SESSION 3:

APPRAISING THE LITERATURE: LEVEL & QUALITY OF EVIDENCE

Tam H Nguyen, PhD, MSN/MPH, RN

Appraising the literature: Level & Quality of Evidence

"A 21st century clinician who cannot critically read a study is as unprepared as one who cannot take a blood pressure or examine the cardiovascular system."

(BMJ 2008; 337: 704-705)

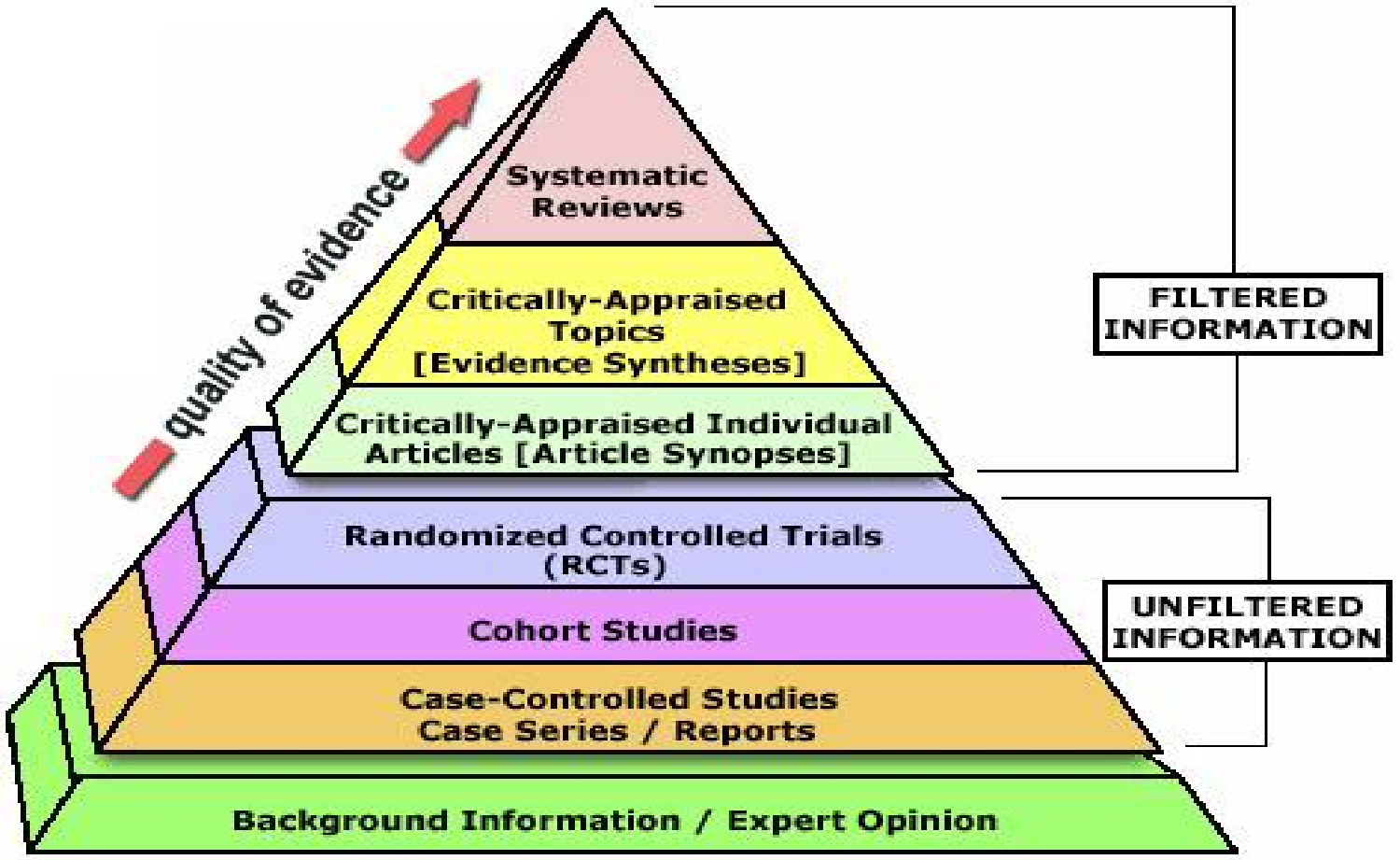
Skills needed to appraise literature

- Be able to:
  - Identify level and quality of evidence
  - Interpret study results
  - Assess for bias

Johns Hopkins EBN Model:
Strength of evidence: Depends on study design

STRENGTH of the Evidence	
Level I	Experimental study/randomized controlled trial (RCT) or meta analysis of RCT
Level II	Quasi-experimental study
Level III	Non-experimental study, qualitative study, or meta-synthesis.
Level IV	Opinion of nationally recognized experts based on research evidence or expert consensus panel (systematic review, clinical practice guidelines)
Level V	Opinion of individual expert based on non-research evidence. (Includes case studies; literature review; organizational experience e.g., quality improvement and financial data; clinical expertise, or personal experience)

Other evidence hierarchies



Types of studies/papers you may find in the literature

- Primary studies (non-filtered information)
  - Quantitative
  - Qualitative
  - Mixed Methods
- Evidence Summaries (filtered information):
  - Meta Analysis
  - Systematic Reviews
  - Integrative Reviews
  - Clinical Practice Guidelines
  - Article synopsis
  - Content Expert Opinion

