Collecting data

MOS5e chapters 5 and 6

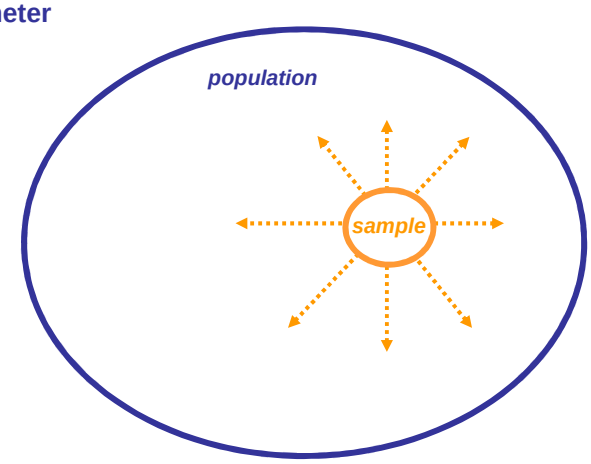
Two sources of information

Population
(data for all individuals)

→ parameter

Sample
(data for some individuals)

→ statistic



The role of randomness

How do you choose the individuals/units to observe for a study?

- **Anecdotal evidence:** haphazardly selected individual cases
- **Voluntary response sampling**: individuals choose to be involved
- **Convenience sampling**: ask whoever is around (mall, street) or take the next 10 units

- **Probability sampling**: individuals or units are randomly selected; the sampling process is **unbiased**

HBO's Last Week Tonight with John Oliver
Oct 2, 2016
Online Post-Debate Polls

www.npr.org/2016/09/28/495805190/no-donald-trump-didnt-win-post-debate-polls

Fox News polling director, Dana Blanton, warned that such polls "**do not meet our editorial standards**":
"As most of the publications themselves clearly state, the sample obviously can't be representative of the electorate because they only reflect the views of those Internet users who have chosen to participate. ... Another problem — we know some campaigns/groups of supporters encourage people to vote in online polls and flood the results."

Observational versus experimental studies

Observational study: record data on individuals without attempting to influence the responses.

Countless confounding variables limit scope of conclusions.

Concluding causation is very difficult.

Experimental study: Deliberately impose a treatment on individuals and record their responses.

Influential factors can be controlled.

Concluding causation is often possible.

Researchers surveyed a representative sample of 350 US executives to rate their diversity-valuing behaviors promoting demographic balance in the workplace. The researchers also obtained the performance evaluations for these 350 executives. They found that "white and male executives weren't rewarded, career-wise, for engaging in diversity-valuing behavior, and nonwhite and female executives actually got punished for it."

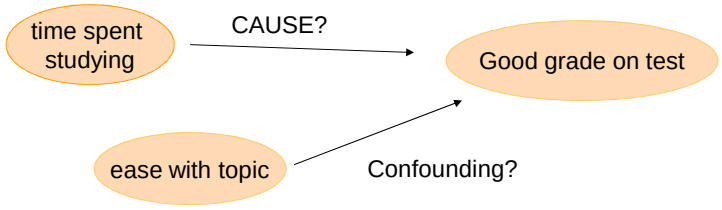
The researchers then asked 307 US working adults to review a hiring decision (between two equally qualified candidates, one being a white male the other not) made by a fictitious manager. Participants read a description of the hiring decision, saw a photo of the manager, and then rated the manager on competence and performance. The sex and race of the candidates and hiring managers were based on random assignment. They found that participants rated nonwhite managers and female managers as less effective when they hired a nonwhite or female job candidate instead of a white male candidate. There was no difference in how participants rated white male managers for different hiring choices.

Academy of Management Journal, March 2016, doi:10.5465/amj.2014.0538

Confounding

Two variables are **confounded** when their effects on a third variable cannot be distinguished from each other.

Observational studies often fail to yield clear causal conclusions because of confounding problems.



A 2013 Gallup study investigated how phrasing affects the opinions of Americans regarding physician-assisted suicide. Telephone interviews were conducted with a random sample of 1,535 national adults. Using random assignment, 719 heard the question in Form A and 816 the one in Form B.

