

An Overview of Data Collection & Survey Design



Lisa V. Smith, MS, DrPH
Chief, Surveillance, Epidemiology and Evaluation
Division of Chronic Disease and Injury Prevention
L.A. County Department of Public Health



If we knew what we were doing, it wouldn't be called research, would it?

Albert Einstein (1879 - 1955)

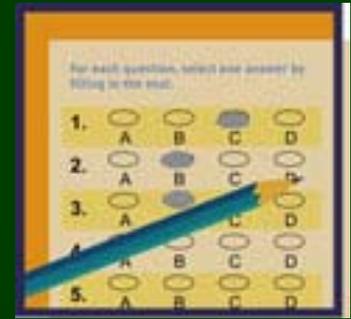
Research is defined as “an original investigation undertaken in order to gain knowledge and understanding .” Source: <http://www.rgu.ac.uk/credo/staff/page.cfm?pge=9471>

Overview of Presentation



- Developing surveys and other data collection tools
 - What is a survey? Why do we survey?
 - What makes a good survey?
 - What are the different types of surveys?
 - What are the steps in developing a questionnaire?
- Role of Epidemiologists & Research Analysts in questionnaire development
 - What is the difference between an Epidemiologist and Research Analyst?
 - Who should do what?

Overview of Presentation



- Available data sources
 - Department of Public Health
 - Academic
 - State
 - Federal
- How to access epidemiologic support
 - How do you find help to collect primary and secondary data?
- IRB considerations
- Topics for future surveys (Special Exercise)

What is a survey?



○ What is a survey?

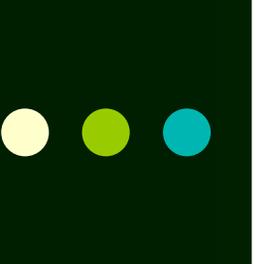
- A system for collecting information to describe, compare, or explain knowledge, attitudes, and behavior
- Survey instruments (questionnaires)
 - Mailed, taped, or self-administered questionnaires
 - In-person (face-to-face) or telephone interviews

Developing surveys and other data collection tools

Why do we survey?



- ☑ To discover what's going on
 - In a non-threatening survey environment, you can learn about what motivates your clients and what's important to them
- ☑ To provide an opportunity to discuss key topics with your target population
 - Communicating with your clients allows for deeper insight into your problems, and can shed light on your research problem
- ☑ To prioritize your actions based on objective data
 - Allows you to make evidence-based decisions rather than those based on “gut” reactions
- ☑ To provide a benchmark
 - Surveying provides a “snapshot” of your target population and helps establish a baseline



Primary vs. secondary data sources

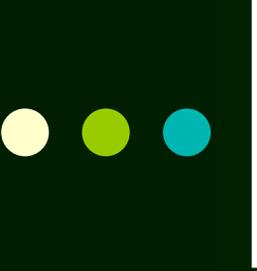
- Primary: Original data collected for a specific research goal
 - Examples: diaries, interviews
 - More costly, but researcher controls collected data and information is timely
- Secondary: Data originally collected for a different study, but used again for a new research question
 - Examples: scholarly books, journal articles
 - Less costly, but researcher has little control over how data were collected and may be obsolete

What are good surveys?

Good surveys have the following characteristics:

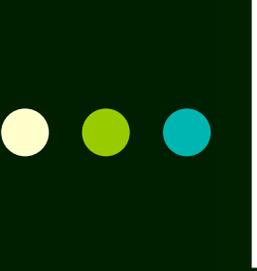
- ☑ Specific, measurable objectives
- ☑ Straightforward questions
- ☑ Sound research design
- ☑ Sound choice of population or sample
- ☑ Reliable and valid survey instruments
- ☑ Appropriate reporting of survey results





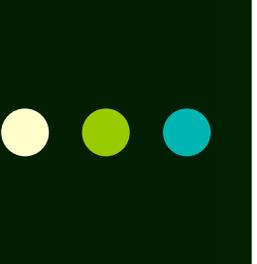
Specific objectives

- Where do survey objectives originate?
 - Need: Someone asks you to do it
 - Literature review: PubMed, Cochrane Library, etc.
 - Focus groups: 10 or less individuals
 - Consensus panel of experts



