Local Health Department Academy of Science

Survey School

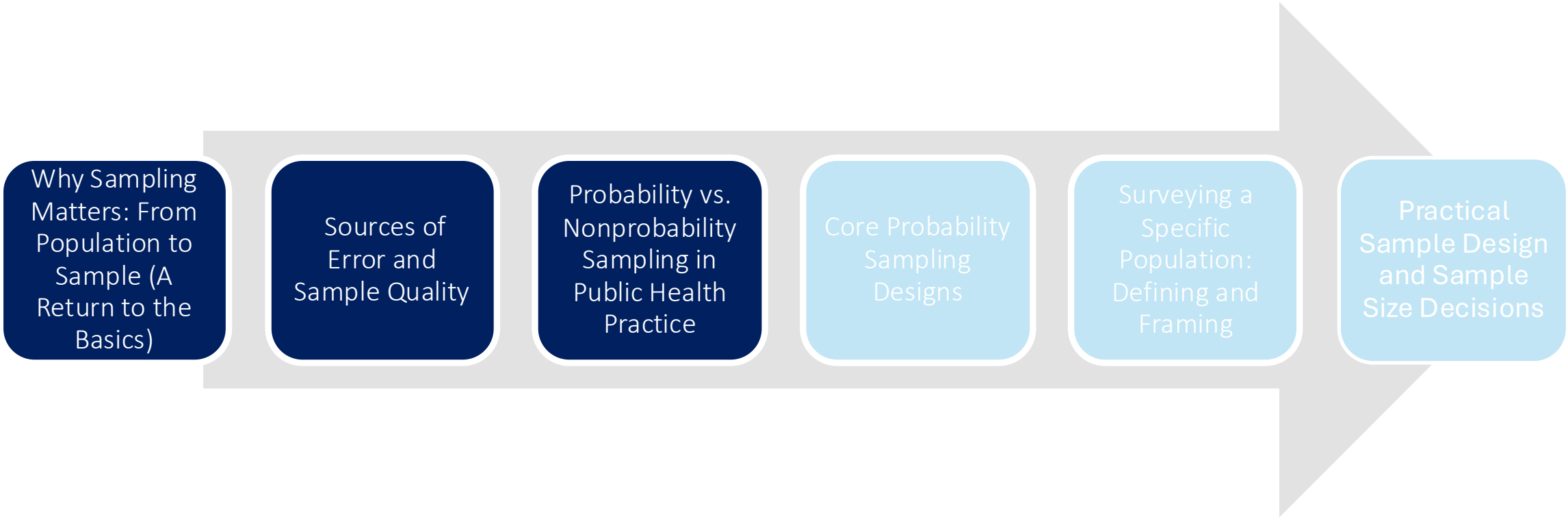
Lesson 03:

Probability vs. Nonprobability Sampling in Public Health Practice

“Building Reliable Public Health Evidence”

