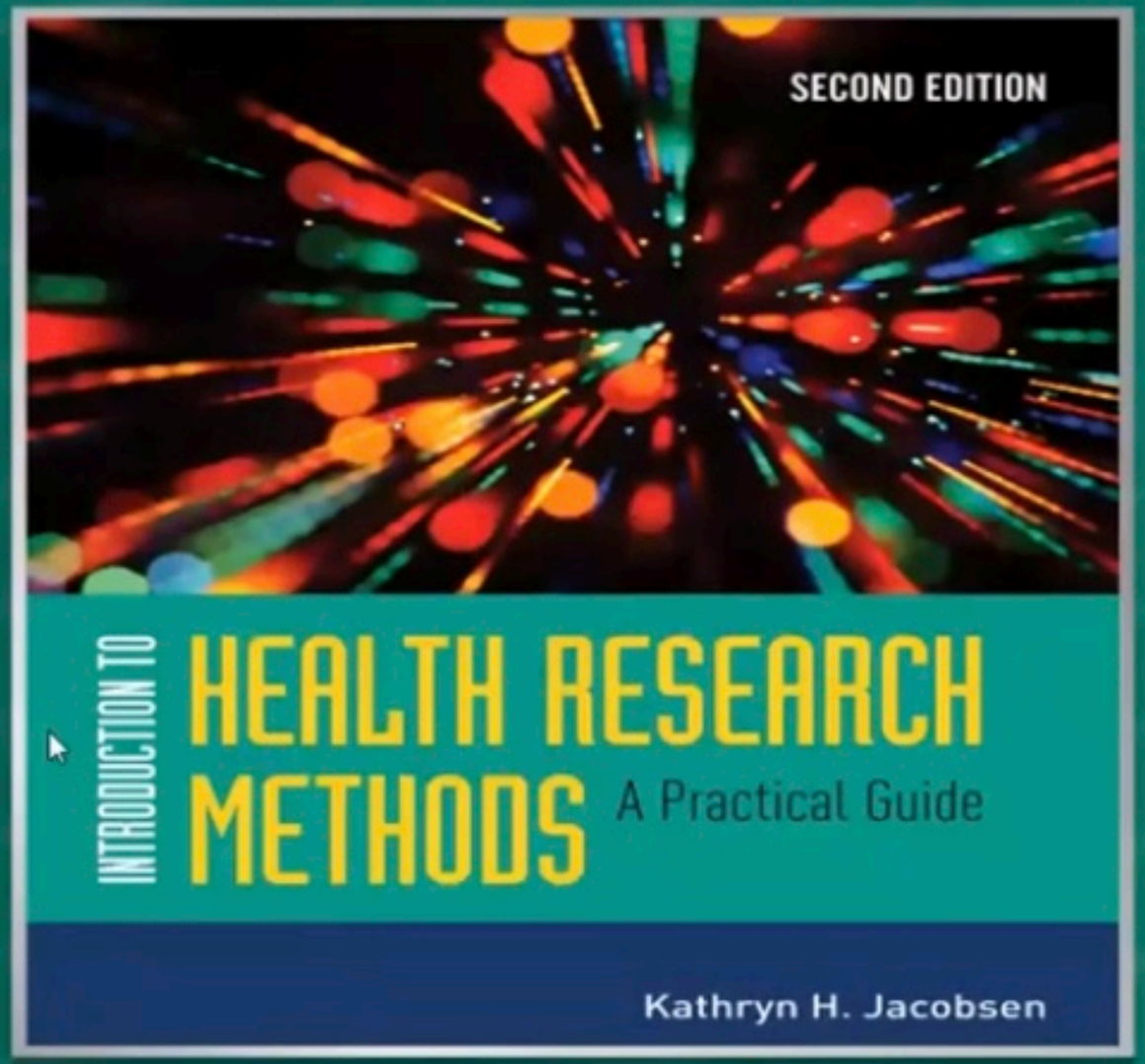
SCIENTIFIC MEDICAL RESEARCH

Week 6

Research Protocols

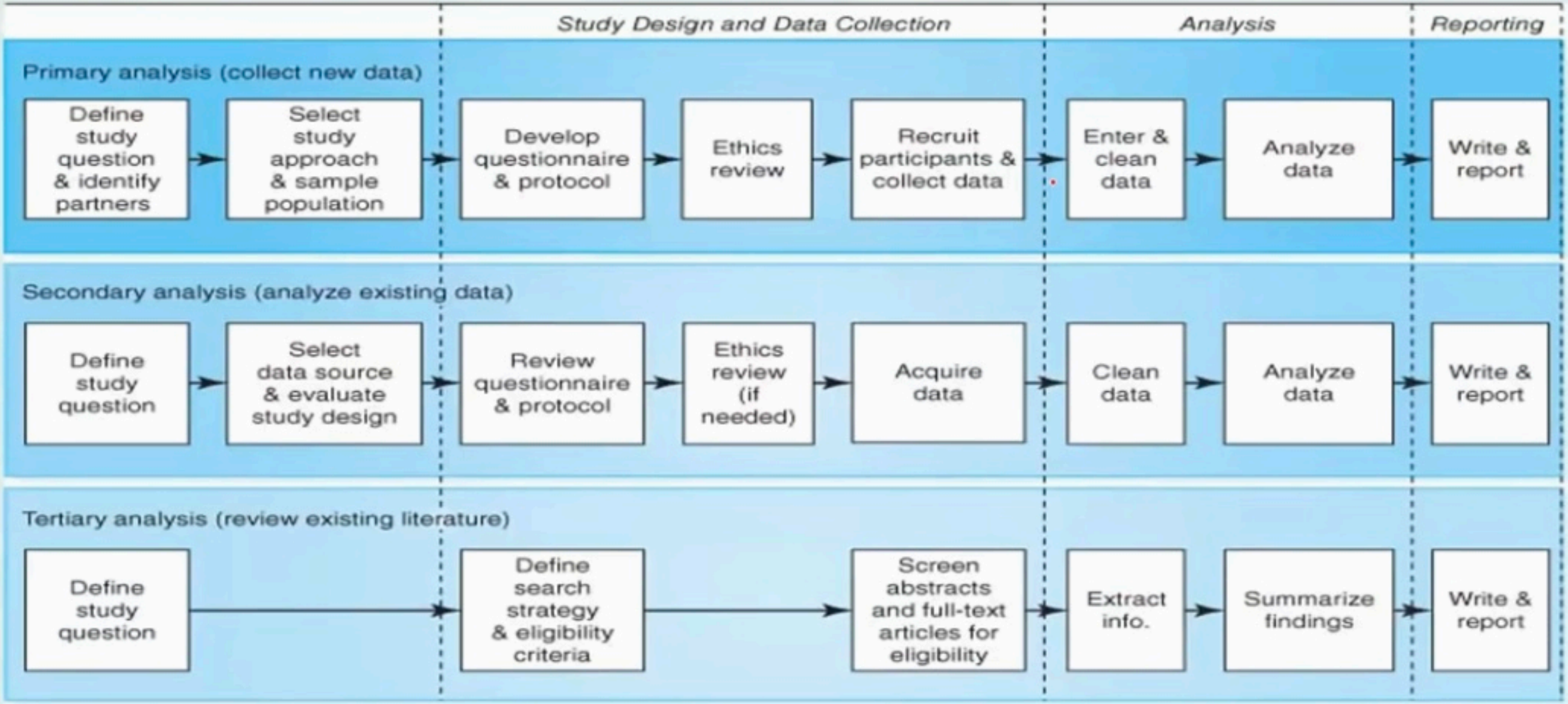
Chapter 15



15.1 Overview of Research Plans by Study Approach

- All types of studies—primary, secondary, and tertiary—require a research protocol to be created **prior** to the collection & analysis of data.

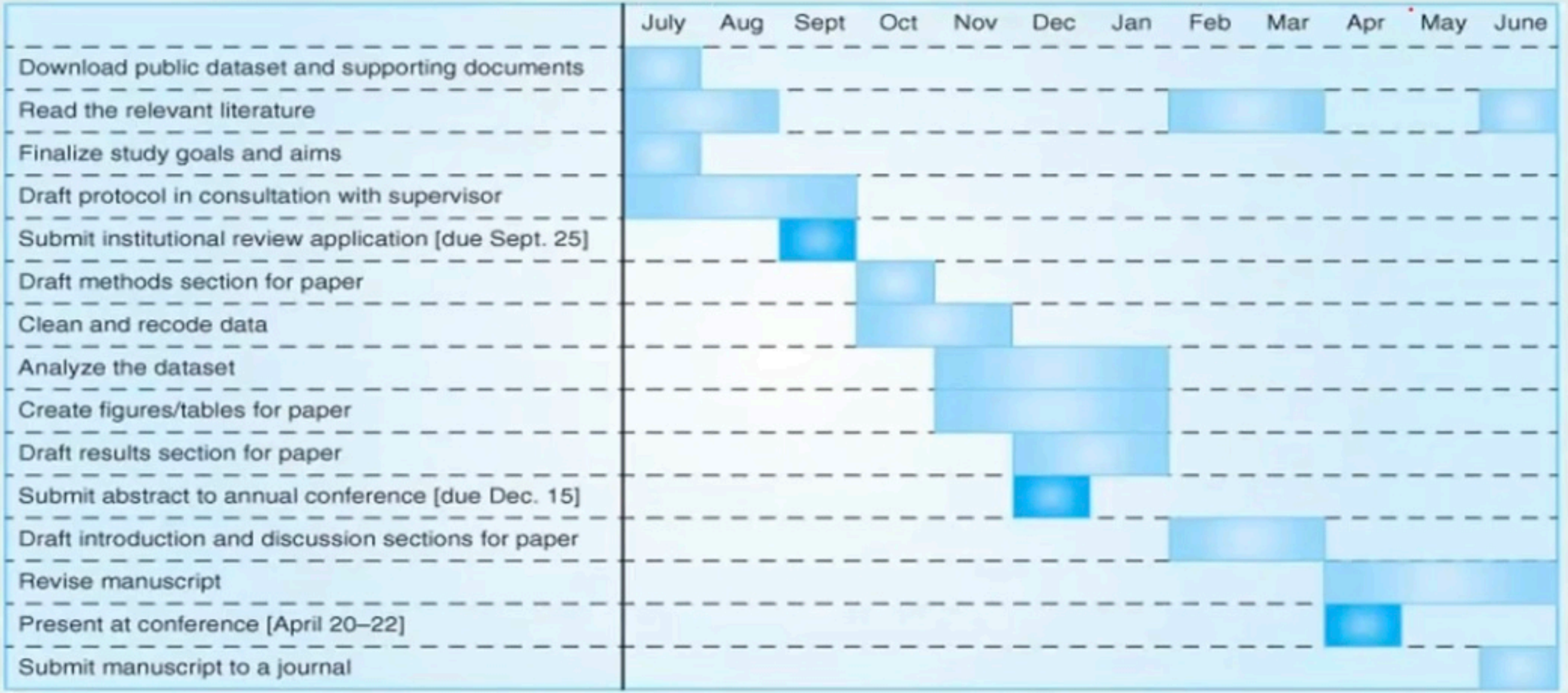
Figure 15-1: **Research Plans for Primary, Secondary, & Tertiary Data Collection**



15.2 Research Timelines

- It is helpful to create a project calendar (such as a ***Gantt chart***) that specifies critical deadlines & other steps toward successful and timely completion.
- Some flexibility is required.

Figure 15-2: **Sample (and Simplified) Gantt Chart for a Year-Long Secondary Analysis Project**



15.3 Researcher Responsibilities

- The roles & responsibilities accepted by each collaborator should be defined early in the project.
- Institutions usually require one person to be designated as the primary investigator (PI) with special responsibilities for ensuring that:
  a. the protocol is followed,
  b. the budget is maintained, and
  c. adverse outcomes are promptly reported.