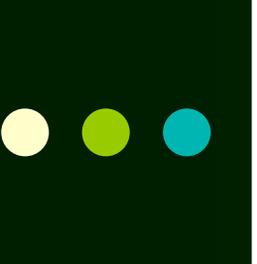
Straightforward questions

- ☑ What are their characteristics?
 - *Purposeful*: Respondent can identify the relationship between the intention of the question and the objectives of the survey
 - *Complete sentences*: One complete thought
 - *Mutually exclusive*: Open or closed



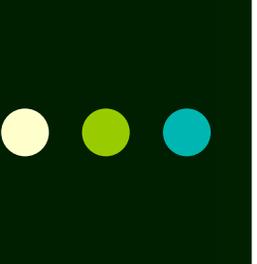
Example of a measurable objective

- Background: As part of its annual student profile, UCLA collects data on students who enter with advanced placement (AP) units
- Objective: To assess the extent of change in the proportion of students entering UCLA with AP units
- Instrument: data abstraction form
- Design: Survey



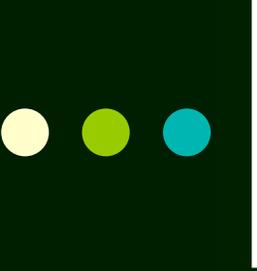
What is wrong with these questions?

- This is a survey about cats. Do you go to amusement parks? What???
- How would you describe your mood?
 - Since when? Birth? Last week?
- Place of birth?
 - City? State? East side? West side?
- How often during the past week were you bored?
 - Not often? 10% of the time? Less than I was last week?



Types of responses & stats: What you ask for matters

- Nominal or categorical (Chi-square, McNemar)
 - Which of these movies have you seen? American Gangster (Yes/No)? The Bee Movie (Yes/No)? Saw IV (Yes/No)?
- Ordinal (Wilcoxon signed ranks test)
 - How would you rate your history teacher? Excellent? Very Good? Good? Poor?
- Numerical or Interval (ANOVA, linear regression)
 - How old are you?



What category is this?

- Names of STDs
- Patient satisfaction scores: excellent, very good, good, average, poor
- Patient's weight and height
- Patient's income (in \$5,000 increments)
- Patient's income (exact numbers)
- Patient's address

Sound research design: Basic designs

Highest

☑ Experimental: randomization and allocation (give treatment)

- Quasi-experiment: allocation only

Adjusts for differences in known and unknown factors

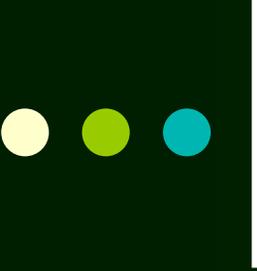
☑ Observational:

- Cohort (exposure → outcome)
- Case-control (outcome → exposure)
- Cross-sectional (surveys) (exposure + outcome simultaneously)

Use statistics to adjust known factors

Evidence for Causality

Lowest



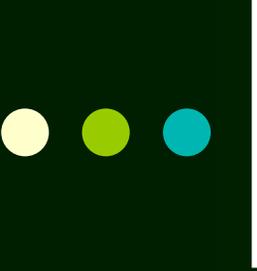
What type of study design is this?

1. A team of experts spent 10 days interviewing all full-time County employees
2. Interviews conducted over 1 month to find out teen's views of the quality of their education
3. The goals and aspirations of the 1990 graduates of the three leading high schools are followed over 15 years. Each year, the 1990 grads are interviewed and filmed

● ● ● | Sound research sampling

- A sample is a subset of a population
- The best sample is representative of target population (e.g., LA County)
 - Eligibility criteria: who's in & who's out
- Methods
 - Probability sampling
 - Simple random sampling, stratified random sampling, systematic sampling, cluster sampling
 - Nonprobability sampling
 - Convenience sampling, snowball sampling, quota sampling
- Remember...no sample is perfect





What type of sampling is this?