> ***Form A:*** When a person has a disease that cannot be cured, do you think doctors should be allowed by law to **end the patient's life by some painless means** if the patient and his or her family request it?
>
> ***Form B:*** When a person has a disease that cannot be cured and is living in severe pain, do you think doctors should or should not be allowed by law to **assist the patient to commit suicide** if the patient requests it?

70% of those given Form A answered "should be allowed", compared with only 51% of those given Form B. What type of study is this?

A. Observational study.
B. Randomized experiment.
C. Neither: This is just anecdotal evidence.

Observational studies: sample surveys

A **sample survey** is an observational study that relies on a random sample drawn *from the entire population* ("**cross-sectional**").

- Opinion polls are sample surveys that typically use voter registries or telephone numbers to select their samples.
- Economic indicators (unemployment, consumer price index, …) are monitored through extensive and closely guarded monthly surveys.
- In epidemiology, sample surveys are used to establish the rate of various medical conditions, diseases, and lifestyles.
- Surveys are used extensively in social sciences since experimentation options are limited.

Don't confuse scientific surveys/polls and quick polls using voluntary samples.

Some survey challenges

- **Selection bias** and **undercoverage:** Parts of the population are systematically left out or underrepresented.

  www.pewresearch.org/2015/11/18/advances-in-telephone-survey-sampling
  www.pewresearch.org/2015/09/22/coverage-error-in-internet-surveys

- **Nonresponse / nonparticipation**: Some people choose not to answer/participate.

- **Wording effects**: Biased or leading questions, complicated/ confusing statements can influence survey results.

- **Response bias**: Fancy term for lying or forgetting (especially on sensitive/personal issues). Can be exacerbated by survey method (in person vs. by phone or online).

	1997	2000	2003	2006	2009	2012
	%	%	%	%	%	%
Contact rate (percent of households in which an adult was reached)	90	77	79	73	72	62
Cooperation rate (percent of households contacted that yielded an interview)	43	40	34	31	21	14
Response rate (percent of households sampled that yielded an interview)	36	28	25	21	15	9

PEW RESEARCH CENTER 2012 Methodology Study. Rates computed according to American Association for Public Opinion Research (AAPOR) standard definitions for CON2, COOP3 and RR3. Rates are typical for surveys conducted in each year.

Figure. Percentages of adults and children living in households with only wireless telephone service: United States, 2003–2015

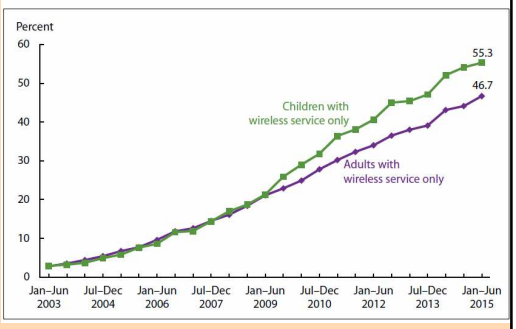


Table 1. Mean Difference Between Self-Reported and Measured Height, Weight, and Body Mass Index, by Sex, National Health and Nutrition Examination Survey, 2001-2006[a]

Characteristic	Self-Reported	Measured	Mean Difference (95% CI)
Men			
Standing height, cm	177.67	176.45	1.22 (1.15, 1.28)
Weight, kg	87.55	87.25	0.30 (0.20, 0.40)
Body mass index (kg/m^2)	27.62	27.96	−0.34 (−0.38, −0.30)
Women			
Standing height, cm	162.99	162.31	0.68 (0.62, 0.74)
Weight, kg	72.46	73.85	−1.39 (−1.48, −1.30)
Body mass index (kg/m^2)	27.20	28.02	−0.82 (−0.85, −0.77)

Some other types of observational studies

Case-control studies start with 2 random samples of individuals with different outcomes. Individuals with the condition are cases, and those without are controls.

Cohort studies enlist individuals sharing a common demographic and study them over a long period of time to examine the compounded effect of various factors over time ("**longitudinal**").

-**Retrospective** studies look for exposure factors in the subjects' past.
-**Prospective** studies collect data over time waiting until some participants eventually develop a condition.