Kevin Brooks, MSc, PhD

Course Overview

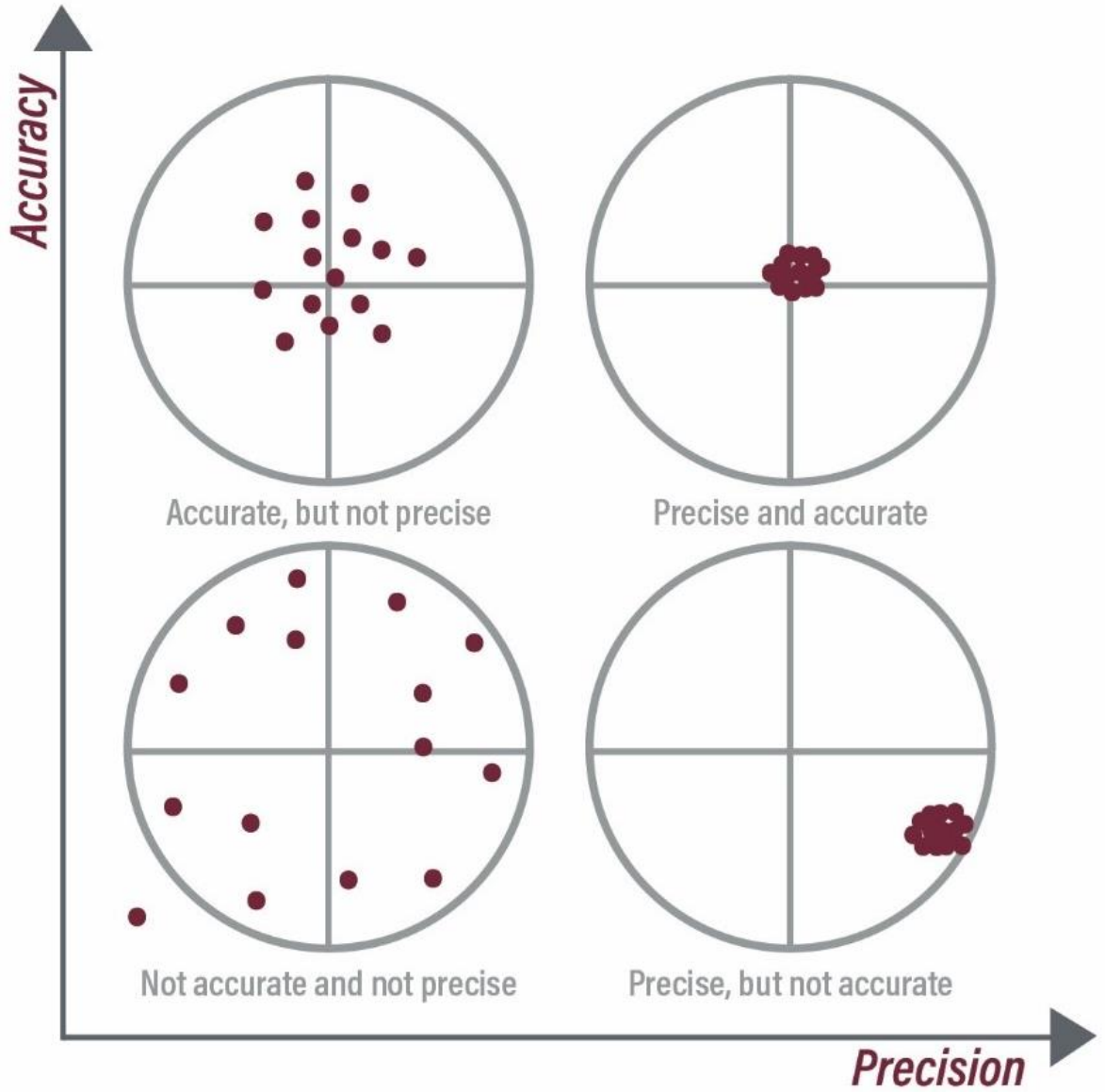


Objectives

- Compare probability and nonprobability sampling approaches
- Identify appropriate situations for each sampling method
- Explain how sampling choice affects generalizability and credibility

Describe strategies for reaching underserved populations and reducing representation bias.

Sample Bias _ Representation v. Sample Size (recap)