1. The rangers at five national parks are each asked to recommend two other rangers
2. The names of all teens who have been incarcerated within the last 6 months will be written individually on a piece of paper. The names will be placed in a glass jar. A blindfolded referee will select 10 names to serve on a focus group

Reliability and validity

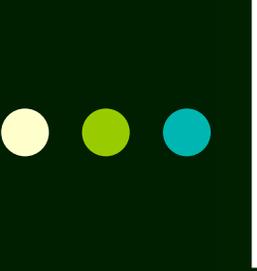
○ **Validity indicates how well an item (question) measures what it sets out to measure**

- Face: Does it look right?
 - Content: What do the experts say?
 - Criterion: How does this item measure up to gold standard?
 - Construct: After many years, does the item still hold up?
- Before survey is conducted
-

○ **Reliability is a statistical measure of how reproducible the survey instrument's data are**

- Alternate-form: Uses two differently worded items to obtain same info
- Test-retest: Measures stability of response among groups
- Intraobserver: Measure stability of response in same individual
- Internal consistency: Measures how well items in a scale vary
- Interobserver (Not used on self-administered surveys): Measures how two or more respondents rate the same phenomenon

○ **A reliable instrument is consistent; a valid one is accurate**



What are the different types of surveys?

- ☑ Self-administered questionnaires
 - Interviews
 - Structured record reviews: data abstraction form
 - Structured observations: observe and record children's use of fitness equipment for 10 minutes each day

Advantages & disadvantages of self-administered questionnaires

- Advantages:
 - Cost: Lower than in-person and telephone interviews
 - Wider coverage (mail surveys)
 - Larger samples
- Disadvantages
 - Lower response rates
 - Problems with literacy and language
 - Absence of skip patterns

Example of a self-administered questionnaire

What are the steps in implementing a questionnaire?

- ☑ Setting objectives for information collection
- ☑ Designing research
- ☑ Preparing a reliable and valid data collection instrument (including questionnaire specifications and the correct translated versions)
- ☑ Administering and scoring the instrument
- ☑ Entering data
- ☑ Analyzing data
- ☑ Reporting the results



Role of Epidemiologists &
Research Analysts in
questionnaire development,
design & analysis

● ● ● | What is the difference between epidemiologists and research analysts (RA) in DPH?

Answer: Do you want the truth?

Respondent: Yes?

Answer: It depends.

Respondent: On what?

Answer: On who holds the position.

Respondent: What???



But, what should I do if I want to conduct a survey?



○ Responsibility of Requestors:

- Literature review, behavioral models and constructs, independent and dependent variables
- Know possible outcomes: What do you think the biggest and smallest possible outcome could be?
- Significance level: What do you want the probability of observing a difference to be when none exists?
1%? 5%? 10%? (↓% ↑N)
- Power: What do you want the probability of detecting a meaningful effect to be, if one were to occur?
80%? 90%? 99%? (↑% ↑N)
- Select group to pilot questionnaire

Sample size estimates

So, what should the Epidemiologist/RA do?

- Determine study design
- Calculate the sample size
- Create a codebook
- Generate database in Access
- Devise an analysis plan
- Generate tables and figures
- Summarize tables and figures (optional)



Los Angeles County Department of Public Health
RAPID ASSESSMENT: NUTRITION SURVEY
Patient Survey

id

ID _ _ _ _

Today's Date: m m d d y y

Site Code

Location Zip Code

date - -

site

locZip

lang: 1-English; 2-Spanish

We are doing a very short survey on nutrition that will take less than 5 minutes to complete. Your responses to our questions will be kept strictly confidential, and we will not ask for your name, address, or telephone number. We appreciate your participation:

First, we need some information about you...

q1
q2
q3
q4

1. Age: _____

2. Gender: Female | Male 2

3. Zip Code of Residence: _____

4. Which race/ethnicity do you identify as? (Check only one)

African American/Black 1

Asian/Pacific Islander 2

Hispanic/Latino 3

White/Non-Hispanic 4

Native American/Alaskan Native 5

Mixed/Multiethnic 6

q5

5. What best describes your education? (Check only one)

Completed less than high school 1

High school graduate or equivalent 2

Some college, junior college or trade school 3

College graduate/Postgraduate 4

Now, please tell us about your experiences at McDonald's...