Types of studies/papers you may find in the literature

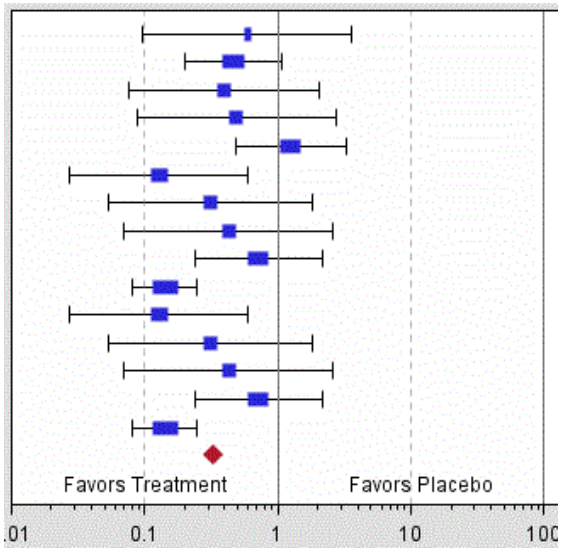
- Research
  - Quantitative
    - Meta-Analysis, Systematic Reviews
    - Randomized Control Trials (RCTs)
    - Quasi-Experimental
    - Non-Experimental (i.e., observational)
  - Qualitative
    - Meta-synthesis
    - Phenomenology
    - Grounded Theory
    - Ethnography
- Non-research
  - Integrative Reviews
  - Clinical Practice Guidelines
  - Content Expert Opinion

Systematic Reviews & Meta-Analysis:

- There is a growing body of information (information overload). Systematic reviews & meta-analysis can serve as a solution

THE COCHRANE COLLABORATION®

- Systematic reviews & meta-analysis do not always exist for all clinical questions



- Not all research is created equally (research hierarchy)

Quantitative Research Designs

Why read the 'Methods' section?

Types of Quantitative Designs

Increase in Control & Causal Association

- Experimental
- Quasi-Experimental

Examining Causality ***through intervention/manipulation***

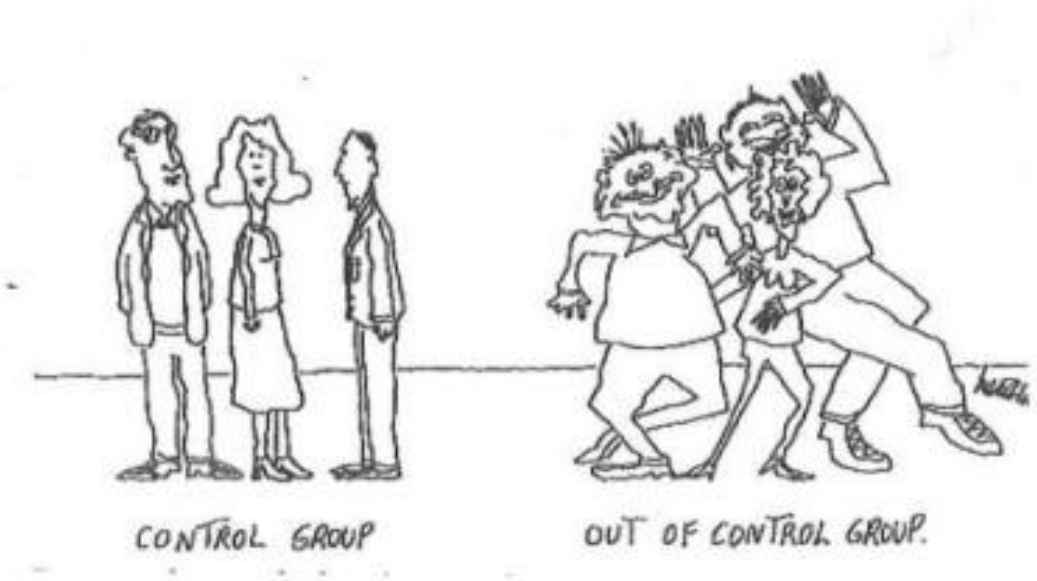
- Non-experimental

Describing, explaining, or predicting relationships between variables ***through observation***

*Critical component: Design should match research question

Elements of True Experiment

1. Randomization (random sampling vs. assignment)
2. Control
3. Manipulation



Randomization (Random Assignment)

- Distribution of subjects to an experimental and a control group on a purely random basis
- Each subject has equal and known probability of being assigned to either group
- Eliminates systematic bias
- Assumes that extraneous variables will occur equally in the two groups, controls for any other possible influence of IV on DV
- Thus, the difference in the dependent variable can be assumed to happen because of the experiment

Example: (Kim et al., 2014)

The Effect of a Community-Based Self-Help Multimodal Behavioral Intervention in Korean American Seniors With High Blood Pressure

Table 1. Sociodemographic characteristics of participants at baseline (n = 369)

Characteristics	Total (n = 369)	Intervention (n = 184)	Control (n = 185)	*P* value
Age, y, mean (SD)	70.9 (5.3)	70.6 (5.0)	71.2 (5.6)	0.29
≤69, no. (%)	155 (42.0)	79 (42.9)	76 (41.1)	
70–79, no. (%)	190 (51.5)	97 (52.7)	93 (50.3)	
≥80, no. (%)	24 (6.5)	8 (4.3)	16 (8.6)	
Sex				0.29
Male, no. (%)	111 (30.1)	60 (32.6)	51 (27.6)	
Female, no. (%)	258 (69.9)	124 (67.4)	134 (72.4)	
SBP, mm Hg, mean (SD)	141 (19)	141 (17)	140 (20)	0.69
DBP, mm Hg, mean (SD)	79 (11)	79 (11)	79 (11)	0.91
BMI, kg/m^2, mean (SD)	256 (3.2)	25.5 (3.2)	25.7 (3.3)	0.70
Education, y, mean (SD)	11.2 (4.3)	11.2 (4.2)	11.1 (4.4)	0.73