15.4 Writing a Research Protocol

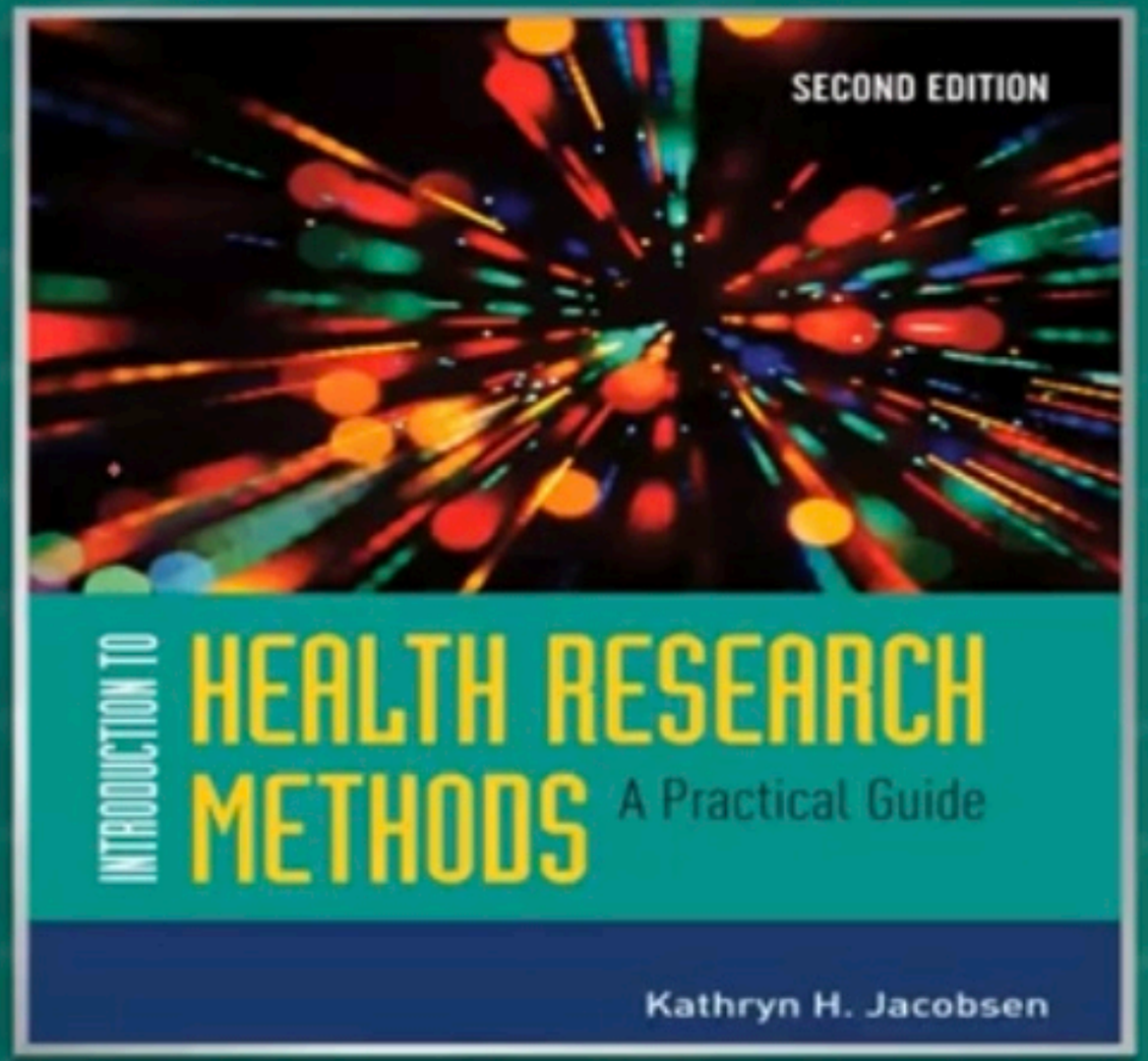
- A research protocol should describe the exact procedures that will be used for every step of the research process.
- A strong protocol provides enough detail that the study could be replicated by other research teams.

15.5 Preparing for Data Collection

- Data collection should only be initiated after a data management plan is in place.

