PERMANENTE MEDICINE®
The Permanente Medical Group

Crafting Data-Driven Investigations in Clinical Learning Environments

Presenter:

Juleon Rabbani, DrPH
Director, Biostatistical Consulting Unit (BSCU)
Kaiser Permanente Northern California Division of Research

KAISER PERMANENTE®

Learning Objectives

After this session, learners will be able to:

- Recognize different types of data commonly collected in health care settings
- Describe epidemiologic reasoning and how it is related to desired level of evidence
- Evaluate feasibility and the resources needed for different types of investigations

Biostatistical Consulting Unit (BSCU)

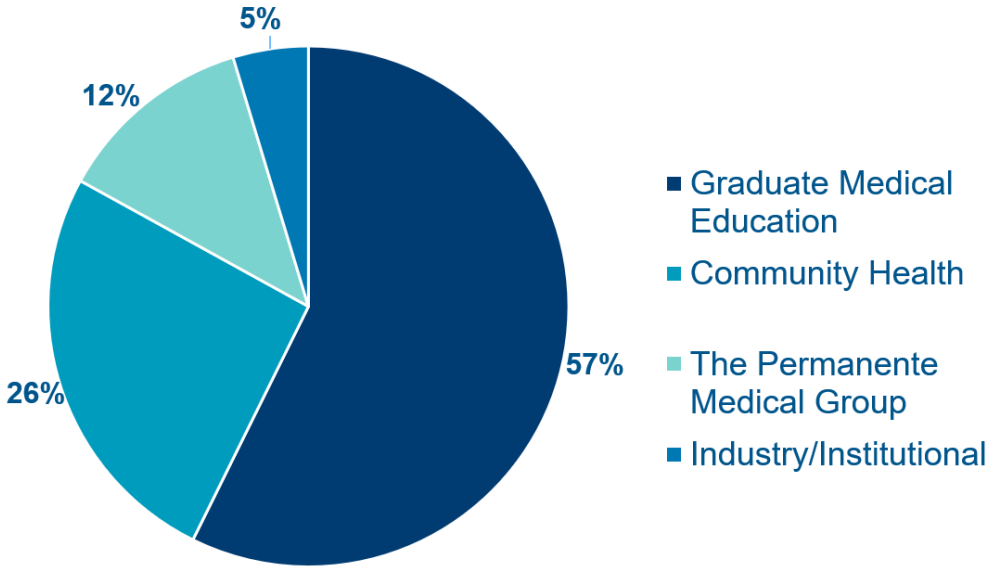
- Founded in 2004 in response to growing:
  - Research interests in our medical centers
  - Programming and biostatistical needs
- 32 staff (23 programmer analysts)

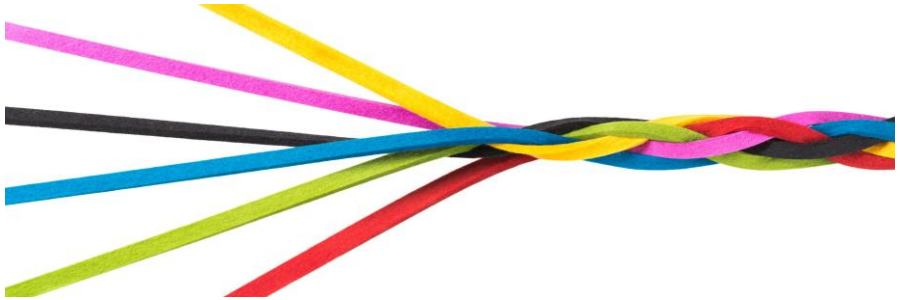
BSCU consultation services at-a-glance:

- Study design
- Statistical methods
- Feasibility data
- Sample size and power calculations
- IRB application development and refinement
- Data collection and management
- Data analyses
- Assistance with preparation of abstracts, presentations, manuscripts (co-authors)

2024 BSCU Research Sponsors

- Kaiser Foundation Hospitals (KFH)
  - Graduate Medical Education (GME)
  - Community Health (CH)
- The Permanente Medical Group (TPMG)
- Industry and Institutional Funding
  - Bayer AG, Physician Researcher Program, other foundation grants





Medical Education Physician Research Leads

Cynthia Campbell, PhD, MPH
Addiction Medicine
Greater Southern Alameda Area

Edward Yap, MD
Anesthesia
UCSF

Gary Grinberg, MD
Bariatric Surgery
Sacramento

Joan Lo, MD
Internal Medicine
Oakland

Choon Hwa (Anne) Goh, MD, MPH
Internal Medicine
San Francisco

Seema Pursnani, MD, MPH
Internal Medicine
Santa Clara

Melinda Bowlby, DPM
Podiatric Surgery
Santa Clara

Matthew Hirschtritt, MD, MPH
Psychiatry
Oakland

Samuel Ridout, MD, PhD
Psychiatry
San Jose

Andrew Ambrosy, MD
Cardiovascular Disease
San Francisco

Cynthia Mendez-Kohlieber, MD
Community Medicine & Public Health
Sacramento

Edward Durant, MD, MPH
Emergency Medicine
Central Valley

Ali Poyan Mehr, MD
Nephrology
San Francisco

Eve Zaritsky, MD
Obstetrics-Gynecology
Oakland

Maryl Sackeim, MD
Obstetrics-Gynecology
San Francisco

Nareg Roubinian, MD, MPH
Pulmonary & Critical Care Medicine
Oakland

Caitlin Painter, DO
Urogynecology & Reconstructive Pelvic Surgery
Oakland/UCSF

Robert Chang, MD
Vascular Surgery
UCSF

Vira Fomenko, MD
Family Medicine
San Jose

Julia Shaver, MD
Family Medicine
Santa Rosa

Kirsten Winnie, MD
Family Medicine
Santa Rosa

Nikhil Joshi , MD
Obstetrics-Gynecology
Santa Clara

Ehsan Tabaraee , MD
Orthopedic Spine Surgery
Oakland

Jonathan Liang, MD, MPH
Otolaryngology Head & Neck
Oakland

Tracy Lieu, MD, MPH
Program Advisor
Regional

Jeffrey Lee, MD, MPH
Gastroenterology
San Francisco

Leslie Wong, MD
General Surgery
Oakland/UCSF

Lisa Yang, MD
General Surgery
Oakland/UCSF

Nardine Riegels, MD
Patient Safety
Oakland

Naomi Adler, MD
Patient Safety
Oakland

Kristie Manning, MD
Pediatric Hospital Medicine
Oakland

Christine Garcia, MD
Gynecologic Oncology
San Francisco

Raymond Liu, MD
Hematology Oncology
San Francisco

Jacek Skarbinski, MD
HIV Medicine
Oakland

Peter Cooch, MD
Pediatrics
Oakland

Daniel Lee, DPM
Podiatric Surgery
Sacramento

Jason Pollard, DPM
Podiatric Surgery
Oakland

PERMANENTE MEDICINE®
The Permanente Medical Group

Oakland

Miranda Ritterman Weintraub, PhD, MPH

Nirmala Ramalingam, MPP

San Francisco

Aida Shirazi, R.Ph, MS., Ph.D.