Some types of probability samples

- A **simple random sample** is made of randomly selected individuals. All possible samples of size *n* have the same chance of being drawn.
  - Draw from a hat (lottery style)
  - Use a table of published random numbers
  - **Use software** that generates random numbers

In a class of 220 students, the instructor uses the roster to randomly pick 5 students' midterms to check that they were graded properly.

- **Systematic:** create your sample by taking every other n^{th} individual on the population list (beware of potential patterns/cycles in population)

- **stratified:** make sure your sample has *x,y,z*% of individuals of certain types, typically to fit the population makeup

The Field Poll is a non-partisan survey of California public opinion that uses a stratified random sample of the state's registered voter population. Sample quotas can be chosen within age groups, or by gender, or across regions so that the composition of the sample matches the characteristics of the overall voter population.

- **cluster and multistage:** select your final sample from a sample or subset of places – *Analysis for multistage samples is more complex than for simple random samples.*

The National Youth Tobacco Survey is administered in schools to students in grades 6–12. Sampling consists of randomly selecting:

1) Counties as Primary Sampling Units (PSU).
2) Schools **within** each selected county.
3) Classes **within** each selected school.

- …

Comparative, randomized experiments

The **individuals** in an experiment are the **experimental units.** If they are human, we call them **subjects.**

Experiments **compare** the response to a treatment versus to:

- other treatments
- the absence of treatment
- a placebo

Experiments **randomize** the assignment of individuals to treatments.

Experiments use **replication**: several individuals are studied.

About the "placebo effect"

Improvement in health or perceived condition due, not to any active treatment, but only to the patient's belief that he or she is being cared for or helped.

- therapeutic results on up to 35% of patients
- neural response to the placebo effect recorded even in spinal chord

Direct Evidence for Spinal Cord Involvement in Placebo Analgesia

Falk Eippert,[1*] Jürgen Finsterbusch,[1] Ulrike Bingel,[2] Christian Büchel[1]

Science (2009), DOI:10.1126/science.1180142

- "negative placebo effect" also observed
- famous placebo: kiss/blow/hug/bandaid to help kids cope with minor injuries

Bias and blinding

Bias when dealing with human subjects: (1) placebo effect and Hawthorn's effect (impact of being studied), (2) experimenter bias, conscious or unconscious. "**Blinding**" can help against bias.

A **double-blind** experiment is one in which neither the subjects nor the experimenter(s) know which individuals got which treatment until the experiment is completed.
Placebos may need to be creative: **sham** procedure, **double dummy**

However, subjects must be informed that they will get one of a number of treatments, possible pros and cons of each, and must give consent.

Efficacy of acupuncture for migraine prophylaxis: A single-blinded, double-dummy, randomized controlled trial

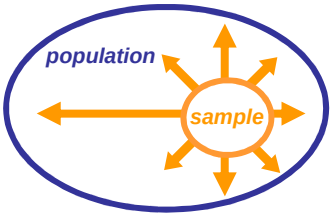
doi: 10.1016/j.pain.2011.04.006

Blinding was implemented by means of double-dummy with verum acupuncture plus placebo medication in the acupuncture group and flunarizine plus sham acupuncture in the control group. The follow-up assessors and statisticians, who were uninvolved in clinical management, were blinded throughout the study. Due to the procedure of the acupuncture technique, acupuncture practitioners in this trial were unable to be blinded.

The sham points were chosen to be unrelated to headache treatment but to have otherwise the same number and same needle insertion method as the acupoints intended for the treatment of headache.
The appearance of the placebo medication was exactly the same as that of flunarizine.

More issues: Lack of realism, intended population

Random sampling is meant to represent the larger population from which we sample. Conclusions apply only to individuals like the ones sampled.