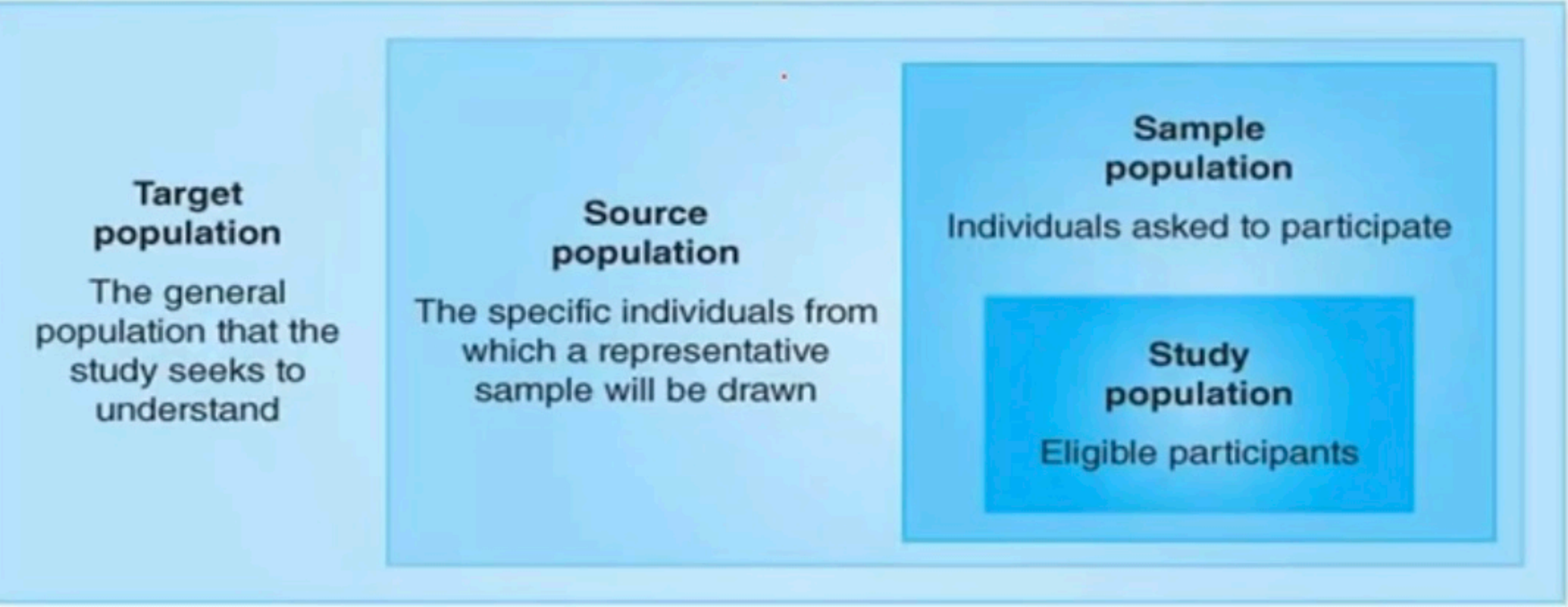
Population Sampling

Chapter 16



16.1 Types of Research Populations

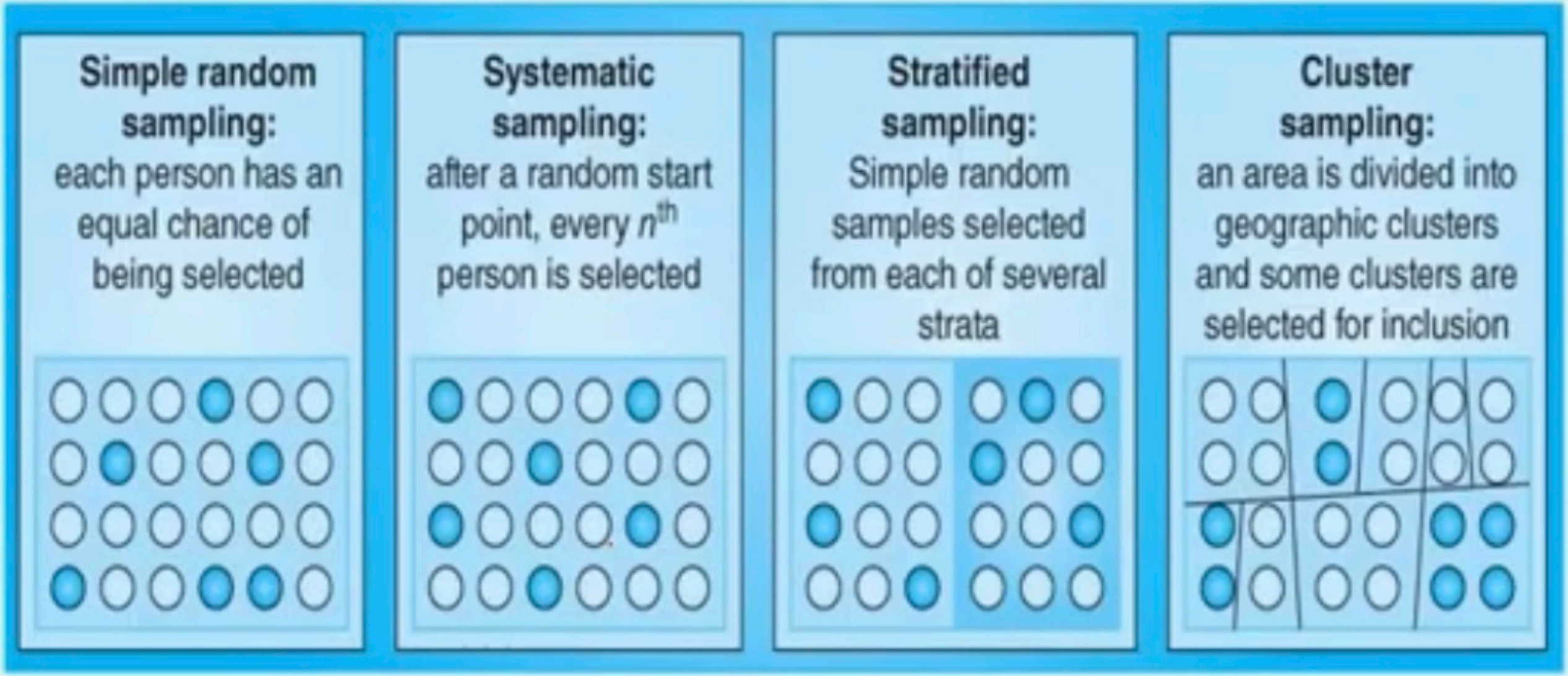
- Target population > source population > sample population > study populations



16.2 Target & Source Populations

- ***Target population***: the population to which the results of the study are intended to apply (for generalization)
- ***Source population (sampling frame)***: a list of particular people from whom a sample population can be drawn
- ***Sample population***: the members of the source population who are invited to participate in the study
- Ideally, probability-based sampling is used to ensure that the sample population is **representative** of the source population.

Figure 16-2


Simple random sampling:
each person has an equal chance of being selected
Systematic sampling:
after a random start point, every n^{th} person is selected
Stratified sampling:
Simple random samples selected from each of several strata
Cluster sampling:
an area is divided into geographic clusters and some clusters are selected for inclusion

Non-probability-based convenience Population

- May be selected based on ease of access.
- Should be used with caution (non-representative), which leads to **non-random sampling bias**.
- participants characteristics should be compared with the broader community intended to represent.
- Researchers must avoid **ascertainment bias** (the convenience sample is not representative of the source population as a whole)

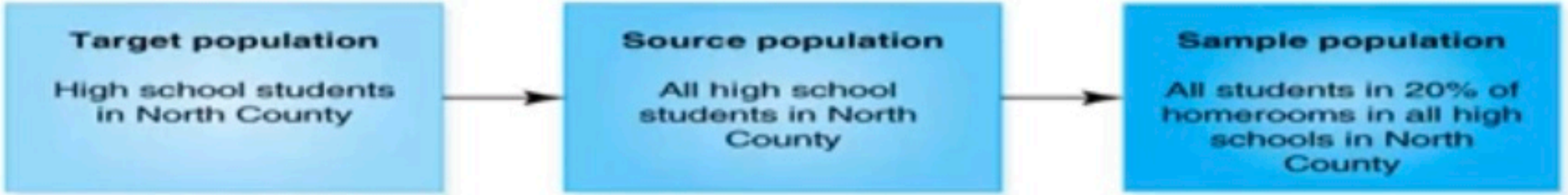
16.4 Study Populations

- ***Study population***: the people who actually participate in a study
- Few studies achieve a 100% participation rate (Response Rate) among those who are invited to participate, but researchers should aim for a high participation rate.
- Low Response Rate leads to **non-response bias**

16.5 Populations for Cross-Sectional Surveys

- Avoid ***convenience samples*** that are not representative of the target population.