q11a 11. a. How often do you go to McDonald's? Do you go there at least... (Check only one)
 Once a day Once a week Once a month A few times a year Almost never/Never

q11b b. When you go to McDonald's, how often do you use the drive-thru? (Check only one)
 All of the time Most of the time Half the time Less than half the time Almost never/Never

q12 12. The last time you went into a McDonald's restaurant, do you remember seeing any calorie information about McDonald's foods and drinks when deciding what to order?
 Yes No

Now, we would like you to answer a few questions about what you think about nutrition and calories...

q13a 13. a. How many calories should a person your age, height, and weight eat each day? (Check only one)
 1000 or less 1001 - 1500 1501-2000 2001-2500 2501-3000 3001 or more

q13b b. How often do you look at calorie information on packages and cans of food and drinks sold in grocery stores when you are buying something for the first time? (Check only one)
 Always Most of the time Sometimes Rarely Never

q13c c. How important is it for you to have calorie information listed on packages and cans of food and drinks sold in grocery stores? (Check only one)
 Very important Important Somewhat important Not important at all

Nutrition Survey Codebook

Position	Variable Name	Variable Label	Type	Informat (Columns)	Coded Values
1	id	ID Number	Numeric	4	N/A
2	date	Date	MM/DD/YYYY	MMDDYY10.	N/A
3	site	Site Code	Numeric	4	1000 = Curtis Tucker HC 2000 = Torrance HC 3000 = South HC 4000 = Central HC 5000 = North Hollywood HC 6000 = Hollywood-Wilshire HC
4	loczip	Location Zip Code (precoded)	Numeric	5	90301 = Curtis Tucker HC 90502 = Torrance HC 90002 = South HC 90012 = Central HC 91601 = North Hollywood HC 90038 = Hollywood-Wilshire HC
5	lang	Survey language	Numeric	1	1 = English 2 = Spanish
6	q1	Age (Q1)	Numeric	2	99 = Missing
7	q2	Gender (Q2)	Numeric	1	1 = Female 2 = Male 7 = Refused 9 = Missing
8	q3	Patient's Zip Code (Q3)	Numeric	5	99999 = Missing

2007 Nutrition Survey
Dr. Paul Simon, Principal
Investigator

- PAGE 1 -

ID NO.:

DATE:

SITE CODE*

LOCATION ZI

LANGUAGE:

q1:

q2:

q3:

q4:

q5:

q6:

q7a:

q7b:

q8:

q9a:

q9b:

q10:

q10a:

*Site Code: 1000 = Tucker HC; 2000 = Torrance
HC; 3000 = South HC; 4000 = Central HC; 5000
= North Hollywood HC; 6000 = Hollywood; 9999
= Missing

- PAGE 2 -

q11a:

q11b:

q12:

q13a:

q13b:

q13c:

q13d:

q13e:

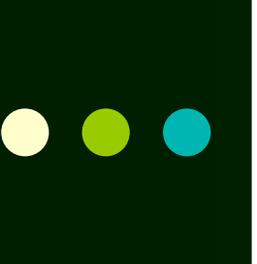
q13f:



ENTERE

ENTRY DATE:

Created: 09/27/07



So, if you don't want to conduct a survey, where can you find data?

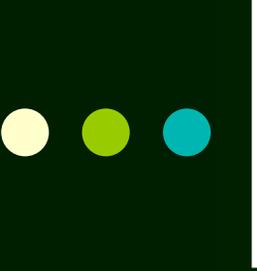
Data Resources: DPH & Academic

DPH

- LA County Health Survey (OHAE)
- Injury Mortality & Hospitalization Data (IVPP)
- STD*Casewatch[®] (STD Program)
- Foodborne Illnesses in Los Angeles County (ACDC)
- HIV/AIDS Surveillance Database (HIV Epidemiology)
- Public Health Information Systems (PHIS)

Academic

- California Health Interview Survey (UCLA)
- Los Angeles County Cancer Surveillance Program [CSP] (USC)
- UCLA Institute for Social Science Research (ISSR) Data Archives: <http://www.sscnet.ucla.edu/issr/da/>