Six types of "True Experiments" (FYI)

1. Two group, pre-test/post-test
2. Two group, post-test only
3. Solomon four group
4. Multiple experimental groups
5. Factorial design
6. Cross-over design

Common Symbols of Experimental Designs

R = random assignment

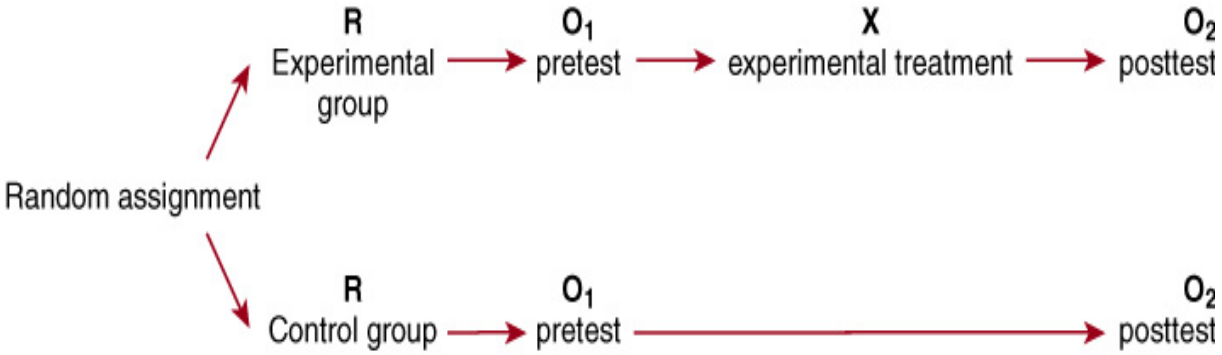
$\mathbf{O_1}$ = 1st data point (pre-test)

X = intervention

$\mathbf{O_2}$ = 2nd data point (post-test)

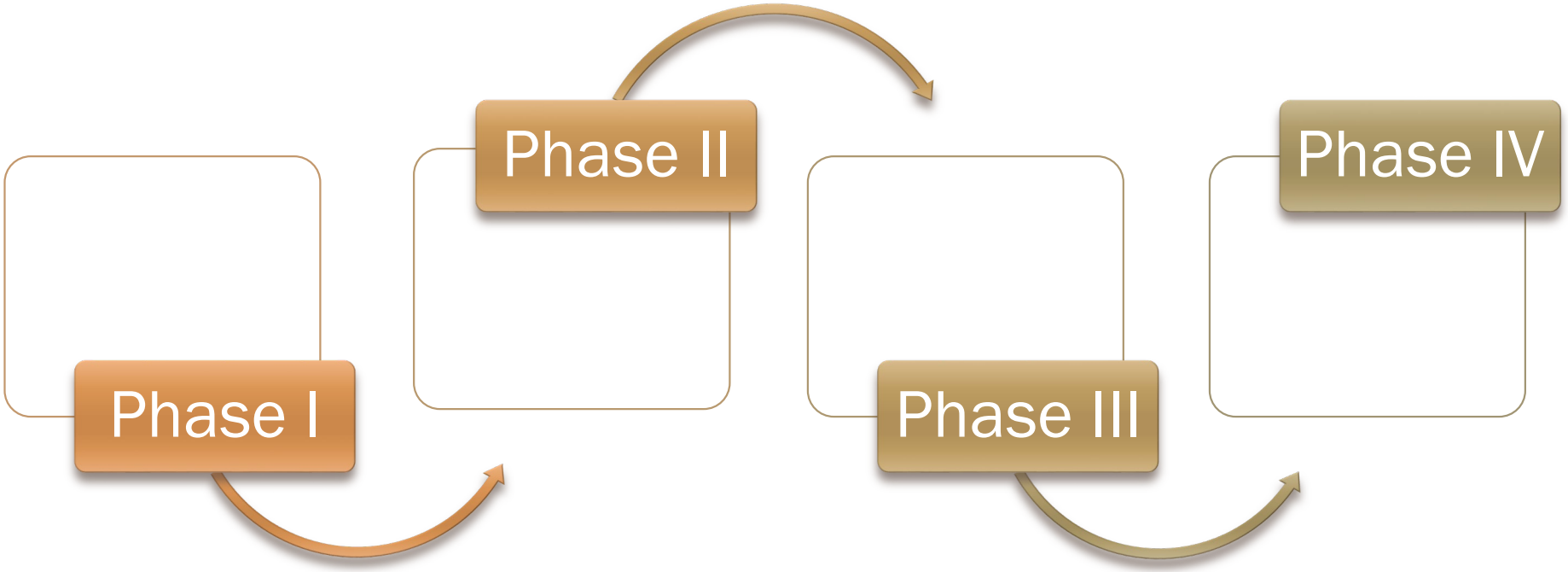
Two group, Pre-test/post-test (Classic RTC)

- Design:



- Pros:
  - Have a baseline to compare
- Cons:
  - Have to collect data twice, washout period, potential drop out

Phases of Clinical Trials



*Phase of Clinical Trial

Pilot Trials

In Phase I trials, researchers test an experimental drug or treatment in a small group of people (20-80) for the first time to evaluate its **safety**, determine a safe dosage range, and identify side effects.