FIGURE 16-3 Population Example for a Cross-Sectional Study

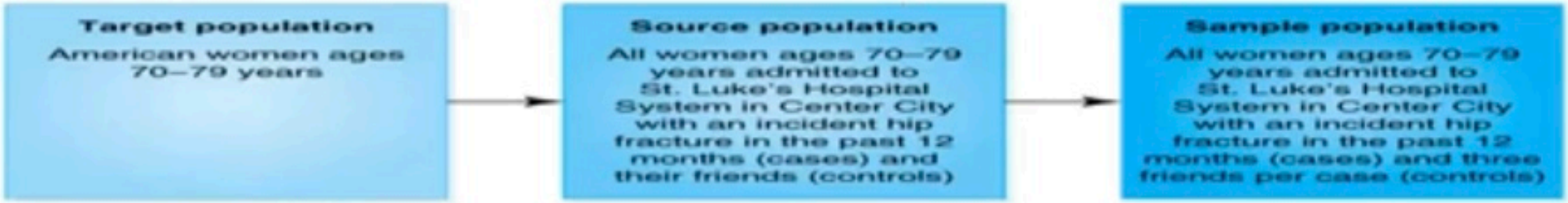


Study approach	Cross-sectional survey
Study question	What proportion of high school students in North County smoke cigarettes?
Data collection method	Participants will complete their own paper-based questionnaires.
Target population	Students in grades 9–12 in North County
Source population	All students enrolled in any of the 14 high schools in North County
Source population list	A list of the number of students in each homeroom provided by each high school
Sample population	Based on estimated sample size requirements, 20% of homerooms will be randomly sampled from the lists provided, and all students in these sampled homerooms will be asked to participate in the study.
Study population	Eligible individuals from the sample population who agree to participate
Confidentiality	No student names will ever be provided to researchers; surveys will be anonymous.

16.6 Populations for Case-Control Studies

- Find the cases first, and then identify an appropriate source of controls.

FIGURE 16-4 Population Example for a Case-Control Study

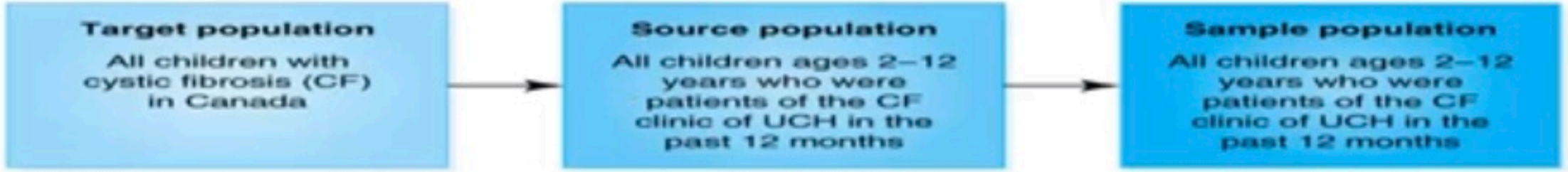


Study approach	Case-control study
Study question	What are the risk factors for hip fractures in adult women in the United States?
Data collection method	Participants will be interviewed in person or by telephone.
Target population	Women ages 70–79 living in the United States
Source population	All women ages 70–79 who were admitted to St. Luke's Hospital System in Center City with an incident hip fracture in the past 12 months
Source population list	A list of the hospital registration numbers for each inpatient female ages 70–79 at admission whose electronic medical record indicates a diagnosis of a hip fracture (ICD10 code S72) during the eligible 12-month period.
Sample population	All members of the source population will be asked to participate as cases, and each case will be asked to provide the names of three female friends in the same age range who might be able to serve as controls.
Study population	Eligible individuals from the sample population who agree to participate
Confidentiality	The hospital will provide the researcher with the names, addresses, and phone numbers of potential participants. Personally identifying information will not be included in the electronic file that contains questionnaire responses.

16.7 Populations for Cohort Studies

- A longitudinal cohort study needs a representative population (like a cross-sectional study).
- Prospective & retrospective cohort studies start by identifying an appropriate exposed population (in the same way that case-control studies start by identifying a source of cases).

FIGURE 16-5 Population Example for a Cohort Study

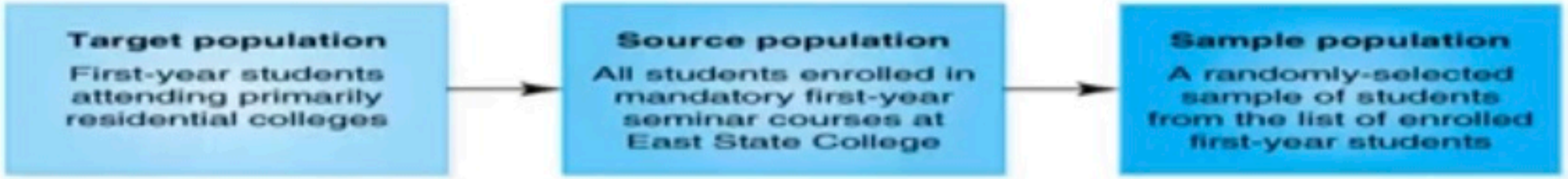


Study approach	Cohort study
Study question	What is the incidence rate for lung infections in children with cystic fibrosis?
Data collection method	Participants' parents will be asked to log all infections throughout the 2-year prospective study period, and these will be checked against the patients' medical records.
Target population	All children with cystic fibrosis in Canada
Source population	All children ages 2–12 years who were patients of the cystic fibrosis clinic of University Children's Hospital (UCH) in the past 12 months
Source population list	A list of all children ages 2–12 who were examined at the UCH cystic fibrosis clinic in the past 12 months
Sample population	The parents of all individuals in the source population will be asked if they will allow their children to participate in the study.
Study population	Eligible individuals from the sample population whose parents agree to let them participate
Confidentiality	All guidelines and regulations for the protection of patient information will be strictly adhered to, and only essential personnel will have access to patient records.

16.8 Populations for Experimental Studies

- Be aware of special ethical requirements associated with interventional studies.
- Safety must be the #1 priority.

FIGURE 16-6 Population Example for an Experimental Study



Study approach	Experimental study
Study question	Does nutritional counseling during the first semester of college prevent weight gain?
Data collection method	Half of the participants will be assigned to meet weekly with a nutritionist during their first semester, and half will have no intervention. All participants will complete nutritional assessments during the first and last weeks of the fall and spring semesters of their first year at college.
Target population	First-year students at primarily residential colleges
Source population	All first-year students at East State College
Source population list	A list of all students enrolled in the mandatory first-year seminar class at East State College
Sample population	A randomly selected sample of students from the source population
Study population	Eligible individuals from the sample population who agree to participate
Confidentiality	Nutritional counseling and assessment sessions will be conducted in a private setting, and only essential personnel will have access to participants' records. Participation in the study will be voluntary, and professors teaching first-year seminars will not known which students have enrolled in the study.