Santa Clara

Cynthia Triplett, MPH, MA

Napa Solano

Kat Dang, MS, MAS

GME Research Project Managers (RPMs)

San Jose

Brooke Harris, PhD

Santa Rosa

Shannon McDermott, PhD

Medical center-based researchers who support local GME research and quality improvement initiatives

Sacramento

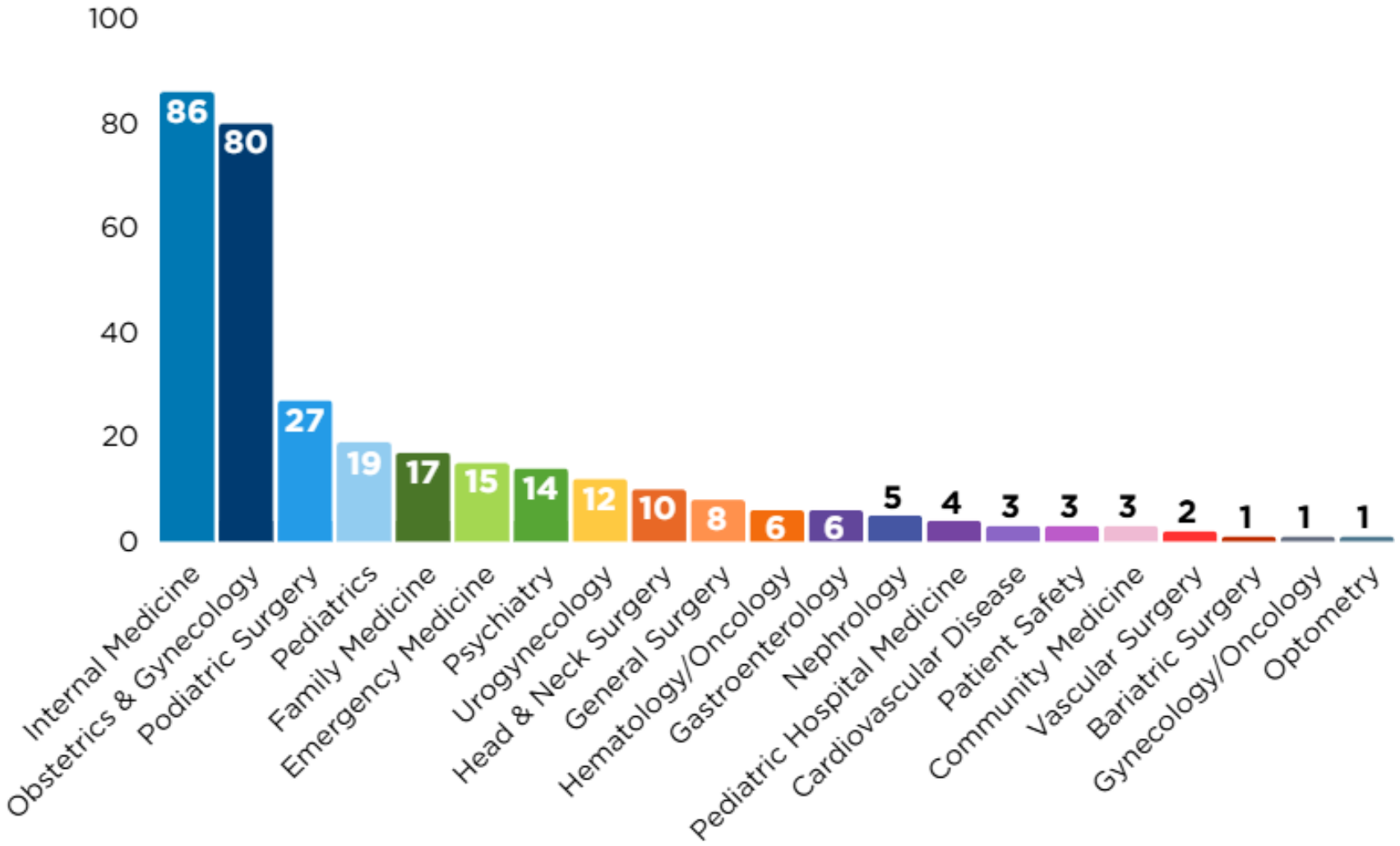
Stacie Lloyd, PhD

Modesto

Lorlelei Lee-Haynes, MPH

PERMANENTE MEDICINE®
The Permanente Medical Group

Recent* Medical Education-led Studies (N=323)

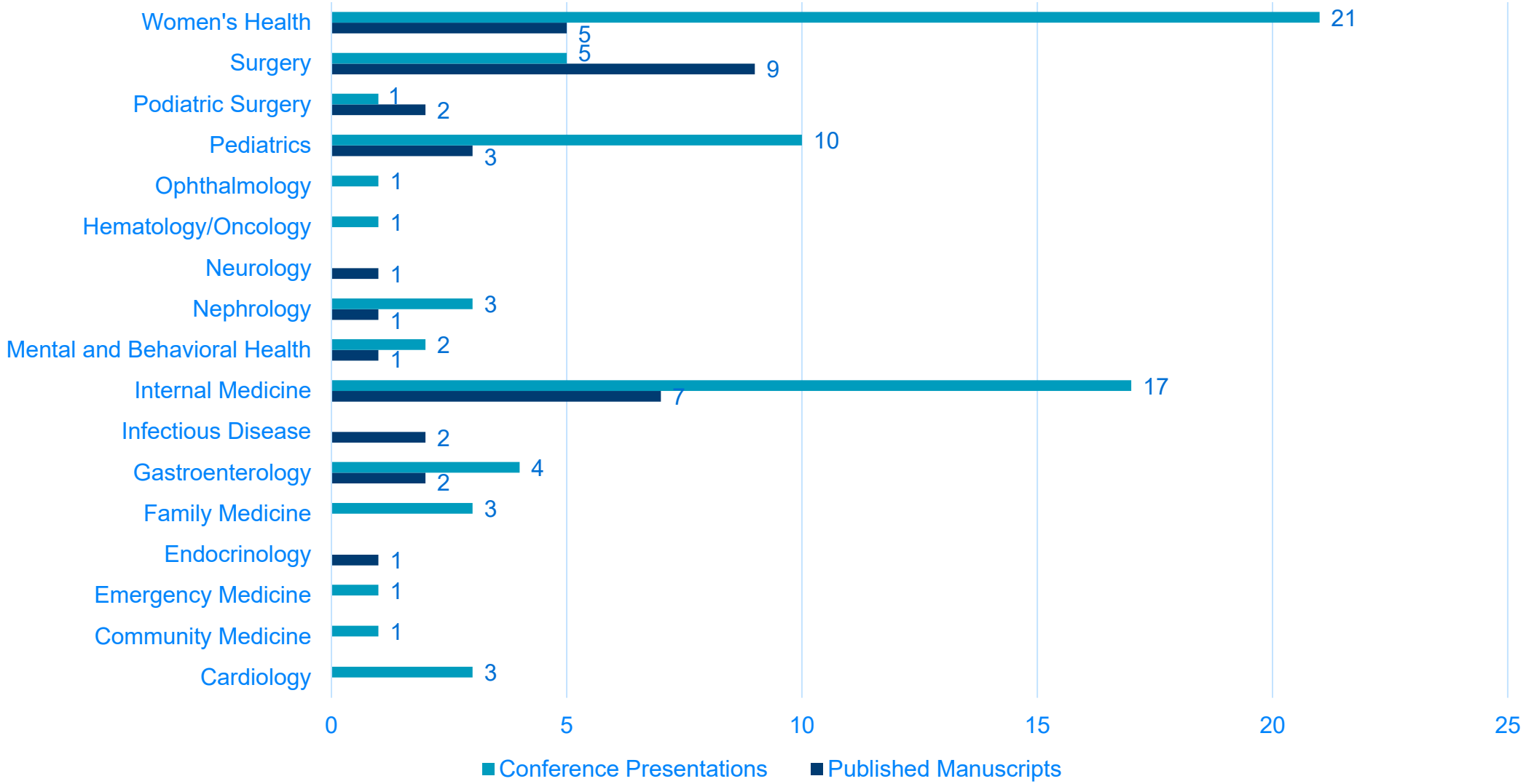


*Jan 2024 - Oct 10, 2025

Scholarly Outcomes by Clinical Specialty

In 2024, the BSCU supported **107 scholarly outcomes**:

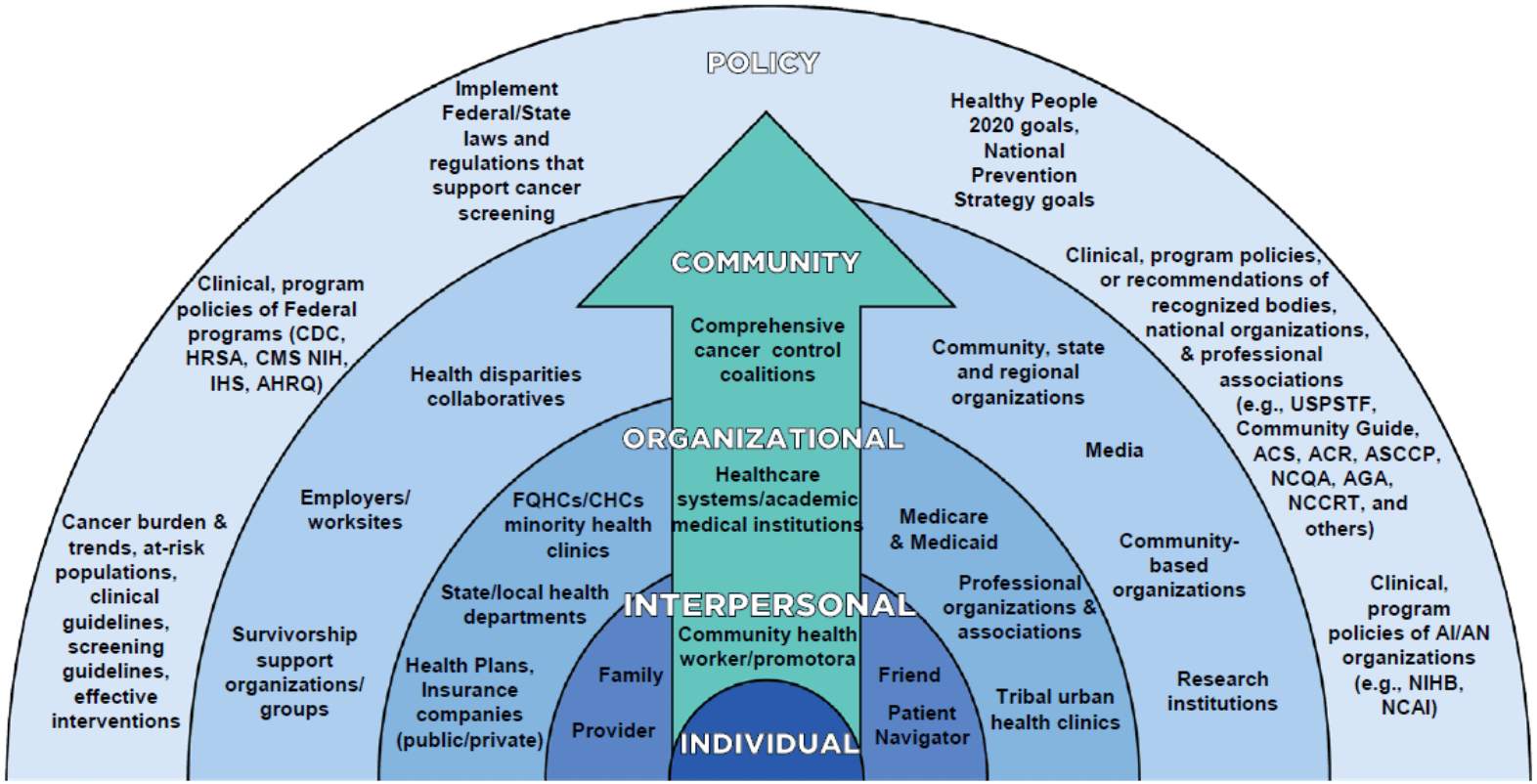
- **34 Published Manuscripts**
- **73 Conference Presentations**



PERMANENTE MEDICINE®
The Permanente Medical Group

Understanding Data

Levels of Data (socio-ecological)



*Some groups may fit within multiple levels of this model.

The CDC Colorectal Cancer Control Program's (CRCCP's) multi-level approach to colorectal cancer prevention

- Outcomes are influenced by factors at **multiple levels**, from genetic to societal
- There are important **interactions** between and within levels
- Change at higher levels **(society)** can cause change at lower levels **(behavior)** and vice versa **(reciprocal determinism)**

Choosing a Level and Scope of Intervention & Investigation

- Suggest limiting to 2 levels:
  - One level: Individual (resident's attitudes) & Individual (resident's clinical competency)
  - Two Levels: Individual (resident's attitudes) & Interpersonal (attending's behavior)
- It takes an entire body of knowledge to address all levels
- Critics can be unrealistic
  - If you don't solve economic inequality, your intervention is trivial
- Thinking at multiple levels will broaden your intervention & investigation options

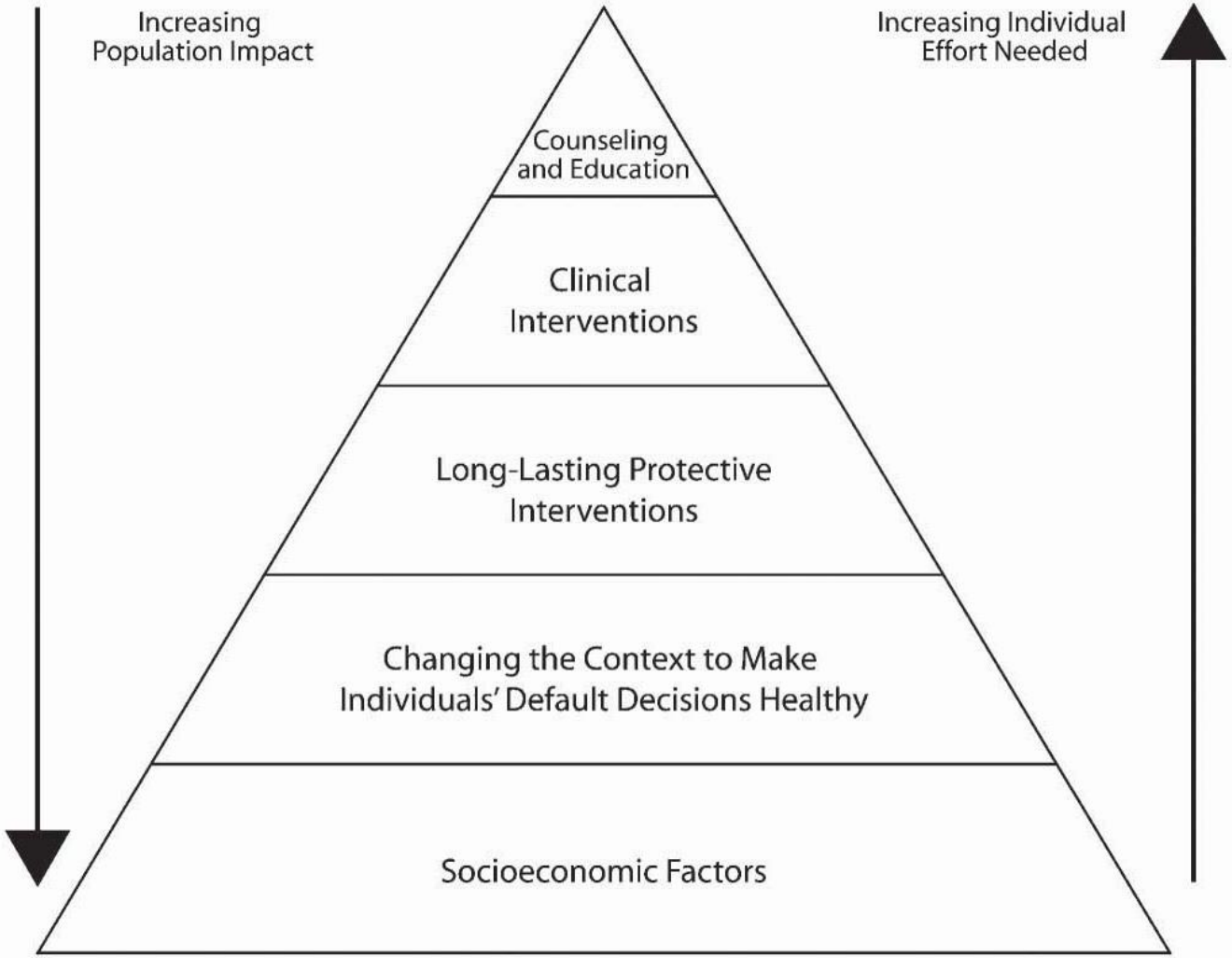
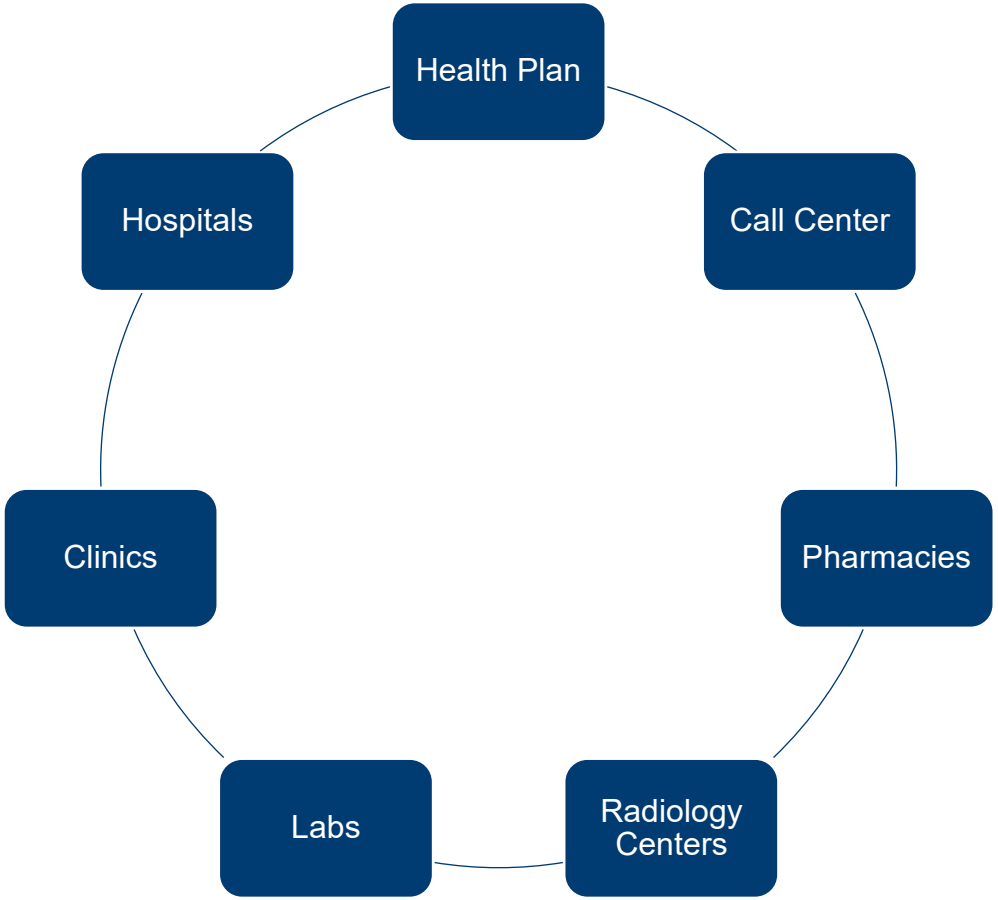