In Phase II trials, the experimental study drug or treatment is given to a larger group of people (100-300) to test its **efficacy** and to further evaluate its safety.

Population Base Trials

In Phase III trials, the experimental study drug or treatment is given to large groups of people (1,000-3,000) to confirm its **effectiveness**, monitor side effects, compare it to commonly used treatments, and collect information that will allow the experimental drug or treatment to be used safely.

Cost Effectiveness; Comparative Effectiveness

In Phase IV trials, **post marketing studies** delineate additional information including the drug's risks, benefits, optimal/off-label use.

*** Impacts quality, rather than level of evidence**

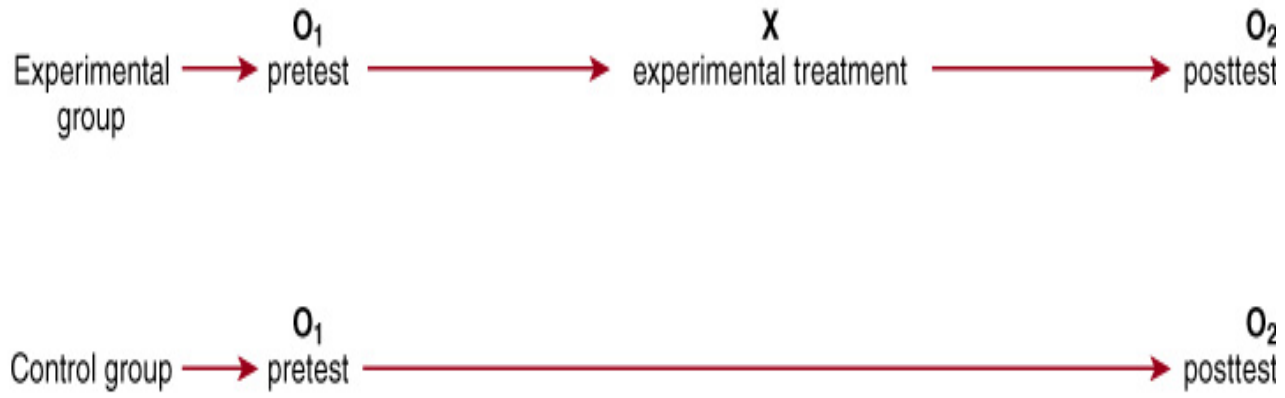
Quasi-experimental designs

Quasi-experimental designs

- Defining feature: lack randomization and/or control group
- Not as strong as experimental studies in determining cause-and-effect
- Types:
  1. Non-equivalent control group (no R)
  2. Time series (no R, no C, several Os)
  3. Pre-experimental, one group (no R, no C)

Non-equivalent control group

- Design:



- Pros:
  - Convenient, participants may be able to pick what group they are in
- Cons:
  - Harder to infer causality, must evaluate how controls were **MATCHED**

Quasi-Experimental Designs: Thoughts

- Useful when full experimental control is not possible; may be the only way to evaluate some treatments.
- They can be practical, feasible, and with lots of evidence, generalizable; adaptable to real-world environments.
- Less able to determine that the intervention/ treatment caused the changes in the dependent variable than experimental studies
- However, they lead to increased knowledge about the treatment and dependent variable (e.g. pilot studies)

Example of Quasi-experimental study

WOUND CARE

The Effect of a Silicone Border Foam Dressing for Prevention of Pressure Ulcers and Incontinence-Associated Dermatitis in Intensive Care Unit Patients

Kyung Hee Park

PURPOSE: We measured the effect of a silicone border foam dressing on the development of pressure ulcers (PUs) and incontinence-associated dermatitis in intensive care unit (ICU) patients.

DESIGN: Nonrandomized comparison cohort (quasi-experimental) study.

SUBJECTS AND SETTINGS: One hundred and two patients (>40 years of age) with a Braden Scale score of 16 or less who were admitted to 2 ICUs at the Samsung Medical Center in Seoul, South Korea, participated in the study.

METHODS: Fifty-two subjects were assigned to the experimental group (standard PU preventive care routine plus application of the silicone border foam dressing), and 50 subjects were assigned to the control group (standard PU preventive care alone). The number of patients who developed PU in the experimental group was compared with that from the control group using the chi-square test (χ^2). The IADS score of the experimental group was measured and compared with those of the control group, using an independent *t* test. Logistic regression was carried out to analyze the relationship between the IADS score and PU development.

Non-experimental Designs

Why are non-experimental designs useful?