16.9 Vulnerable Populations

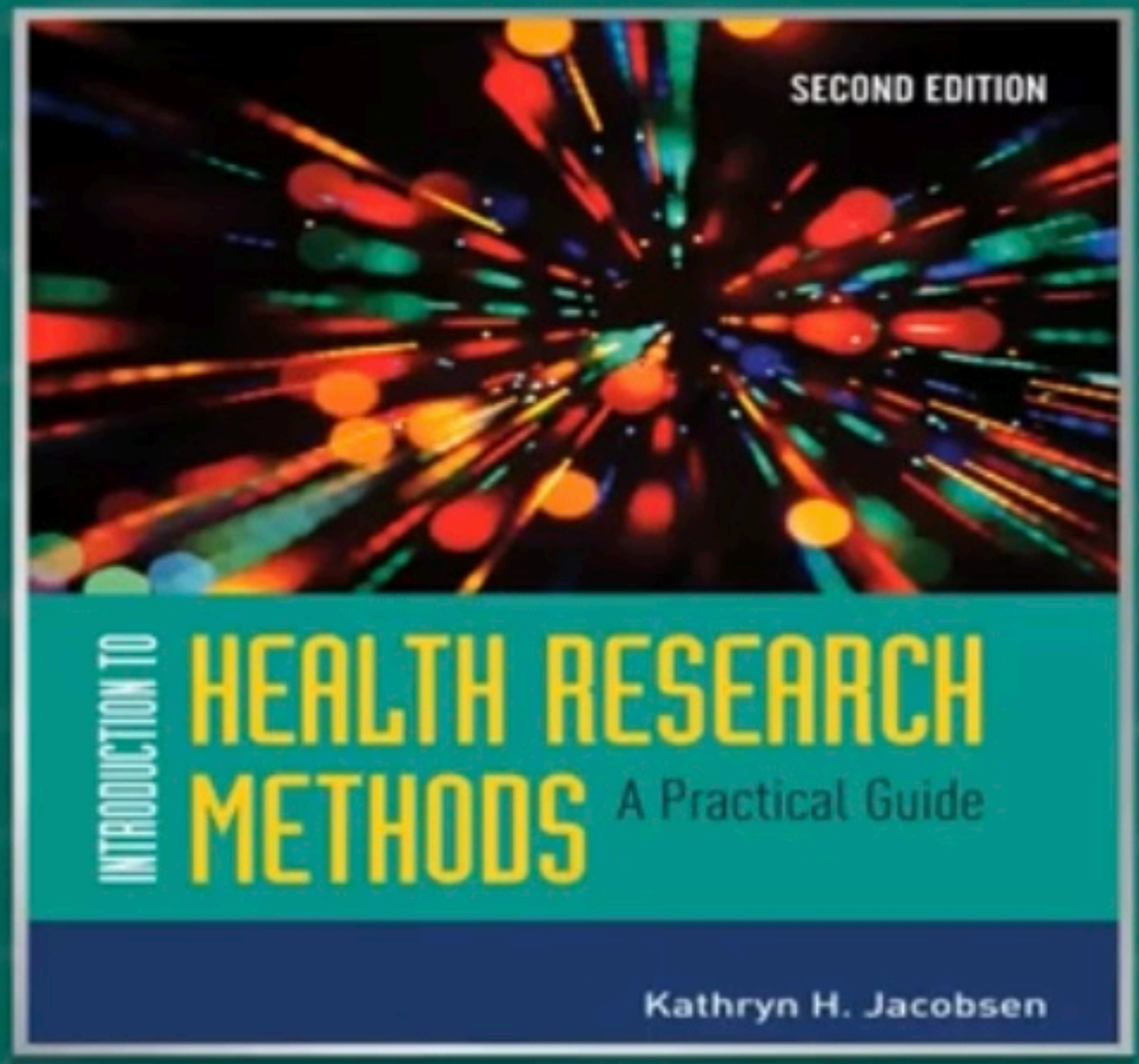
- Young children, people in prison, & some other groups of people may have limited ability to make an informed autonomous decision about participation.
- Should not be selected for studies that do not require their participation.
- Pay close attention to the ethics of research with vulnerable groups.

16.10 Community Involvement

- Community involvement improves response rate, improves cultural competence of the research team, and ensures that outcomes are valuable to the community
- ***Community-based participatory research (CBPR)*** is based on partnerships in which community members identify research priorities and are involved in every stage of the research process.

Sample Size Estimation

Chapter 17



17.1 Importance of Sample Size

- Should be based on estimation
- Recruiting too many participants wastes resources.
- Recruiting too few participants makes the study invalid.

17.2 Sample Size & Certainty Levels

- Larger sample sizes produce narrower confidence intervals (**an estimate of how close to the population value a sample of a particular size is expected to be**) for statistical measures.
- Larger sample sizes yield more statistically significant results.