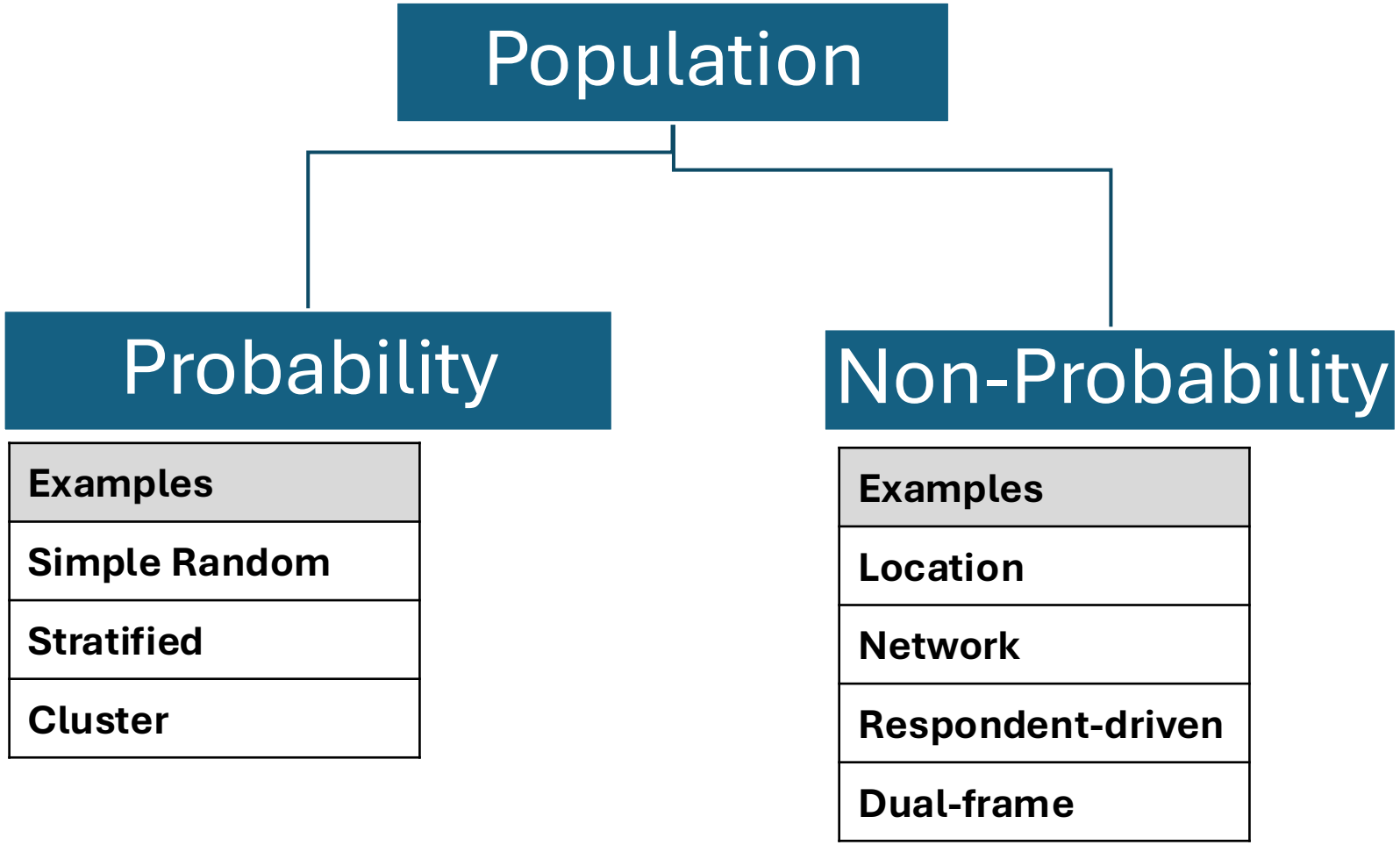
Sources of Error_ (recap)

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_ (recap)

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

Two Major Sampling Approaches



Probability Sampling

Simple Random



Every member of the population has an equal chance of selection.

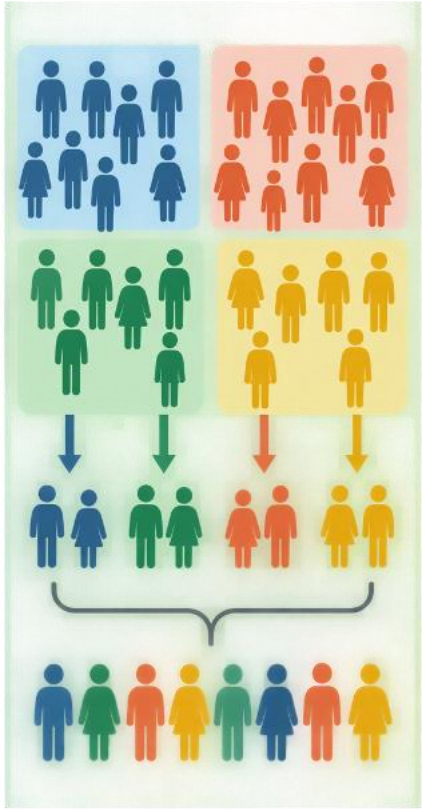
Examples

- Randomly selecting residents from an address list
- Randomly selecting patients from a clinic roster
- Drawing names from a registry