Non-experimental designs

- Correlational
  - Descriptive Cross-sectional
  - Cohort design
  - Retrospective case-control design

Non-experimental designs

- Independent variable is not manipulated
- Do not enable the investigator to establish cause-and-effect relationship between variables (“*correlation does not prove causation*”)
- The researcher has the least amount of control in this type of quantitative study
- Usually cross-sectional, but can be longitudinal or retrospective

Descriptive Designs

- Search for accurate information about the characteristics of particular subjects, groups, institutions or situations or about the frequency of a phenomenon's occurrence.
- Used when little is known about the phenomenon.
- Variables such as opinions, attitudes, facts

Surveys

- Data collected through questionnaire or interview
  - "Self-Report"
  - A survey interview uses structured questions with "forced choice" responses (rather than open-ended questions as in a qualitative study)
- Pros:
  - Obtain a great deal of data from a large population
- Cons:
  - Data/findings can be superficial since focus is on breadth, not depth

Cross-sectional correlation

Relationship Between Nurses' Organizational Trust Levels and Their Organizational Citizenship Behaviors

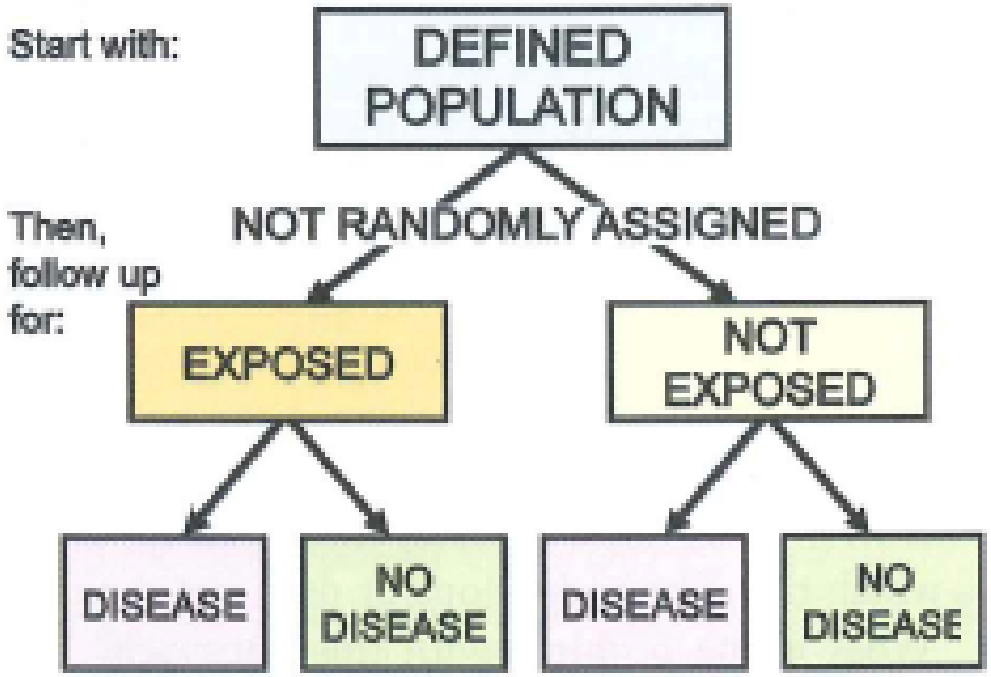
Serap Altuntas, BSN, PhD[1] & Ulku Baykal, BSN, PhD[2]

Design and Methods: Nurses who had completed their orientation from a total of 11 hospitals with bed capacities of 100 and located in the European district of Istanbul were included in the sample for this study. Formal, written applications and approval of the ethical committee were obtained from concerned institutions before proceeding with the data collection step. The Organizational Trust Inventory and the Organizational Citizenship Level Scale, a questionnaire form including five questions regarding nurses' personal characteristics, were used in data collection. Data collection tools were distributed to 900 nurses in total, and usable data were obtained from 482 nurses. Number and percentage calculations and Pearson correlation analysis were used to assess research data.

Conclusions: The findings attained in this study indicated that the organizational trust the staff had in their institutions, managers, and coworkers influenced the organizational citizenship behaviors of conscientiousness, civic virtue, altruism, and courtesy, whereas it had no effect on sportsmanship behavior. Nurse managers should introduce studies to improve their subordi-

Cohort study

- Good for prognosis & etiology questions

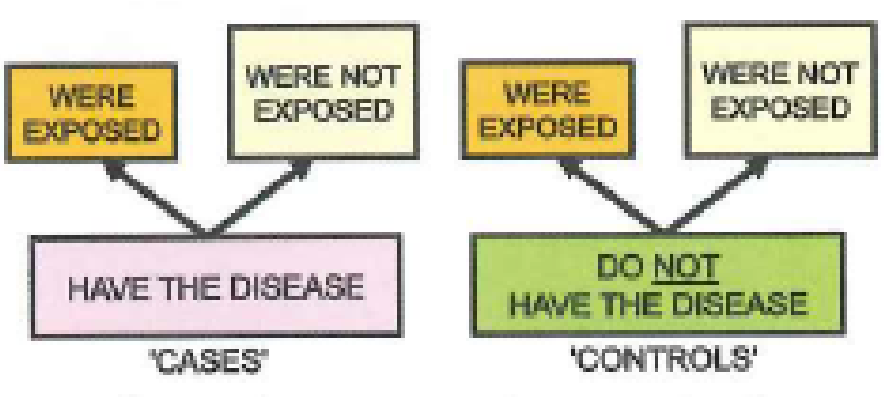


Example of Cohort Study: Nurses' Health Studies

http://www.channing.harvard.edu/nhs/

- One of the largest and longest running investigations of factors that influence women's health
- Started in 1976 and expanded in 1989
- Sample: 238,000 dedicated nurse-participants
- New insights on health and disease
  - Prevention of cancer
  - Cardiovascular disease
  - Diabetes and other conditions
- Have shown that diet, physical activity and other lifestyle factors can powerfully promote better health

Retrospective Case control



- Important point is that cases need to be matched on important variables
- Less expensive than cohort studies, especially for rare diseases

Let's practice: What study design was used?

