

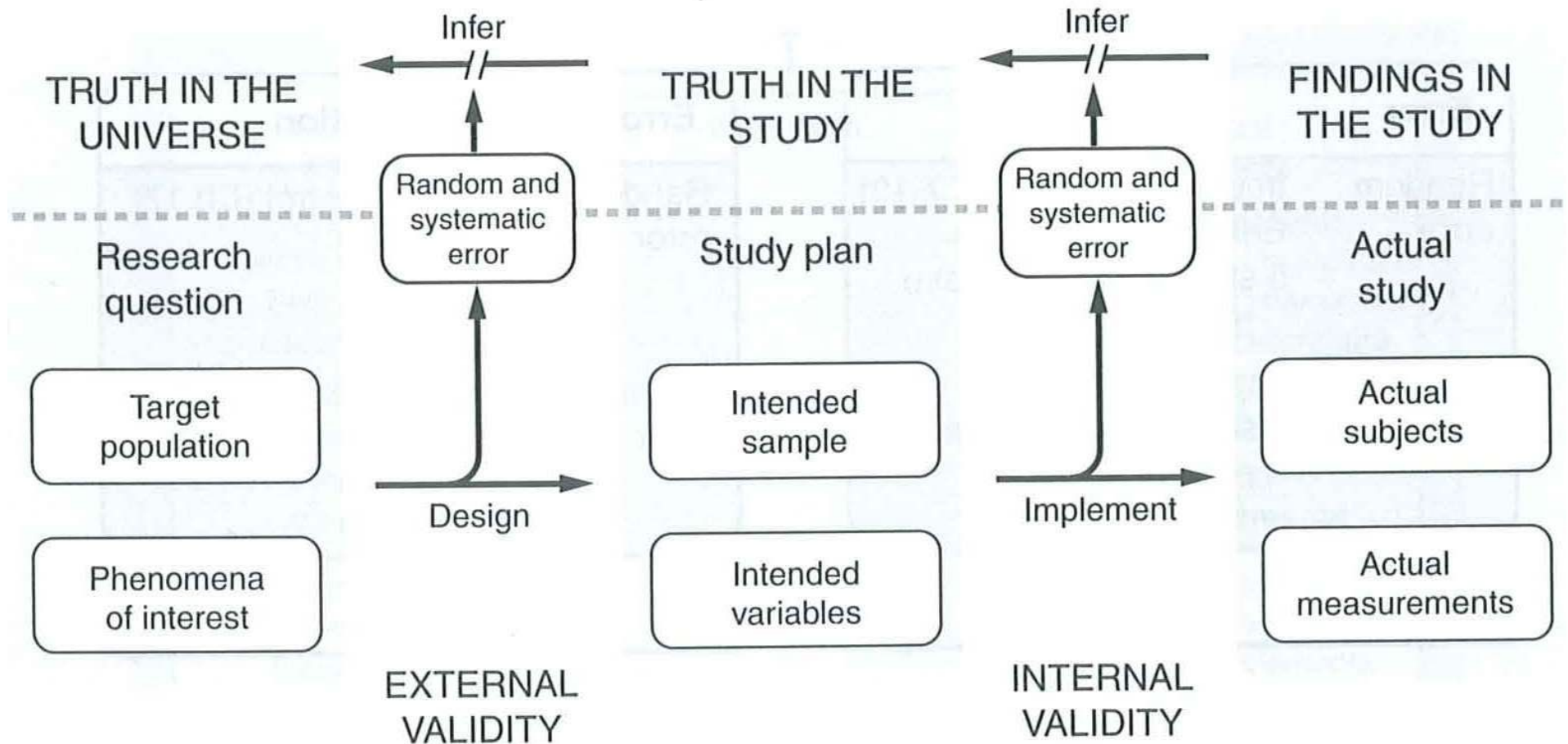
Public Health Sciences 310  
Epidemiologic Methods