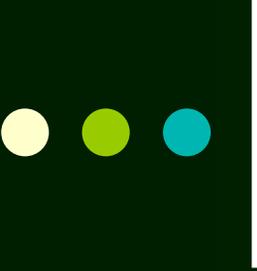
Data Resources: State & Federal

State

- California Birth Defects Monitoring Program (CBDMP)
- Healthcare Information Resource Center, Office of Statewide Health Planning and Development
- Fitnessgram (OHAE)
- The 2001 California Dietary Practices Survey (CDPS)
- The California Children's Healthy Eating and Exercise Practices Survey (CalCHEEPS), 2003

Federal (see <http://www.cdc.gov/DataStatistics/>)

- National Behavioral Risk Factor Surveillance System (BRFSS)
- Youth Risk Behavior Surveillance (YRBSS)
- National Health Interview Survey (NHIS)
- National Health and Nutrition Examination Survey (NHANES)
- National Immunization Survey (NIS)
- National Asthma Survey (NAS)



How do you request data from analysts in DPH?

- In DPH, go to Program website and obtain phone number
 - Ask to speak to Epidemiologist or Research Analyst responsible for generating the program reports
 - If you are requesting a data run, be prepared to give the following information:
 - Specific disease or condition (e.g., chlamydia, measles)
 - Time period (e.g., month and/or year)
 - Age categories (e.g., 5-9, 10-14, 15-19)
 - Gender (e.g., male and/or female)
 - Race/Ethnicity (e.g., African-Americans, Asians)
 - County (e.g., Los Angeles)
 - Census tract, health district, SPA

● ● ● | But, I don't know any
Epidemiologists...

Lisa V. Smith, MS, DrPH

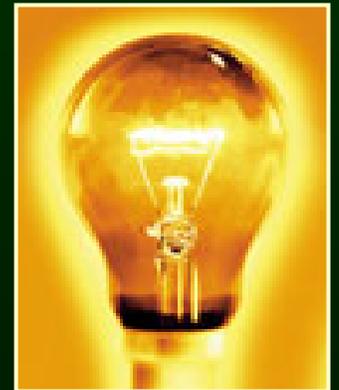
Chief, Surveillance, Epidemiology and Evaluation
Division of Chronic Disease and Injury Prevention

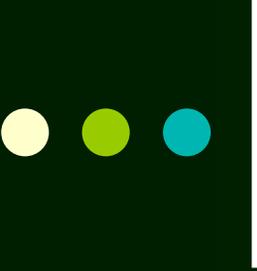
3530 Wilshire Boulevard, Suite 800

Los Angeles, CA 90010

213.351.7332

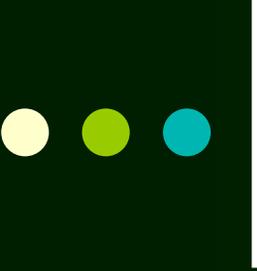
lismith@ph.lacounty.gov





Institutional Review Board (IRB) Approval

- ☑ It is strongly recommended that you submit an application to the County IRB before you begin your research
 1. Required any time you are going to interact with patients and their information for the purposes of research. This includes pilot data.
 2. Studies evaluating existing projects are usually exempt from IRB, but you should obtain a written waiver from the IRB committee.
 3. If you are affiliated with more than one agency, you will have to get clearance from each one.
 4. Most scientific journals require that authors submitting manuscripts show evidence of IRB clearance.
- ☑ For additional information, contact Olga Coronado at ocoronado@ph.lacounty.gov



Acknowledgments

Unless indicated, all survey information was abstracted from the *The Survey Kit* (Sage Publications, 1995):

1. Fink A. *The Survey Handbook*
2. Fink A. *How to Ask Survey Questions*
3. Bourque LB and Fielder EP. *How to Conduct Self-Administered and Mail Surveys*
4. Fink A. *How to Design Surveys*
5. Fink A. *How to Sample Surveys*
6. Litwin MS. *How to Measure Survey Reliability and Validity*



Thank you very much!!!

Lisa V. Smith, MS, DrPH

Chief, Surveillance, Epidemiology and Evaluation

Division of Chronic Disease and Injury Prevention

3530 Wilshire Boulevard, Suite 800

Los Angeles, CA 90010

213.351.7332

lismith@ph.lacounty.gov





Topic Exercise

TASK: In groups, identify 5 topics that can be explored using a survey design.