[redacted] To test the efficacy of the intervention protocol, we used a [redacted] trial with the intervention delayed for the control group. Using adaptive stratified randomization, we selected 22 Korean American churches and senior centers as intervention and control group sites, depending on size or location. Potential participants were screened, enrolled, and tested at each site.

Qualitative Research

Each type of qualitative research is guided by a particular philosophical stance, and results in different "knowledge"

Research design	Discipline	Domain	Outcome
Ethnography	Anthropology	Culture	Rich & holistic description of a culture
Phenomenology	Psychology/philosophy	Lived experiences	Meaning/ essence of lived experience(s)
Grounded Theory	Sociology	Social settings	A theory, driven by a core variable and the basic social process/ patterns around it

Polit & Beck, 2012

Common Elements

- Purpose: identify, describe and understand
- Inductive methods
- Data are words or text
- Small sample size
- Preserve the *subjective experience* of the study participants
- Findings are new concepts, definitions, themes, or theories

Mixed Methods: Philosophical Pluralism

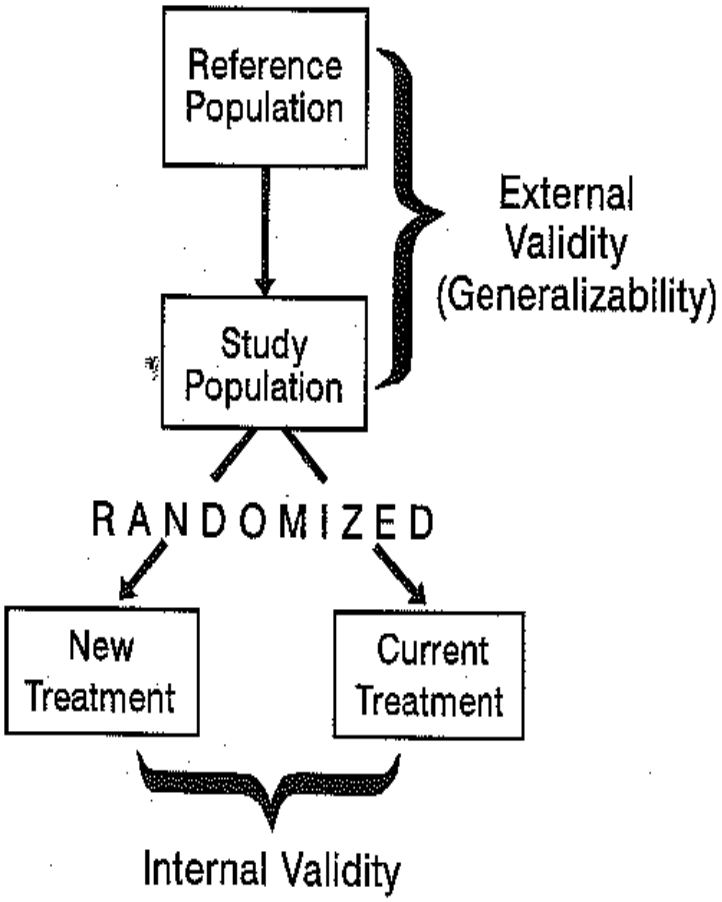
- Mixed-methods research (John Creswell)
  - Integrates qualitative and quantitative data and strategies in a single study or coordinated set of studies
  - Advantages:
    - Use of words and numbers, the two languages of human communication
    - Quicker feedback loops between hypothesis generation and testing
    - Method triangulation strengthens the ability to make inferences

Another critical skill needed to appraise evidence:

Assessing for bias

Internal Validity vs. External Validity

- **Internal Validity**: refers to the accuracy of study of results. It is determined by assessing how well a study was conducted (research design, operational definitions used, how variables were measured, what was/wasn't measured, etc.). Ultimately we want to know how confidently one can conclude that the observed effect(s) were produced solely by the independent variable and not extraneous ones.
  - Experimental designs: Did the IV cause the change?
  - Descriptive designs: Is this relationship real?
- **External Validity**: refers to the generalizability of study findings to other samples.
  - Can we apply the results?



Main sources of bias impacting internal validity

- Confounding
- Mortality/Attrition
- Lack of Power (low sample size)

Main sources of bias impacting external validity

- Study Sampling Strategy

Another critical skill needed to appraise evidence:

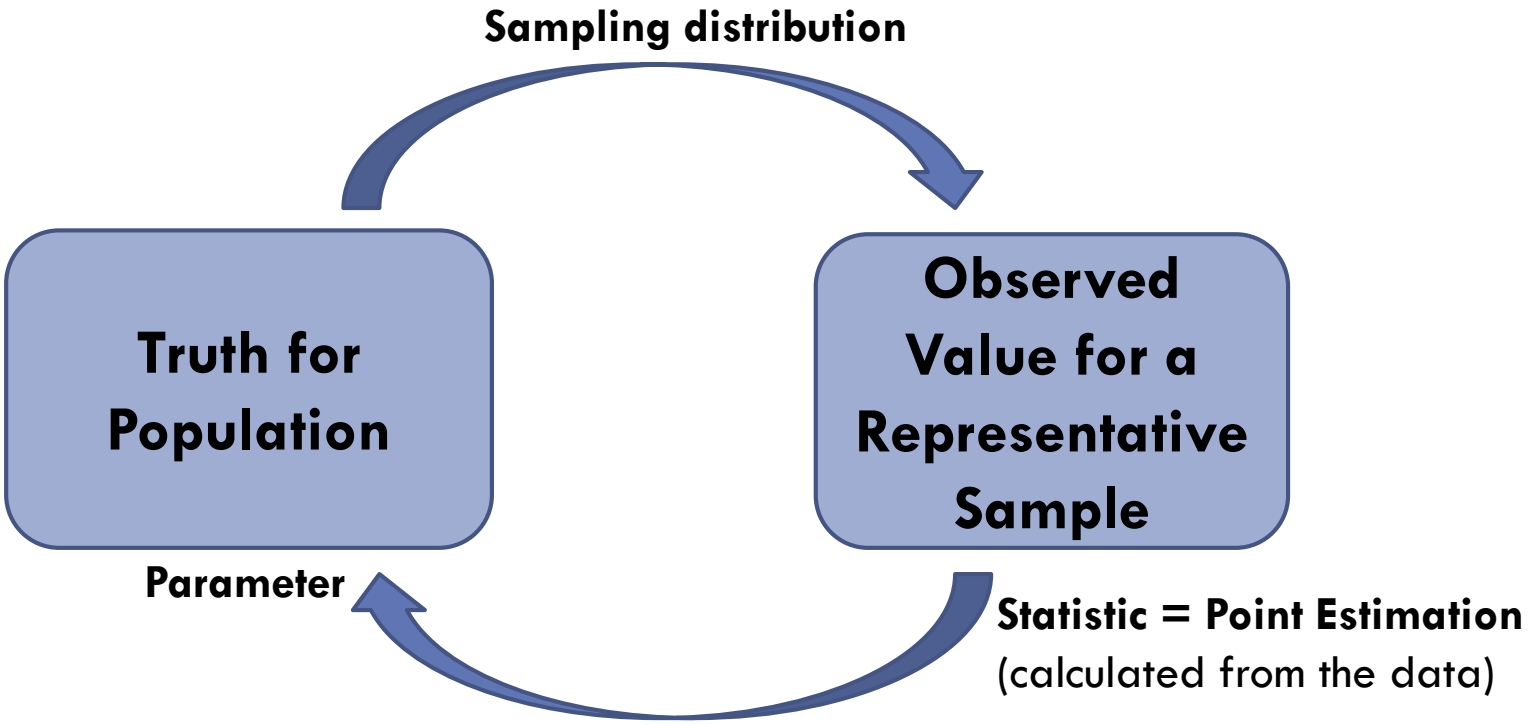
Understanding hypothesis testing & interpreting results (p-values, 95% CI)

Critically considering “clinical significance”

What is hypothesis testing?

- TOOL that helps researchers, clinicians, or administrators reach a decision concerning a population by examining a sample from that population
  - Given the observed data, do we reject or fail to reject a pre-specified null hypothesis in favor of an alternative?
  - “Significance testing”

Search for "Truth" given limited data/resources



Statistical Inference

- 95% Interval estimate
- Hypothesis testing

Steps in Hypothesis Testing

- Set up a null hypothesis (Ho) and alternate hypothesis (Ha) based on your research question.
  - Note : the null hypothesis is often stated as "no relationship" or "no effect" "no association" between the IV and DV

    Ho: There is no difference
  - The alternate hypothesis is that "there is a difference" or "there is an association" between the IV and DV

    Ha: There is a difference
  - Start by assuming null is true, such that we can only "reject the null" or "fail to reject the null"

P-value

- P-value is a measure of how much evidence we have against the null hypothesis.
- The smaller the p-value the more evidence we have against the null hypothesis.
- Traditionally, researchers will reject the null hypothesis if the p-value is less than 0.05.

Let's practice

- Do adults who participate in an 8-week exercise program have lower LDL levels?

  Direction (1 sided or 2 sided):

  Ho:

  Ha:

- Does gender, functional status, number of factories within a geographic community, and number of days above 80 degrees have an impact on the respiratory status of elderly individuals?

  Direction (1 sided or 2 sided):

  Ho:

  Ha:

More Practice

- Among patients with dementia, does pain management decrease agitation?

  Direction (1 sided or 2 sided):

  Ho:

  Ha:

  p-value = 0.06

  Decision/conclusion:

More Practice

- Among adults with a terminal diagnosis, does the initiation of palliative care at the time of diagnosis impact quality of life?

Direction (1 sided or 2 sided):

Ho:

Ha:

p-value = 0.014

Decision/conclusion:

Confidence Intervals

- Definition:
  - An interval that expresses the uncertainty or variability (precision) associated with an estimate
  - 95% CI =
    $X \pm (1.96 * \mathbf{SEM})$
    $X \pm (1.96 * \mathbf{SD}/\sqrt{n})$
  - NOTE: SEM ≠SD!