P (Selection) = n/N

➔500/5000 = 0.10 or 10%

Stratified Sampling



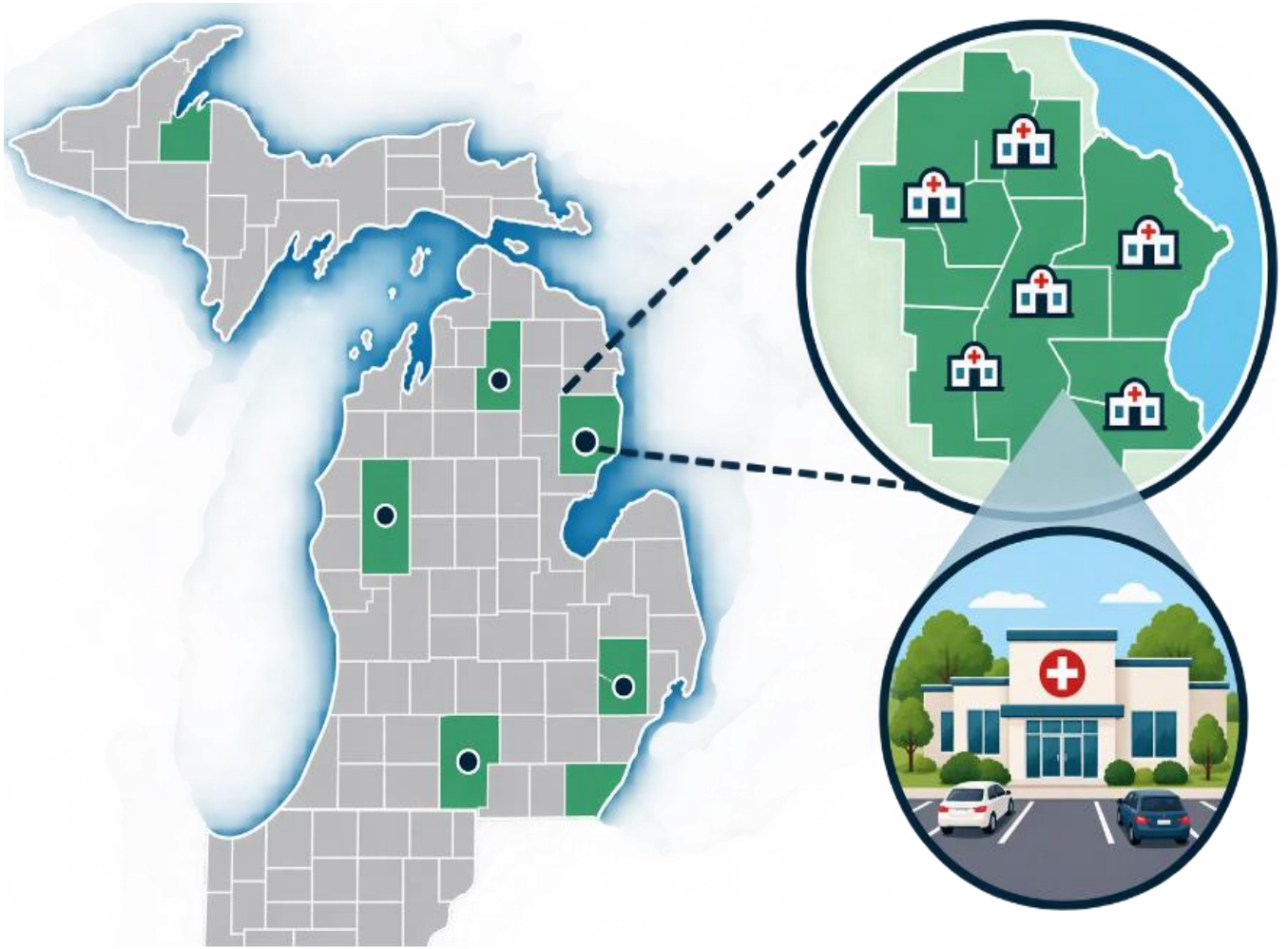
Population is divided into meaningful subgroups before sampling.

Examples

- Urban vs. rural residents
- Medicaid vs. non-Medicaid populations
- Age categories
- Race and ethnicity groups