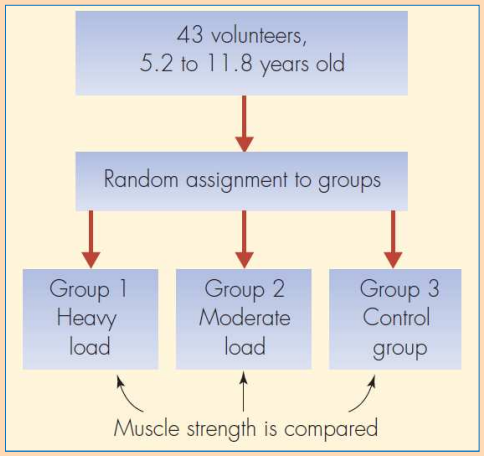
U.S. Food and Drug Administration
Protecting and Promoting *Your* Health

- *FDA requiring lower recommended dose for certain sleep drugs containing zolpidem* (Jan. 10, 2013): "Since women eliminate zolpidem from their bodies more slowly than men, the FDA has notified the manufacturers that the recommended dose should be lowered for women"

- Carcinogenicity studies administer high doses of a potential carcinogen to lab rats. Results don't always apply to humans (e.g. saccharin delisted in 2000).

Some common experimental designs

In a **completely randomized experimental design** individuals are randomly assigned to groups, then the groups are randomly assigned to treatments.



In **matched pairs / repeated measures** designs, either

- we choose **pairs of subjects** that are closely matched (like twins). Within each pair, *randomly* assign who will receive which treatment.

- **OR** we give the two (or more) treatments to **each subject over time**, in *random* order, so we have repeated measures for each subject.

Each subject is given two chili bowls: version A and version B. Subjects eat and rate both versions.

Half of all subjects are given version A first then version B.

The other half are given version B first then version A.

Data collection for an experiment with two conditions (A, B) under a completely randomized design versus a matched pairs design

Individual ID	Condition tested	Value recorded
01	A	
02	A	
03	B	
04	A	
05	A	
06	B	
07	B	
08	A	
09	B	
10	B	
...		

Individual or pair ID	Value under condition A	Value under condition B
01		
02		
03		
04		
05		
06		
07		
08		
09		
10		
...		

The Clinical Efficacy of Kinesio Tape for Shoulder Pain: A Randomized, Double-Blinded, Clinical Trial

doi:10.2519/jospt.2008.2791

A: cross-sectional survey
B. longitudinal cohort
C. case-control study
D. completely randomized experiment
E: matched pairs experiment

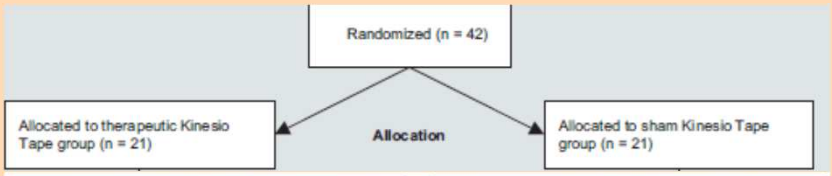


FIGURE 2. Therapeutic Kinesio Tape application.

FIGURE 3. Sham Kinesio Tape application.

Researchers studied the 2010 epidemic of pertussis in California.

- They selected a random sample of 682 medical records of California children ages 4 to 10 who had been diagnosed with pertussis.
- They also selected a random sample of 2016 medical records of California children in the same age group who had received care from the same clinicians on the same day but were not diagnosed with pertussis.

They found that children diagnosed with pertussis were much more likely to have not received any pertussis vaccine, to have not received all recommended doses of the vaccine, or to have had a longer interval of time since their last vaccination.

A: cross-sectional survey
B. longitudinal cohort
C. case-control study
D. completely randomized experiment
E: matched pairs experiment

doi:10.1001/jama.2012.14939

Does cell phone use while driving impair reaction time?

32 subjects took a familiar driving skill test twice, once while operating a cell phone and once while listening to the radio. The order of the conditions was randomized and reaction time (in ms) was assessed for each subject in both tasks:

A: cross-sectional survey
B. longitudinal cohort
C. case-control study
D. completely randomized experiment
E: matched pairs experiment

Cell	Radio
636	604
623	556
615	540
672	522
601	459
600	544
542	513
554	470
543	556
520	531
609	599
559	537
595	619
565	536
573	554
554	467
626	525
501	508
574	529
468	470
578	512
560	487
525	515
647	499
456	448
688	558
679	589
960	814
558	519
482	462
527	521
536	543