FIGURE 1—The health impact pyramid.

Frieden, 2010

Types of Health Care Data

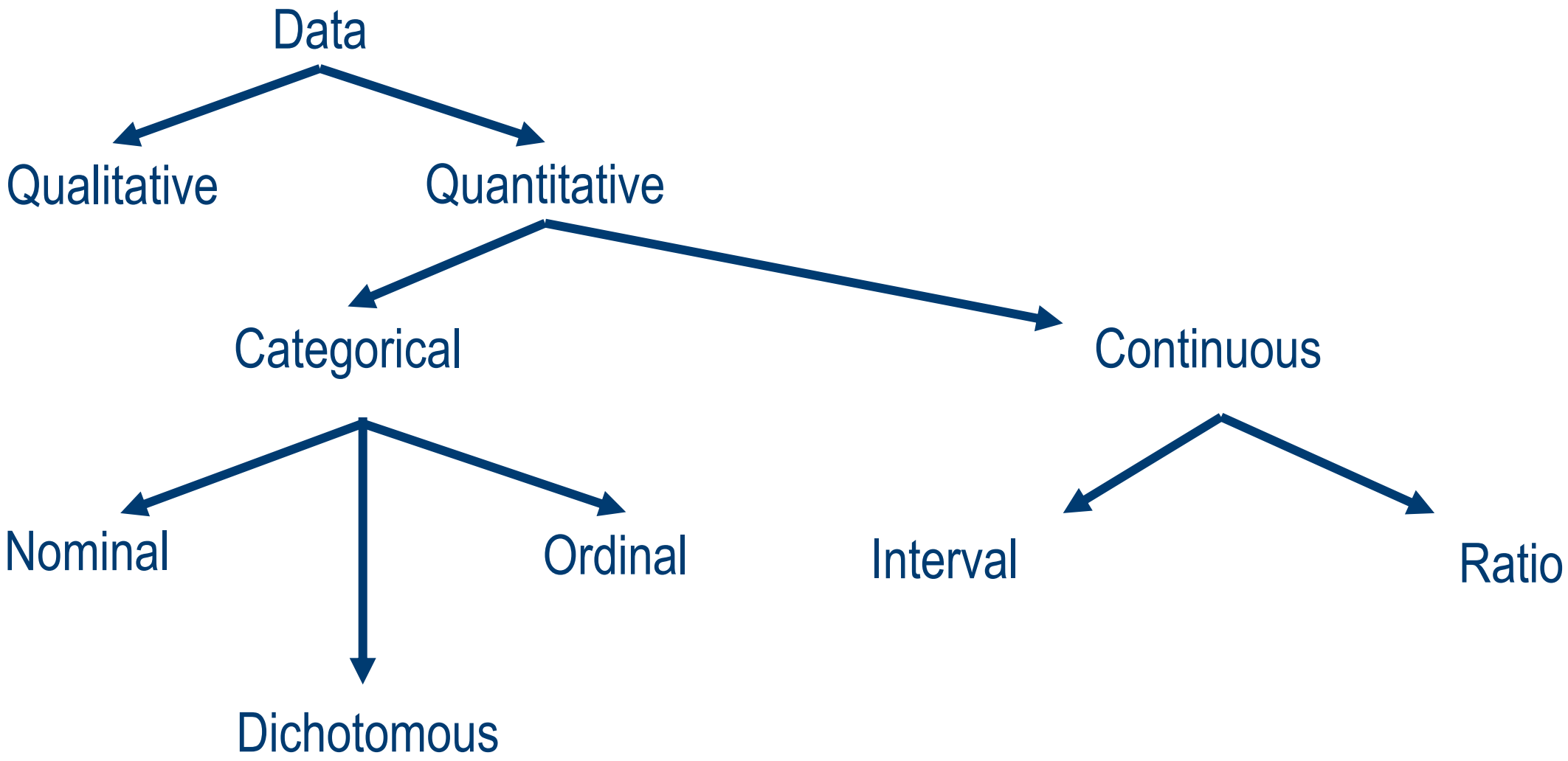
- Patient-level (more common in research vs QI)
  - Medical records (diagnoses, procedures, labs, pharmacy, radiology)
  - Patient demographics
  - Patient satisfaction
  - Utilization (visits, calls)
  - Costs (copays, premiums)
  - Health insurance (type of insurance, public vs commercial)
- Provider-level (residents, fellows, faculty)
  - Administrative (human resource data on employees)
  - Behaviors
  - Facilitators/barriers associated with behaviors (psychosocial, environmental, structural, legal)
  - Clinical learning environment:
    - Training activities
    - Clinical competency



Why Understanding Data Matters

The type of data you have, along with how it is distributed and varied, will dictate how you analyze and interpret your data

Types of Data



Qualitative Data

- Words (scripts, interviews)
- Pictures
- Videos

Qualitative research asks:

- *"What is X?"*
- *"How does X vary in different circumstances?"*
- And *"Why?"*

Rather than *"How many X's are there?"*

Qualitative Methods

Common designs:

- Focus groups
- Surveys
- Interviews
- Nominal group process
- Delphi Technique
- Community forum
- Participant observation
- Photovoice
- Video ethnography

Gathers ideas and themes about:

- Knowledge/Beliefs
- Attitudes
- Practices
- Priorities
- Available resources
- Program preferences
- Distribution outlets
- Choice of language

Quantitative Data

- Data that can be quantified:
  - **Discrete** = counts (whole numbers).
  - **Categorical** = labels you can sort into groups.
  - **Continuous** = measurements (can include decimals).
    - **Interval** = equal spacing, no true zero.
    - **Ratio** = equal spacing, with a true zero.