Probability Sampling

Cluster Sampling

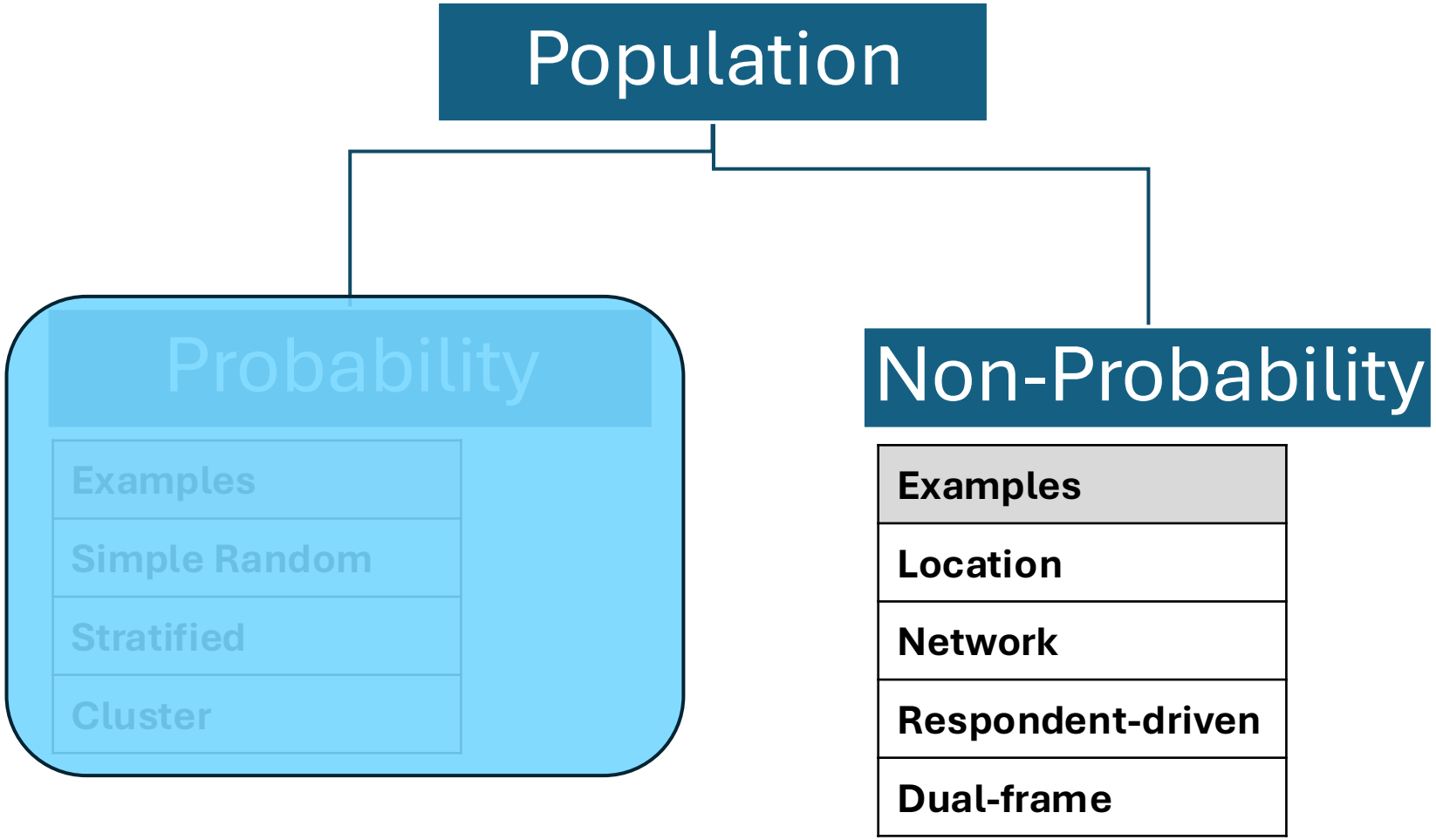


Clusters are selected first, then individuals are sampled within selected clusters.

Public Health Examples

- Selecting counties
- Selecting schools
- Selecting clinics
- Selecting neighborhoods

Two Major Sampling Approaches



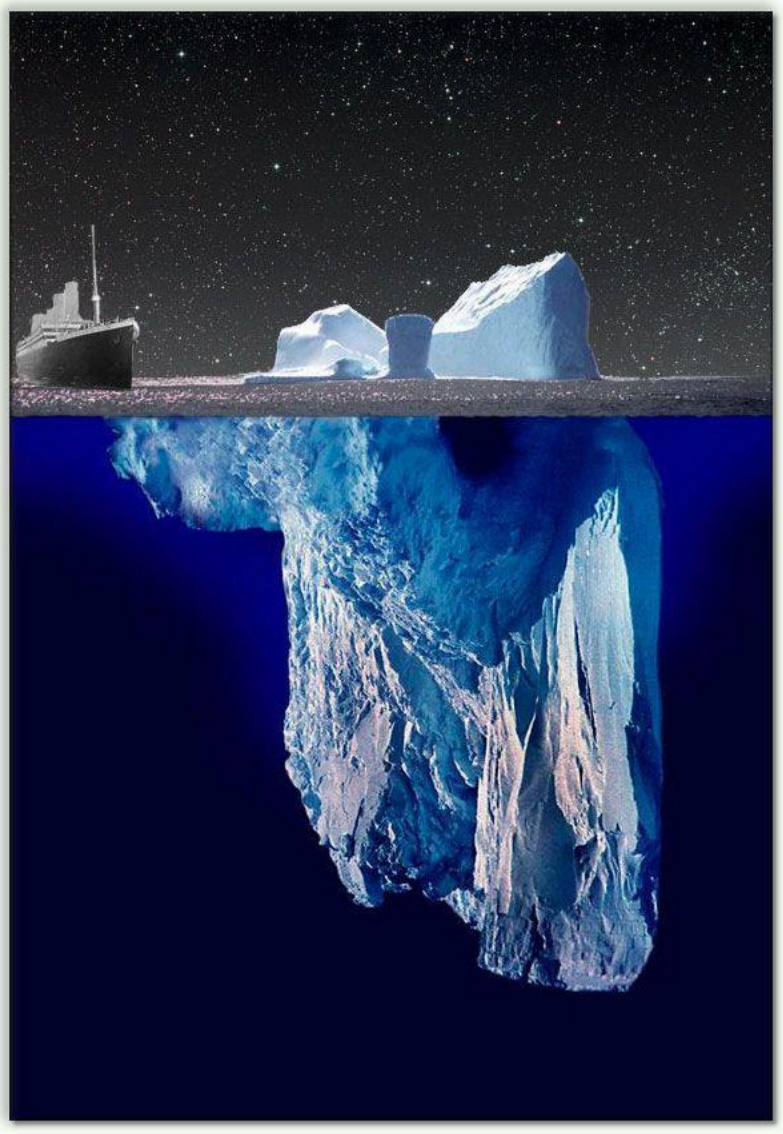
Comparing Probability and Nonprobability

