nonsampling Error: Vaccination Survey Example

Suppose a state health department conducts a survey to estimate seasonal influenza vaccination coverage among adults.

The survey results indicate 70% report receiving a flu shot during the past year. At first glance, that seems straightforward

But consider the possible sources of nonsampling error:

- ✓ Some respondents may not accurately recall
- ✓ Others may report that they were vaccinated because they believe vaccination is the "right" answer.
- ✓ Interviewer might unintentionally clarify differently across respondents.
- ✓ Data processing, a few vaccination responses could be entered incorrectly.

Sample Bias

Sample Bias

- ✓ Systematic differences between the sample and population
- ✓ Occurs when the people included in a survey differ systematically from the population we want to understand
- ✓ Coverage Bias, Selection Bias, and Nonresponse Bias

Sample Bias _ Case Study

County Community Health Assessment _ Coverage Bias

- ✓ **Health-needs survey distributed**
  - ➢ Online only
  - ➢ Through social media and email lists
- ✓ **May be missing**
  - ➢ Older adults with limited internet access
  - ➢ Low-income households
  - ➢ Residents experiencing housing insecurity
  - ➢ Residents in areas with limited broadband
  - ➢ Individuals with limited English proficiency

Sample Bias _ Case Study

Potentially underestimates:

- ✓ Access-to-care barriers
- ✓ Transportation challenges
- ✓ Housing insecurity
- ✓ Food insecurity
- ✓ Healthcare affordability concerns

When some groups are underrepresented, their health needs may be underrepresented as well.

Sample bias is not just a statistical issue. It is an equity issue.

Sample Bias _ Selection Bias and Nonresponse Bias

Selection Bias

Some groups more likely to be selected

- ✓ Example Recruitment occurs primarily through healthcare organizations
  - ➢ Overrepresents people already connected to care
  - ➢ Underrepresents individuals with limited healthcare access

Nonresponse Bias

Some groups are less likely to respond.

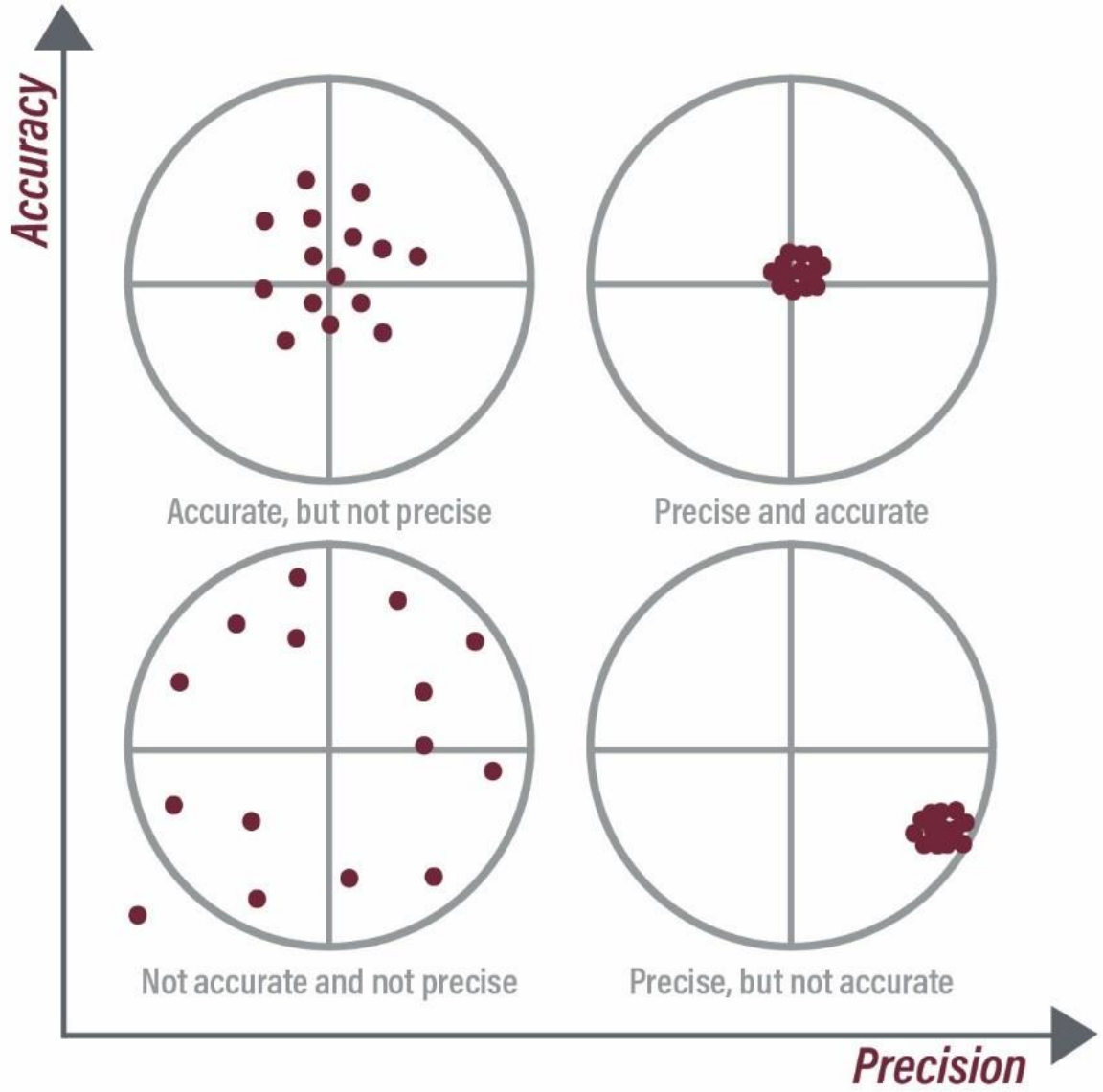
- ✓ Younger adults and working families respond at lower rates.
  - ➢ Community priorities may reflect only the most engaged residents

Sample Bias _ Representation v. Sample Size

Sample	Result
5,000 respondents, but underrepresents vulnerable populations	Potentially biased
500 respondents, representative of the community	More accurate

Most important question in survey research is not 'How many people responded?' It is 'Whose voices are represented in the data?

Sample Bias _ Representation v. Sample Size



Sources of Error _ Pulling it Together

Source of Error	What Causes It?	Can Larger Samples Fix It?	Public Health Example	Threat
Sampling Error	Random chance	✓ Yes, reduced with larger samples	Vaccination rate estimate varies from 43% to 51% across samples	Precision
Nonsampling Error	Data collection, response, or coding problems	X No	Respondents misreport vaccination status	Measurement
Sample Bias	Missing or underrepresented populations	X No	Community survey underrepresents residents experiencing housing insecurity	Representation

Sources of Error

If we don't hear from the right people, it doesn't matter how many people we hear from.

Thank you

~ Sources of Error & Sample Quality

"Ensuring Survey Results Reflect the Communities we Serve"~

Next →Lesson 3.

Probability vs. Nonprobability Sampling in Public Health Practice

Kevin Brooks, MSc, PhD