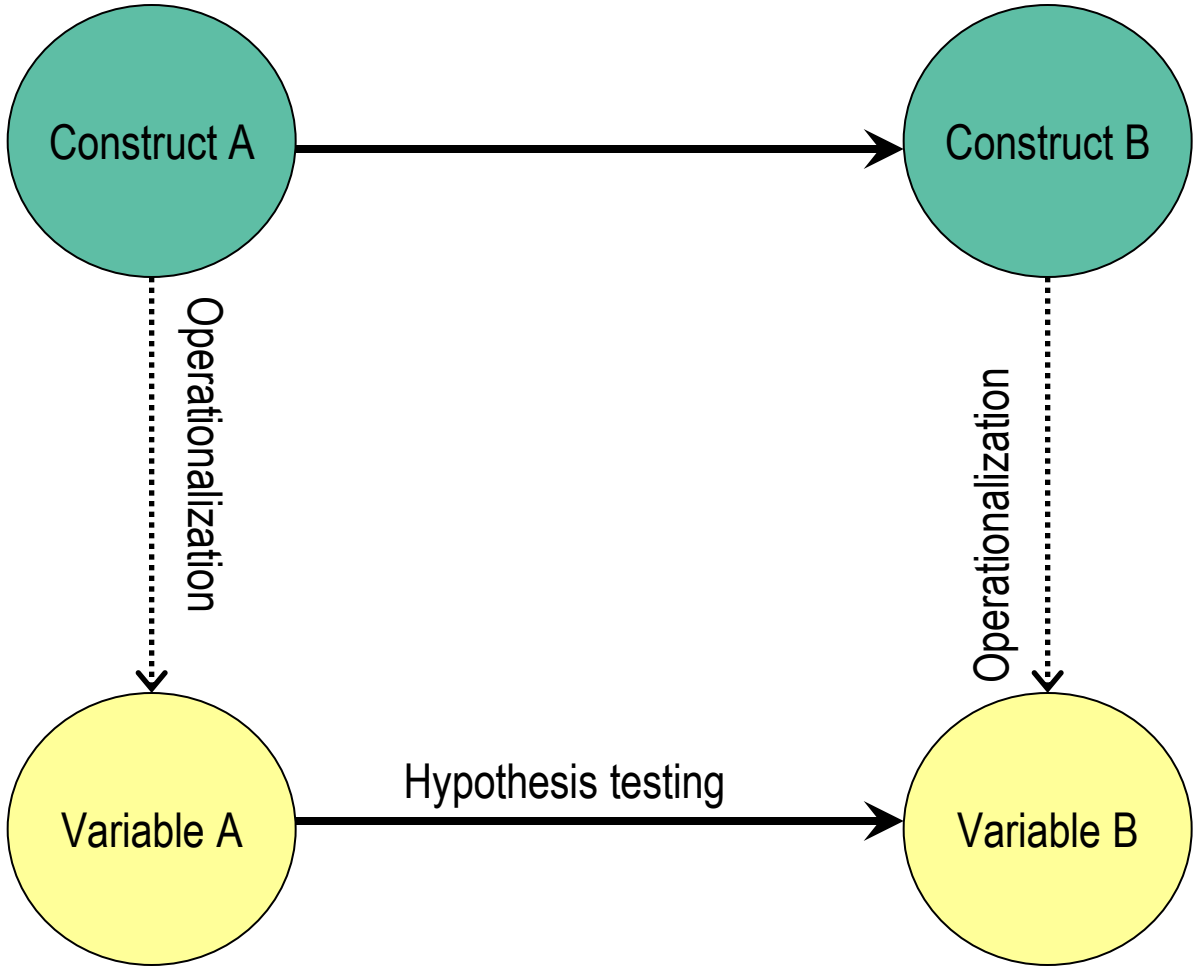
Quantitative research asks:

- **Descriptive Questions** *(What is happening with X?)*
- **Comparative Questions** *(Is there a difference between X and Y?)*
- **Relationship Questions** *(Is there a connection between X and Y?)*
- **Causal/Experimental Questions** *(Does X cause Y?)*