Feature	Probability	Nonprobability
Random selection	✓	✗
Known selection probabilities	✓	✗
Confidence intervals possible	✓	✗
Reduced risk of selection bias	✓	✗
Faster and less expensive	✗	✓
Useful for hard-to-reach populations	Sometimes	Often

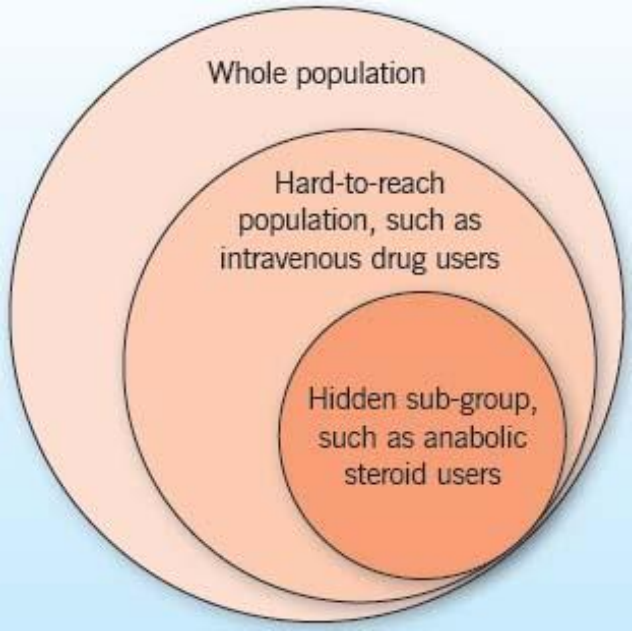
Examples Probability and Nonprobability

PROBABILITY	NONPROBABILITY
Community health assessments	Focus groups
Behavioral Risk Factor Surveillance System (BRFSS)-type surveys	Pilot studies
Immunization assessments	Housing and homelessness initiatives (when recruiting hard-to-reach populations through convenience, venue-based, or snowball sampling)
Maternal and child health projects (population-based surveys)	
Community health needs assessments	
Program evaluations when population estimates are needed	Outreach and community engagement efforts
Used when results need to be generalized to a broader population	Used when rapid feedback, exploration, or input from specific groups is the primary goal

Nonprobability Sampling — Reaching Hard-to-Reach Populations







Nonprobability Sampling — Reaching Hard-to-Reach Populations

Definition: group that comprises **10% or less of the general population**, and often much less.

Examples include:

- Men who have sex with men (MSM)
- Low-income households
- Users of specialized services
- Certain disease populations

Primary challenge: screening cost. If a target group represents only 2% of the population, approximately 50 screening interviews may be required to locate one eligible respondent.

Nonprobability Sampling — Reaching Hard-to-Reach Populations

First Determine Whether a List / Sampling Frame Exists

YES

- Registry available?
- Administrative data?
- Membership list?