Lectures 13  
Data Collection Instruments and Field  
Operations

February 15, 2024

Brian C.-H. Chiu

# How Research Works



# 14-STEP Program for Epidemiology Studies

- Define questions/hypotheses based on current literature
- Choose appropriate study design
- Define groups for comparison
- Define exposure and outcome variables
- Define extraneous variables to be measured
- Calculate sample size
- Develop/choose measurement instruments
- Develop protocol and train staff
- Recruit subjects, collect data, quality control procedures
- Process data
- Analyze data using appropriate statistical methods
- Determine if valid statistical association (assess chance, confounding, interactions and/or bias)
- Judge if association is causal
- Disseminating the results

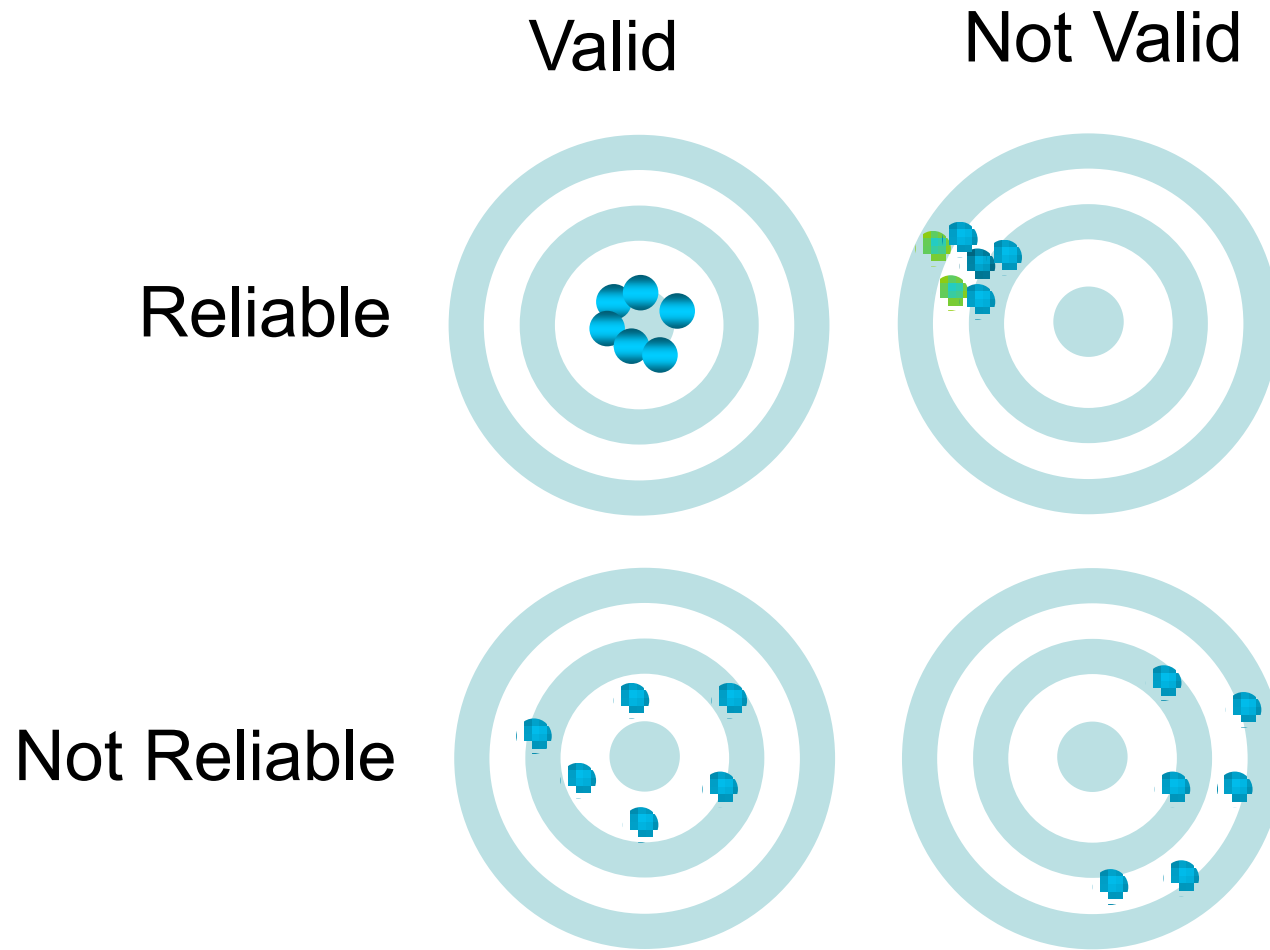
# Methods of exposure measurement

Measurement method	Data		Time		Type of exposure	
	Subjective	Objective	Present	Past	Personal attribute	Environmental exposure
Personal/phone interview	+		+	+	+	+
Self-administered questionnaire	+		+	+	+	+
Diary	+		+		+	+
Reference to records		+	+	+	+	+
Physical or chemical measurements on subject		+	+	?	+	+
Physical or chemical measurements of environment		+	+	+		+