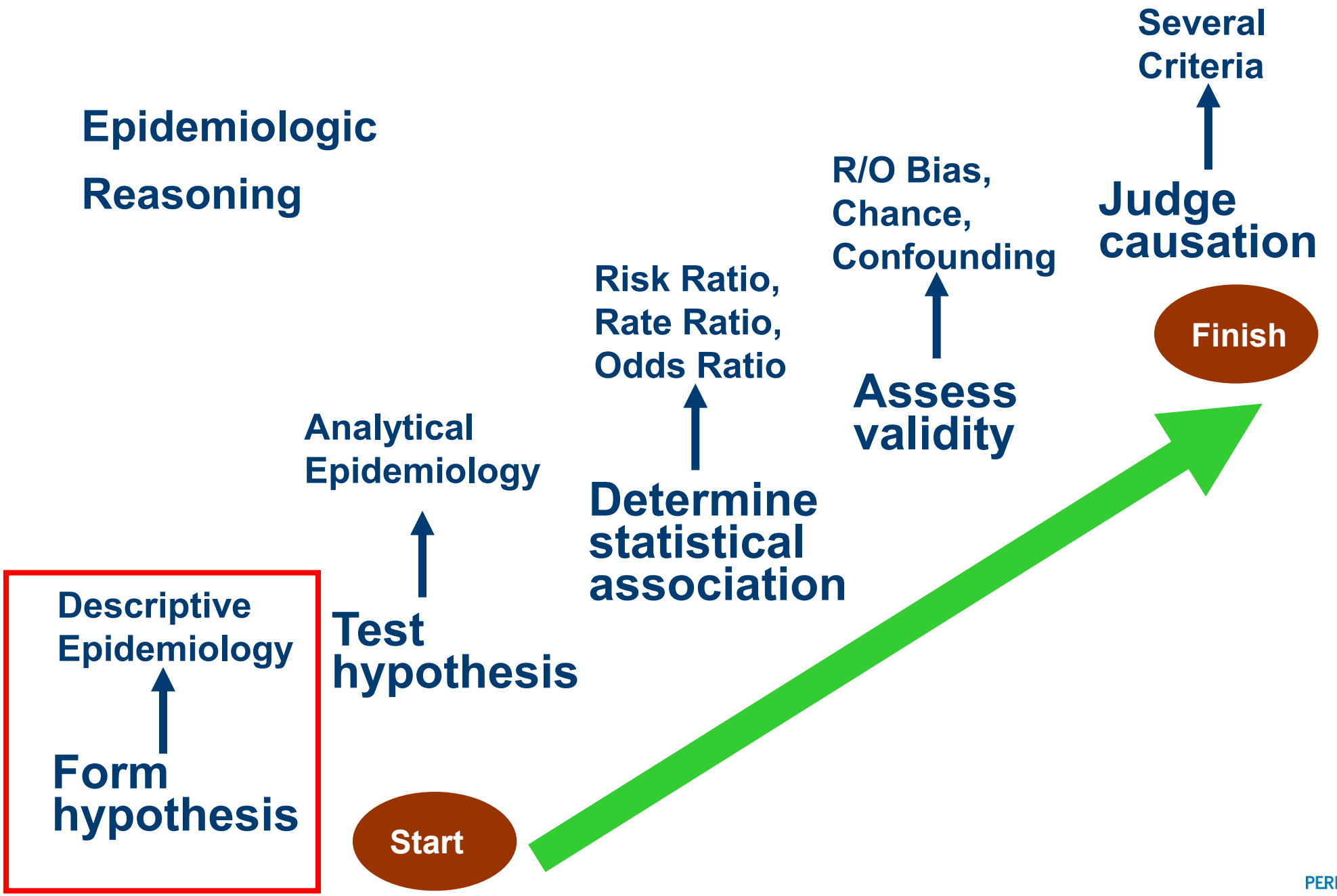
Rather than *"What is X?"*

Clinical Learning Environment Example

- Broad Question: How are knowledge and wellbeing related to clinical competency among residents?
- Operationalize factors and outcomes of interest
- Factors affecting clinical competency:
  - Knowledge → # professional development sessions attended
  - Well-being → mental health score (PHQ9)
- Outcome of interest
  - Clinical Competency -> medical knowledge score cutoffs (low vs high score)
- Operationalized question: How are number of professional development sessions attended and PHQ9 scores related to resident clinical competency in medical knowledge?



Understanding Level of Evidence

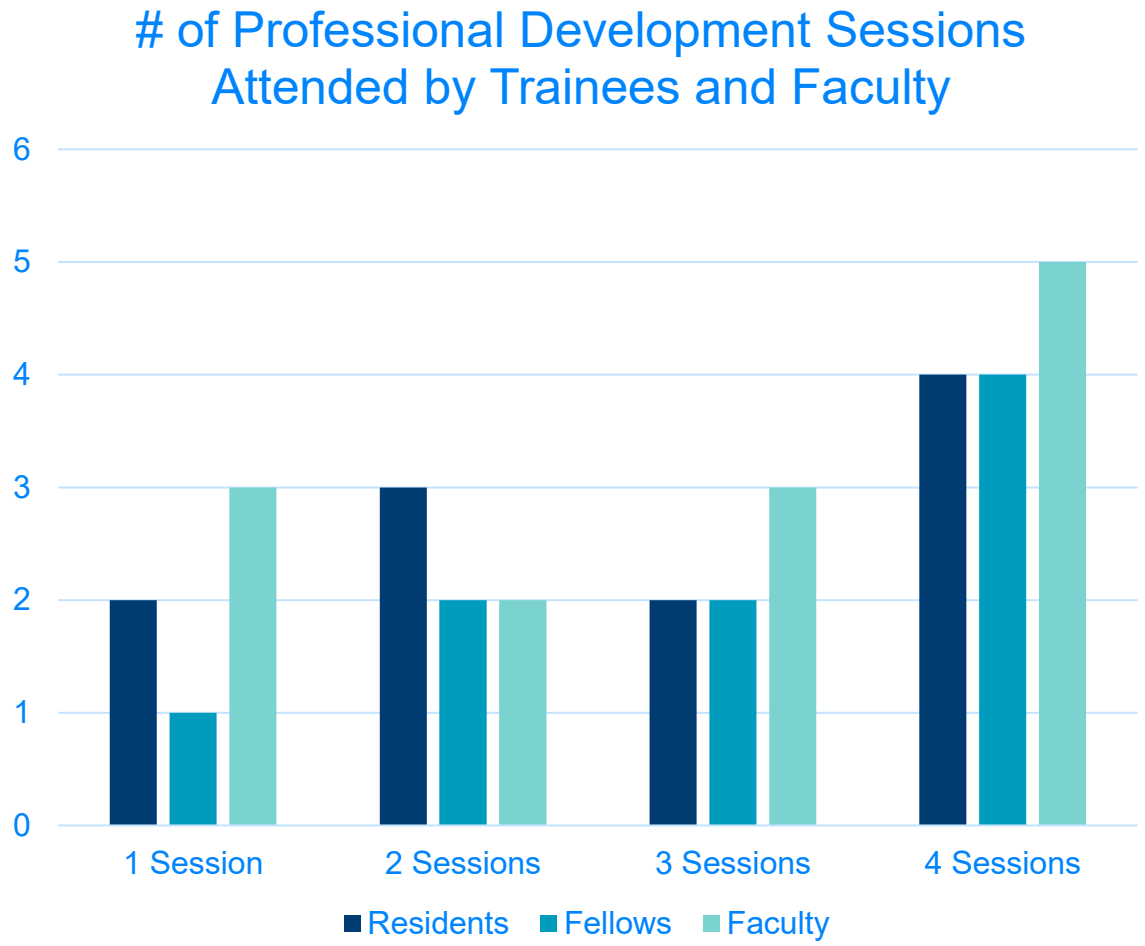


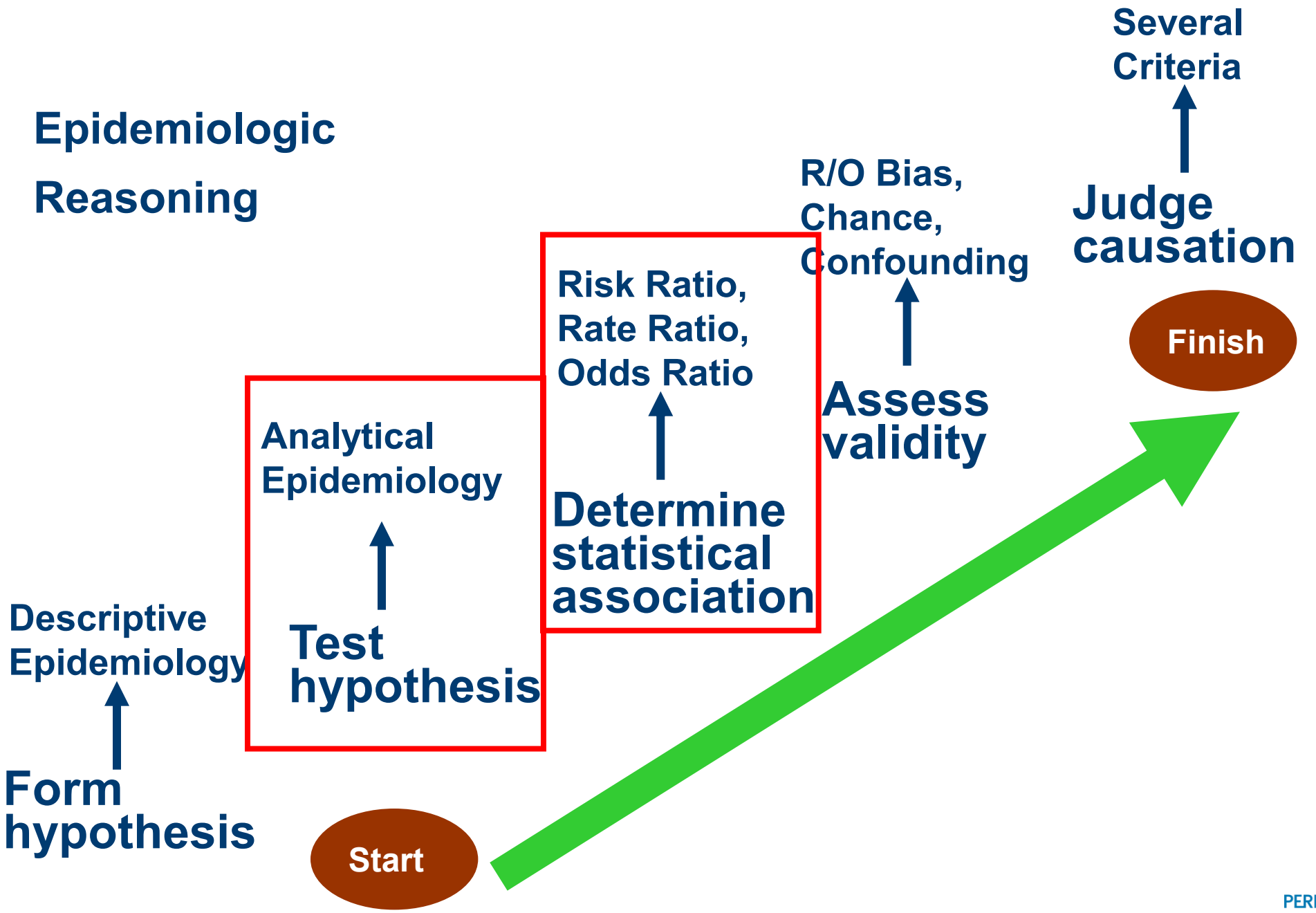
Descriptive Statistics (Univariate Analysis)