Probability Sampling

NO

Alternative Approaches

- Location sampling
- Network sampling
- Respondent-driven sampling
- Dual-frame approaches

Reaching Hard-to-Reach Populations—Location Sampling

Sample People Where They Gather

- Community clinics
- Shelters
- Pride events
- Recovery meetings
- Harm reduction centers



Strengths	Limitations
Efficient	Misses non-attendees
Lower screening burden	Coverage bias
Builds trust	

Reaching Hard-to-Reach Populations _ Network Sampling

Leveraging Existing Social Connections

1. Members know each other well
2. Eligibility is known to network members
3. Contact information is available

Examples – Recruiting members of a migrant-worker community through referrals from existing participants.

or

Recruiting people with a rare disease through patient-support networks.

Major Challenge – Individuals with larger social networks have a greater chance of selection and require weighting adjustments.

Reaching Hard-to-Reach Populations—Respondent-Driven Sampling /Snowball

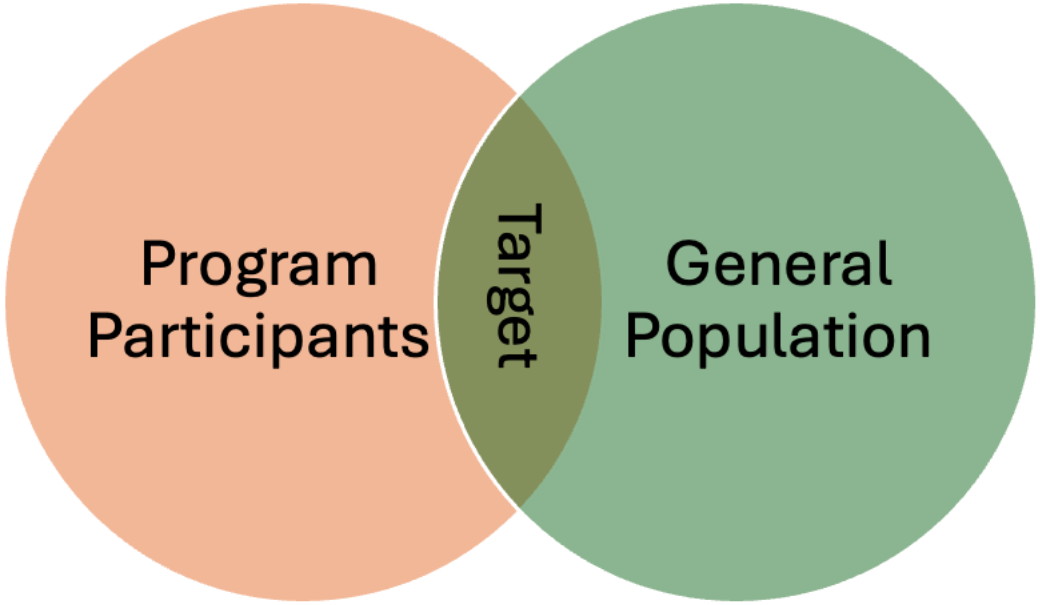
Recruitment Through Social Networks

1. Recruit "seed" participants
2. Participants recruit peers
3. Peer recruitment continues
4. Data are weighted statistically

Examples:

- HIV surveillance
- Substance use studies
- Hidden populations

Reaching Hard-to-Reach Populations—Dual-Frame Sampling



- Treatment Programs
- Clinic Registries
- Service Recipients

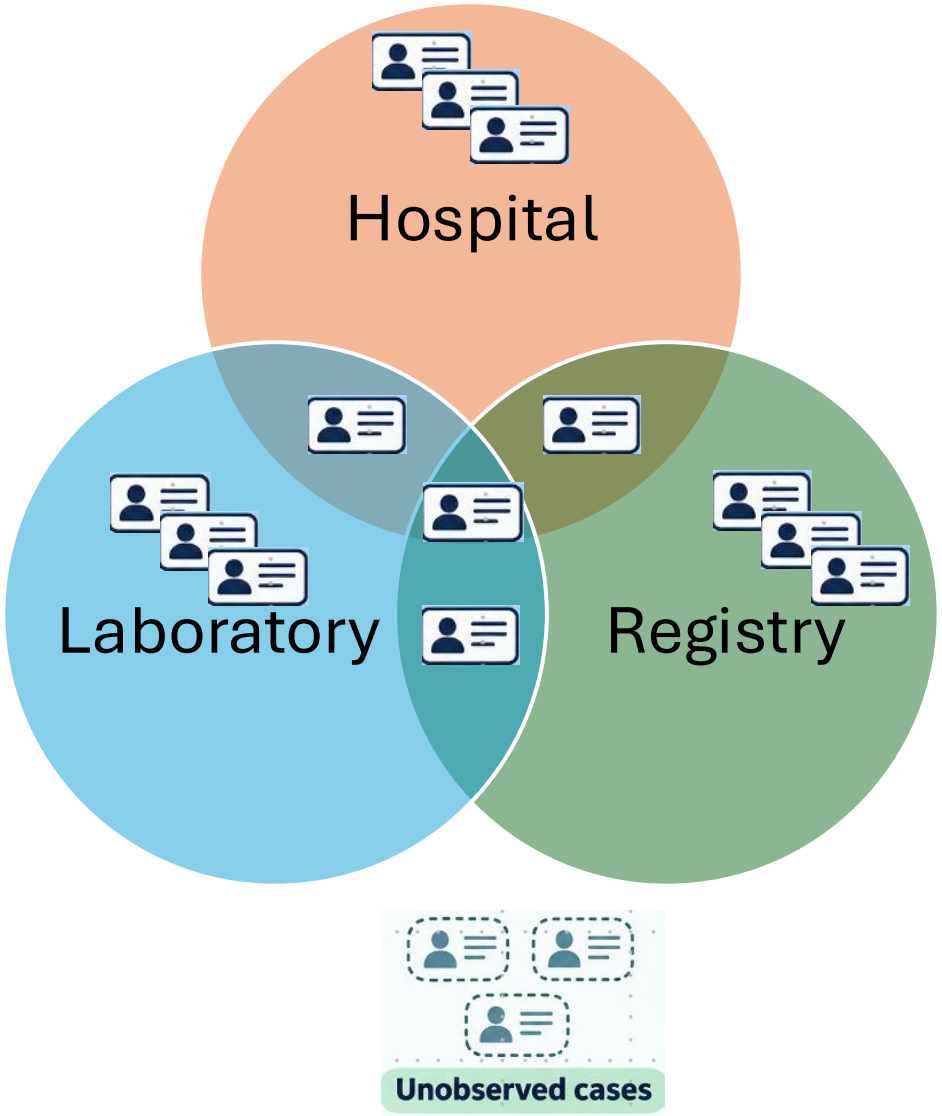
- Address-Based Sampling
- Household Surveys
- Telephone Surveys

Major advantage is that it reduces the risk of systematically excluding important segments of the population.

Useful for studying hard-to-reach populations where no single sampling frame provides complete coverage.

One of the most practical compromises between representativeness, feasibility, and cost

Reaching Hard-to-Reach Populations_Capture-Recapture



Disease registries rarely capture every case

Capture-recapture estimates cases missed by all sources

Helps estimate the true number of people affected by a disease

Challenge - not everyone has the same chance of being captured by these sources

Reaching Hard-to-Reach Populations—Capture-Recapture

Example

Suppose a state wants to estimate the number of people living with hepatitis C:

- Hospital records identify 3,000 people.
- Laboratory reports identify 2,500 people.
- 1,500 people appear in both systems.

Simple capture-recapture analysis would estimate how many cases were missed. However, if people in rural areas are less likely to appear in either source, simple analysis could underestimate the true burden.

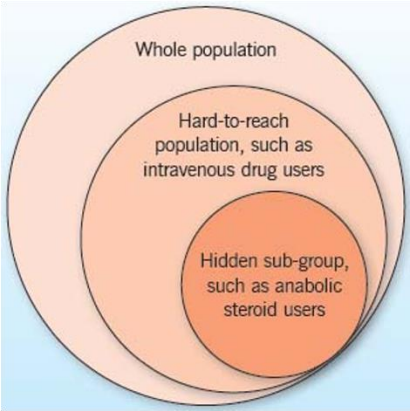
By incorporating factors such as age, geography, or healthcare utilization, researchers can produce a more realistic estimate of the total population affected

Reaching Hard-to-Reach Populations—Dual-Frame Sampling

Five Lessons

1. Hard-to-reach populations need specialized sampling approaches.
2. No single method eliminates bias.
3. Location-based and network-based methods improve efficiency.
4. Dual-frame designs often provide the best balance of coverage and cost.
5. Capture-recapture with adj for covariates, complicated and can greatly improve population estimates





Who is missing from the sample may be just as important as who is included.



Thank you

~ Probability vs. Nonprobability Sampling in Public Health Practice

"Building Reliable Public Health Evidence"~

Next →Lesson 4.

Core Probability Sampling Designs

Kevin Brooks, MSc, PhD