Figure 17-1

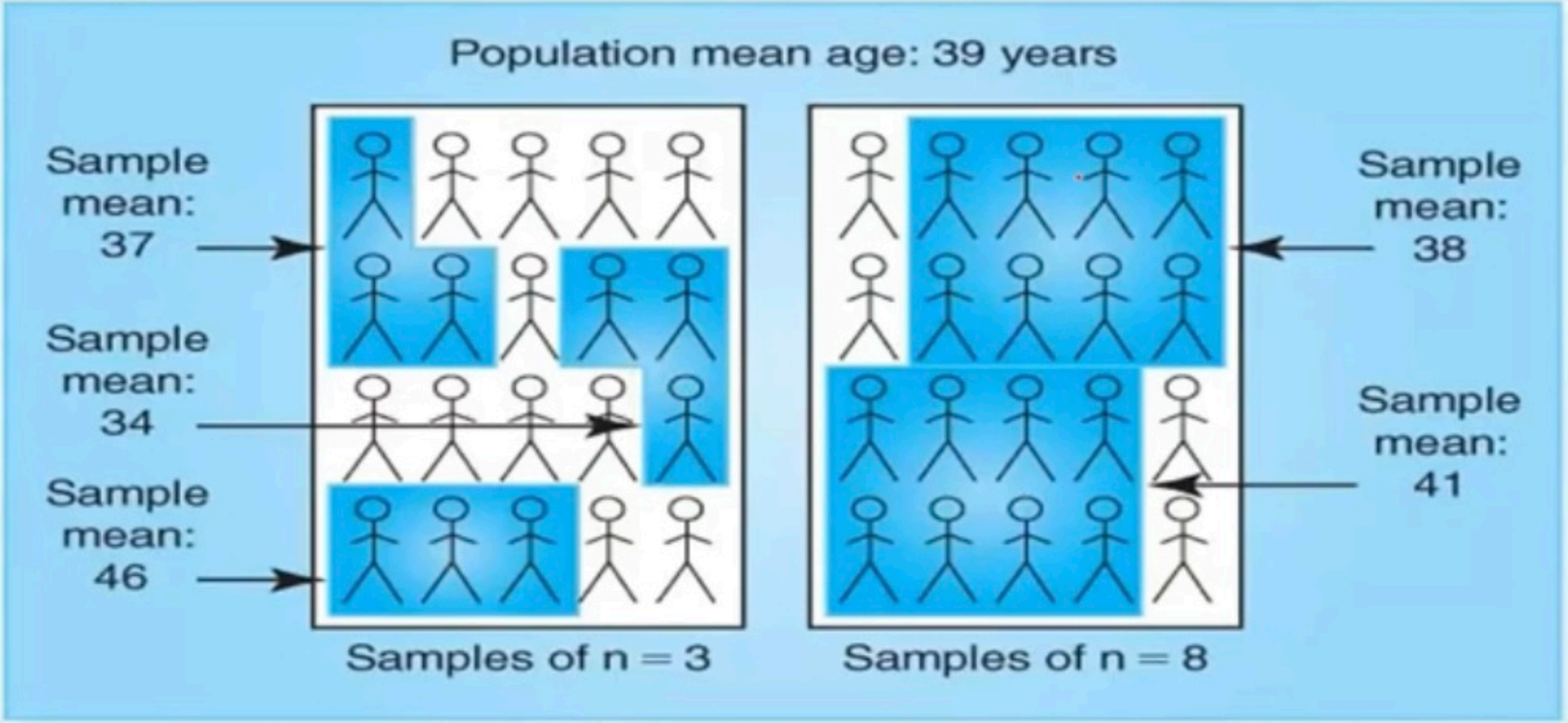
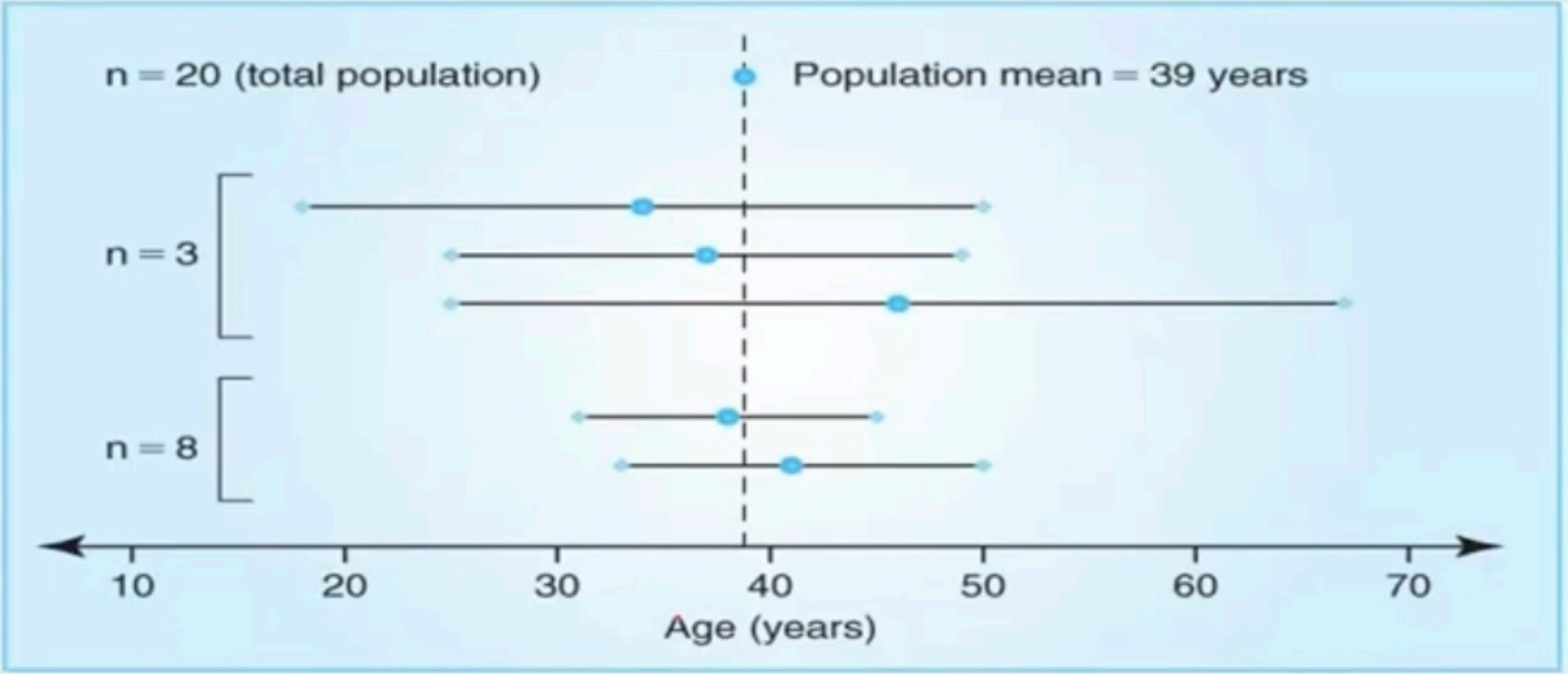

Population mean age: 39 years
Sample mean: 37
Sample mean: 34
Sample mean: 46
Sample mean: 38
Sample mean: 41
Samples of n = 3
Samples of n = 8

Figure 17-2: **Larger Samples from a Population Have a Narrower 95% CI Than Smaller Samples**



17.3 Sample Size Estimation

- Sample size calculators (estimators) estimate the number of participants necessary for a study based on *guesses* about the likely results of the study.
- When the level of certainty about inputs is low, it is better to err on the side of a larger sample size.