- Distribution (normal, skewed)
- Central Tendency (mean, median)
- Variability (standard deviation, variance)

Clinical Learning Environment Example (descriptive)

- Question: How many trainees and faculty have attended at least four professional development sessions on medical knowledge?
  - # of trainees/faculty who attended sessions
  - Distribution of session attendance among trainees and faculty

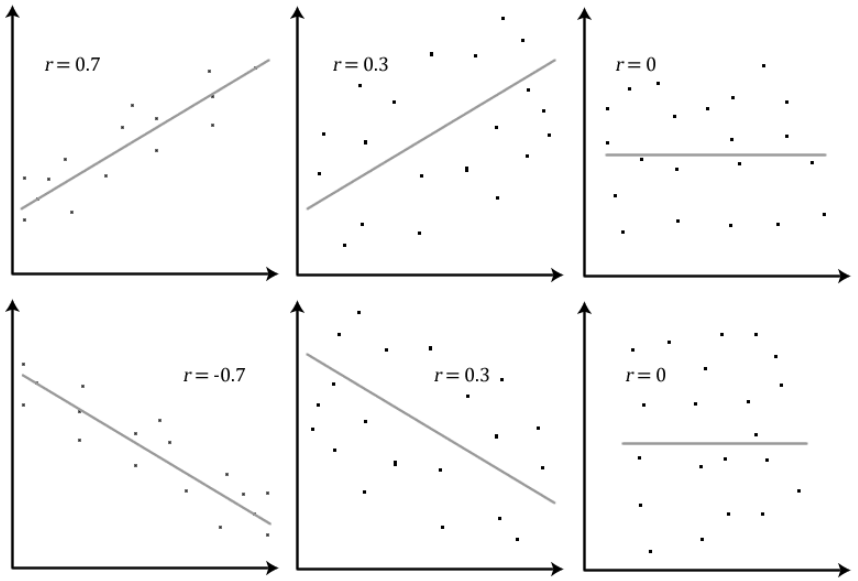




Bivariate Analysis

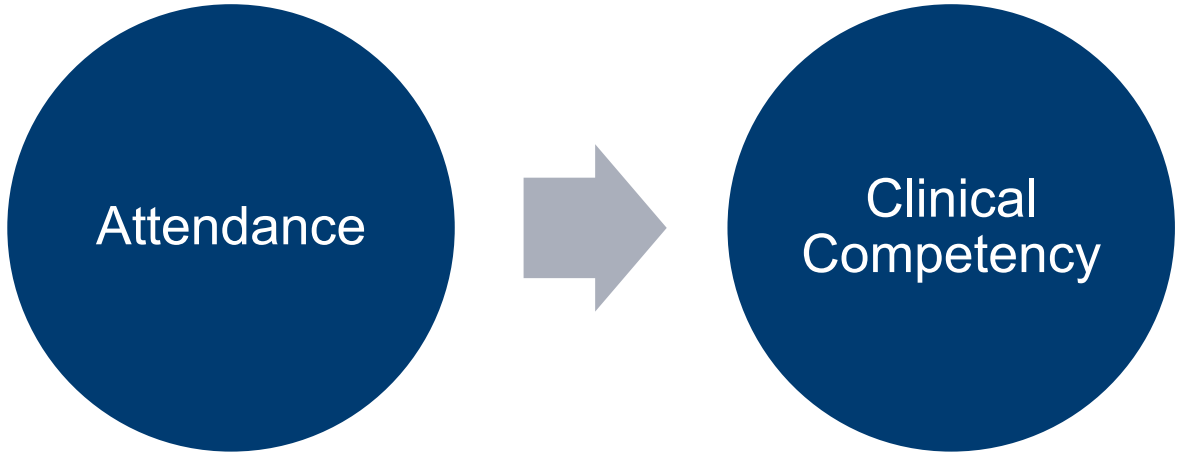
- The simultaneous analysis of two variables
- Explores relationship
- Determines:
  - Whether an association exists
  - Strength of this association
  - If there are differences between the variables
  - The significance of these differences
- Chi-square, t-test, correlation tests

	DISEASE		
RISK FACTOR		+	-
	+	a	b
	-	c	d

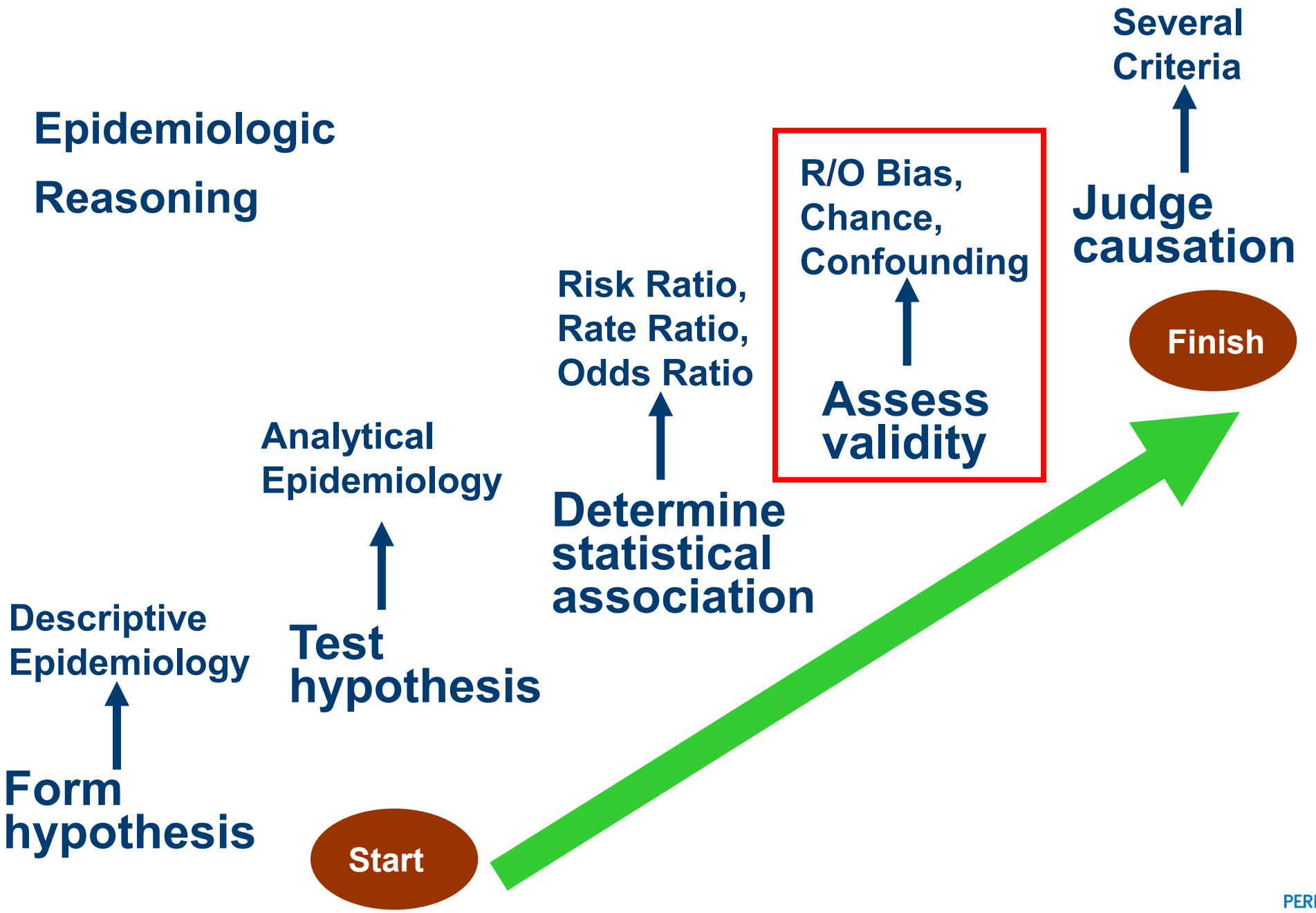


Clinical Learning Environment Example (bivariate)