- Interpretation:
  - "We are 95% confident that the interval covers (overlaps, contains) the true but unknown population mean."
  - "In repeated sampling from a normally distributed population, 95% of all intervals will include the population mean"

Confidence intervals

- Express the researcher's risk of being wrong
- With a 95% CI, researchers accept the probability that they will be wrong 5 times out of 100
- A 99% CI sets the risk at only 1 time out of 100

Practice interpreting 95% CI

- Mean LDL-cholesterol (mg/dL)
  - 130 (115 to 145)
  - 130 (127 to 133)
- Mean difference in LDL-cholesterol (mg/dL)
  - 10 (8 to 12)
  - 10 (-2 to 22)
- Prevalence of Type 2 DM
  - 5% (2% to 8%)
  - 5% (1% to 9%)

Zero in on findings…

The New England Journal of Medicine

VOLUME 347 JULY 11, 2002 NUMBER 2

A CONTROLLED TRIAL OF ARTHROSCOPIC SURGERY FOR OSTEOARTHRITIS OF THE KNEE

J. Bruce Moseley, M.D., Kimberly O'Malley, Ph.D., Nancy J. Petersen, Ph.D., Terri J. Menke, Ph.D., Baruch A. Brody, Ph.D., David H. Kuykendall, Ph.D., John C. Hollingsworth, Dr.P.H., Carol M. Ashton, M.D., M.P.H., and Nelda P. Wray, M.D., M.P.H.

Results At no point did either of the intervention groups report less pain or better function than the placebo group. For example, mean (±SD) scores on the Knee-Specific Pain Scale (range, 0 to 100, with higher scores indicating more severe pain) were similar in the placebo, lavage, and débridement groups: 48.9±21.9, 54.8±19.8, and 51.7±22.4, respectively, at one year ($P=0.14$ for the comparison between placebo and lavage; $P=0.51$ for the comparison between placebo and débridement) and 51.6±23.7, 53.7±23.7, and 51.4±23.2, respectively, at two years ($P=0.64$ and $P=0.96$, respectively). Furthermore, the 95 percent confidence intervals for the differences between the placebo group and the intervention groups exclude any clinically meaningful difference.

Conclusions In this controlled trial involving patients with osteoarthritis of the knee, the outcomes after arthroscopic lavage or arthroscopic débridement were no better than those after a placebo procedure.

(N Engl J Med 2002;347:81-8.)

Not all evidence is created equally

Identifying quality of evidence: JHN EBP Model

- Depends on study design and how the study was carried out

Evidence Levels	Quality Guides
Level I Experimental study, randomized controlled trial (RCT) Systematic review of RCTs, with or without meta-analysis	A **High quality:** Consistent, generalizable results; sufficient sample size for the study design; adequate control; definitive conclusions; consistent recommendations based on comprehensive literature review that includes thorough reference to scientific evidence B **Good quality:** Reasonably consistent results; sufficient sample size for the study design; some control, fairly definitive conclusions; reasonably consistent recommendations based on fairly comprehensive literature review that includes some reference to scientific evidence C **Low quality or major flaws:** Little evidence with inconsistent results; insufficient sample size for the study design; conclusions cannot be drawn
Level II Quasi-experimental study Systematic review of a combination of RCTs and quasi-experimental, or quasi-experimental studies only, with or without meta-analysis	
Level III Non-experimental study Systematic review of a combination of RCTs, quasi-experimental and non-experimental studies, or non-experimental studies only, with or without meta-analysis Qualitative study or systematic review with or without a meta-synthesis	

Identifying quality of evidence: JHN EBP Model

- Continued…

Evidence Levels	Quality Guides
Level IV Opinion of respected authorities and/or nationally recognized expert committees/consensus panels based on scientific evidence Includes: • Clinical practice guidelines • Consensus panels	A **High quality:** Material officially sponsored by a professional, public, private organization, or government agency; documentation of a systematic literature search strategy; consistent results with sufficient numbers of well-designed studies; criteria-based evaluation of overall scientific strength and quality of included studies and definitive conclusions; national expertise is clearly evident; developed or revised within the last 5 years B **Good quality:** Material officially sponsored by a professional, public, private organization, or government agency; reasonably thorough and appropriate systematic literature search strategy; reasonably consistent results, sufficient numbers of well-designed studies; evaluation of strengths and limitations of included studies with fairly definitive conclusions; national expertise is clearly evident; developed or revised within the last 5 years C **Low quality or major flaws:** Material not sponsored by an official organization or agency; undefined, poorly defined, or limited literature search strategy; no evaluation of strengths and limitations of included studies, insufficient evidence with inconsistent results, conclusions cannot be drawn; not revised within the last 5 years

Identifying quality of evidence: JHN EBP Model

- Continued…

Level V Based on experiential and non-research evidence Includes: • Literature reviews • Quality improvement, program or financial evaluation • Case reports • Opinion of nationally recognized experts(s) based on experiential evidence	**Organizational Experience:** A **High quality:** Clear aims and objectives; consistent results across multiple settings; formal quality improvement, financial or program evaluation methods used; definitive conclusions; consistent recommendations with thorough reference to scientific evidence B **Good quality:** Clear aims and objectives; consistent results in a single setting; formal quality improvement or financial or program evaluation methods used; reasonably consistent recommendations with some reference to scientific evidence C **Low quality or major flaws:** Unclear or missing aims and objectives; inconsistent results; poorly defined quality improvement, financial or program evaluation methods; recommendations cannot be made **Literature Review, Expert Opinion, Case Report, Community Standard, Clinician Experience, Consumer Preference:** A **High quality:** Expertise is clearly evident; draws definitive conclusions; provides scientific rationale; thought leader(s) in the field B **Good quality:** Expertise appears to be credible; draws fairly definitive conclusions; provides logical argument for opinions C **Low quality or major flaws:** Expertise is not discernable or is dubious; conclusions cannot be drawn

Other sources of evidence

- Clinical Practice Guidelines
- Pre-Processed Information