# Data Collection

- Questionnaires (most common method)
  - Major administration methods:
    - Mailed questionnaire
    - Telephone interview
    - In-person interview
    - Computer-assisted interview (CATI or CAPI)
    - Combinations
- Physical Examinations
- Biospecimen Collection

# New instrument: should assess validity (accuracy) vs reliability (precision)



# Assessing Validity

- Continuous variables
  - Scatter plots
  - Regression
  - Comparison of means
  - Analysis of variance
- Categorical variables
  - Percent agreement
  - Sensitivity/Specificity
  - Positive/Negative Predictive Value
  - Kappa statistic

# Assessing Reliability

- Test-retest, inter-rater agreement, replicate samples
- Continuous variables
  - Scatter plots
  - Regression
  - Coefficient of variation
  - Intraclass correlation coefficient
- Categorical variables
  - Percent agreement
  - Kappa statistic



# Percent Agreement

- The number of concordant observations divided by the total.

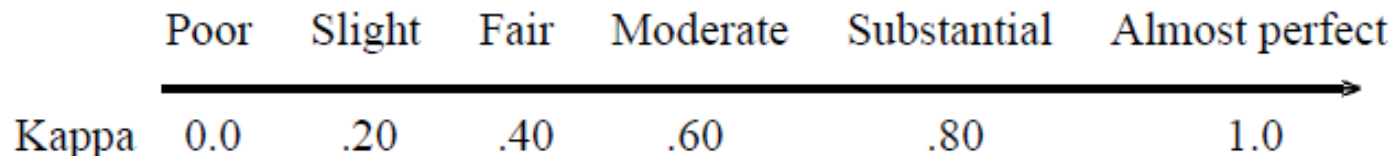
		Observation 2	
		Test +	Test -
Observation 1	Test +	a	b
	Test -	c	d

$$PA = \frac{(a + d)}{(a + b + c + d)} \times 100$$

# Kappa Statistic

- Measures whether agreement exceeds chance for categorical variables
- $K = (P_o - P_e) / (1 - P_e)$

## Interpretation of Kappa



<u>Kappa</u>	<u>Agreement</u>
< 0	Less than chance agreement
0.01–0.20	Slight agreement
0.21– 0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–0.99	Almost perfect agreement

# Kappa Set Up

		Observation 1		
Observation 2		Test +	Test -	Total
	Test +	15	5	20
	Test -	10	70	80
	Total	25	75	100

$$\text{Observed agreement } (P_o) = \frac{15+70}{100} = 0.85$$

$$\text{Expected agreement } (P_e) = \left[ \left( \frac{25}{100} \right) \times \left( \frac{20}{100} \right) + \left[ \left( \frac{75}{100} \right) \times \left( \frac{80}{100} \right) \right] \right] = 0.65$$

$$\text{Kappa, } k = \frac{(P_o - P_e)}{(1 - P_e)} = \frac{(0.85 - 0.65)}{(1 - 0.65)} = 0.57$$

# Field Operations

- Prepare the instrument for use in the field
- Pretest and pilot test the instrument
- Develop protocol
- Recruit and train staff
- Purchase equipment
- Identify and recruit subjects
- Administer the instrument and collect biological samples
- Manage the data

# Ensuring Good Data Quality

- Ensure adequate training, retraining, and supervision of field workers
- Check samples of data-collection forms to assess their completeness and accuracy
- Assess interviewer's performance regularly by watching/listening to interview
- Assess reliability of the data obtained in the two interviews
- Tabulate the most important variables by interviewer to assess inter-interviewer variability
- Re-interview a random sample of subjects