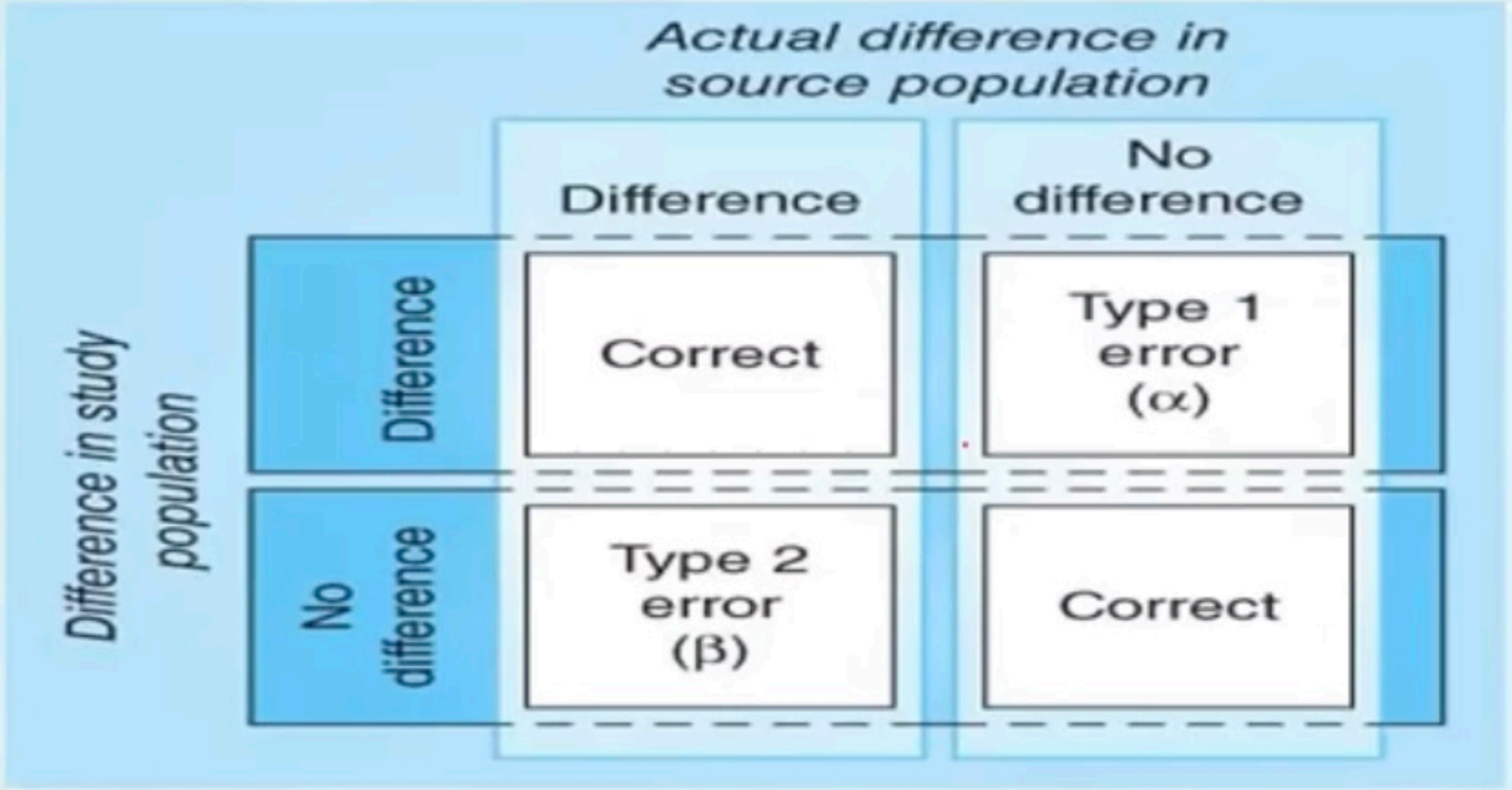
FIGURE 17-3 Examples of Sample Size Calculation

Characteristic	Cross-Sectional Survey	Case-Control Study	Cohort Study	Experimental Study
Study question	What proportion of the population has the exposure or disease?	Are cases more likely than controls to have the exposure?	Are exposed people more likely than unexposed people to develop the outcome?	Are exposed people more likely than unexposed people to have a favorable outcome?
Population size	5000	—	—	—
Anticipated percentage with exposure or disease	15%	—	—	—
Confidence for anticipated exposure percentage	±3%	—	—	—
Ratio of controls to cases	—	2	—	—
Ratio of unexposed to exposed	—	—	1	1
Anticipated percentage of controls exposed	—	25%	—	—
Anticipated percentage of unexposed with disease or outcome	—	—	10%	70%
OR worth detecting	—	1.5	—	—
RR worth detecting	—	—	1.3	1.25
Confidence level ($1 - \alpha$)	95%	95%	95%	95%
Power ($1 - \beta$)	—	80%	80%	80%
Estimated sample size	~500	~350 cases and 700 controls	~1850 exposed and 1850 unexposed	~90 exposed and 90 unexposed

FIGURE 17-4 Sample Size Estimates for a Case-Control Study

Situation	Ratio of Controls to Cases	Anticipated Percentage of Controls Exposed	Anticipated Percentage of Cases Exposed	Odds Ratio (OR) Worth Detecting	Estimated Sample Size Required	Estimated Number of Cases Required	Estimated Number of Controls Required
Base case (Figure 17-3)	2	25%	33%	1.5	1050	350	700
1:1 ratio	1	25%	33%	1.5	950	475	475
10:1 ratio	10	25%	33%	1.5	2750	250	2500
Lower % exposed	2	10%	14%	1.5	2025	675	1350
Higher % exposed	2	30%	39%	1.5	976	325	650
Lower OR	2	25%	29%	1.2	5400	1800	3600
Higher OR	2	25%	40%	2.0	345	115	230

Figure 17-5



17.4 Power Estimation

- ***Type I error (α):*** usually =5% and corresponds to 95 CI, and is written as p-value of p = 0.05
- **Type II error (β):** 20% is the acceptable level (Power= 1- β= 80%)

17.4 Power Estimation

- ***Power*** is the ability of a statistical test to detect significant differences between subgroups of a population when differences really do exist.
- Studies with more participants have more power.

FIGURE 17-6 Examples of Power Calculation

Characteristic	Cross-Sectional Survey	Case-Control Study	Cohort Study	Experimental Study
Number of exposed	100	—	2500	70
Number of unexposed	250	—	1500	70
Number of cases	—	250	—	—
Number of controls	—	490	—	—
Percentage of exposed with disease or outcome	40%	—	13%	85.7%
Percentage of unexposed with disease or outcome	26%	—	9%	64.3%
Percentage of cases with exposure	—	32%	—	—
Percentage of controls with exposure	—	25.5%	—	—
Confidence level ($1 - \alpha$)	95%	95%	95%	95%
Estimated power ($1 - \beta$)	~70%	~45%	~97%	~80%

17.5 Refining the Study Approach

- Recruit larger sample size than the estimated
- The study approach must be adjusted if the number of likely participants is much lower than the estimated number of people required for adequate power.

The End
Good Luck