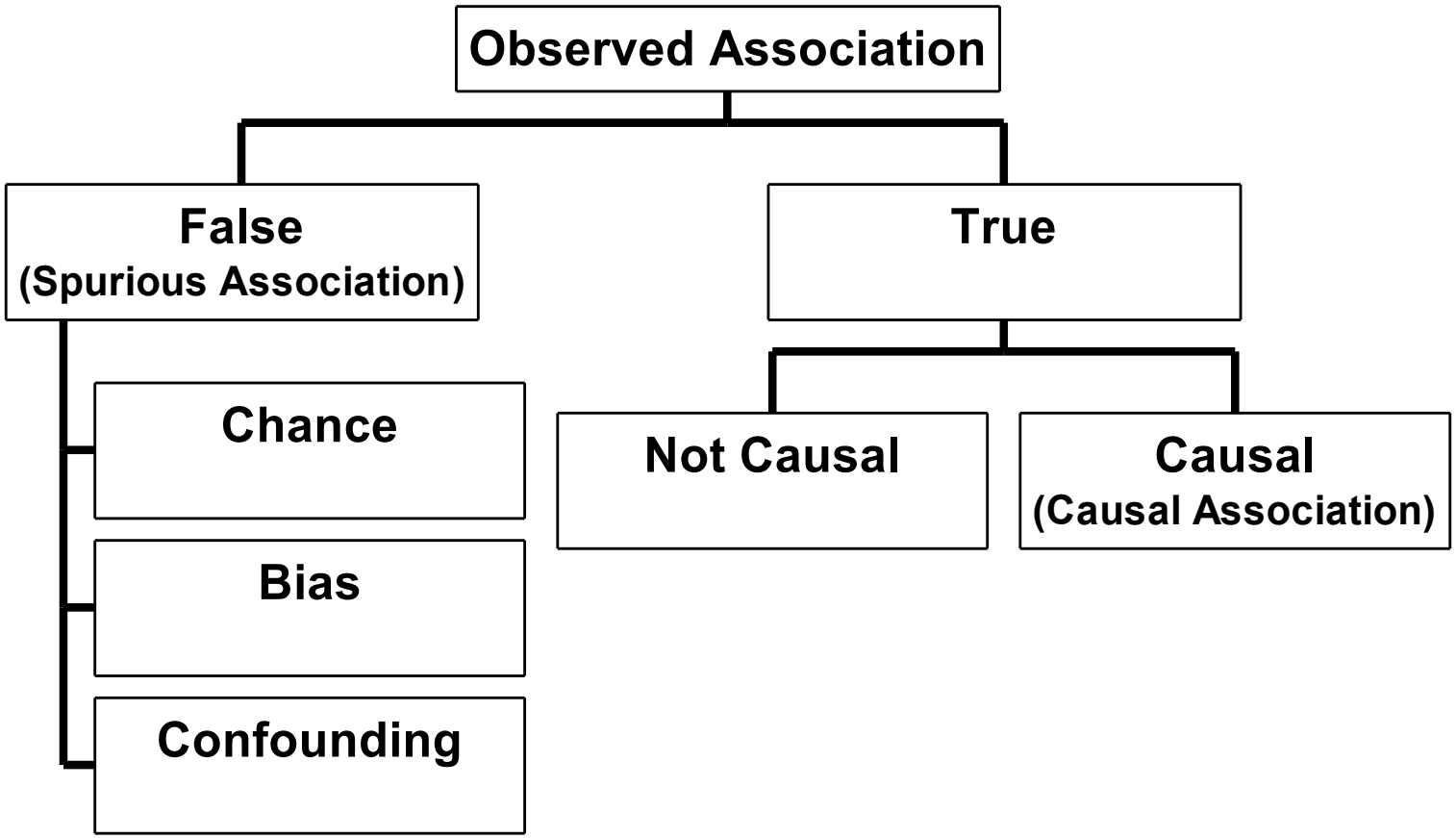
- Question: Is there *a relationship* between residents who have attended at least three professional development sessions and having a high clinical competency score?
  - Two variables:
    - Attended at least 3 session (y/n)
    - High medical knowledge competency score (y/n)
  - Analysis: Chi-square (comparing two categorical variables) or Fischer exact test (if small cell counts); simple logistic regression



Example interpretation: We found that residents who attended at least 3 professional development sessions had 2.5 greater odds of having high medical knowledge competency.

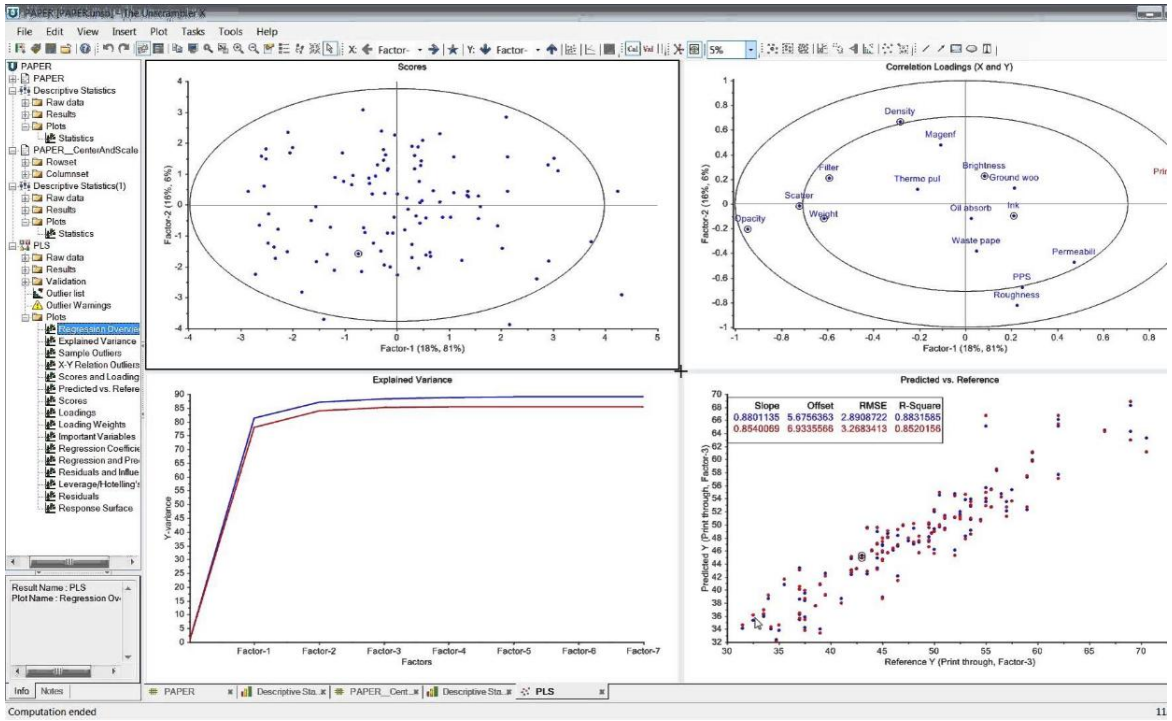


Associations and Threats to Validity



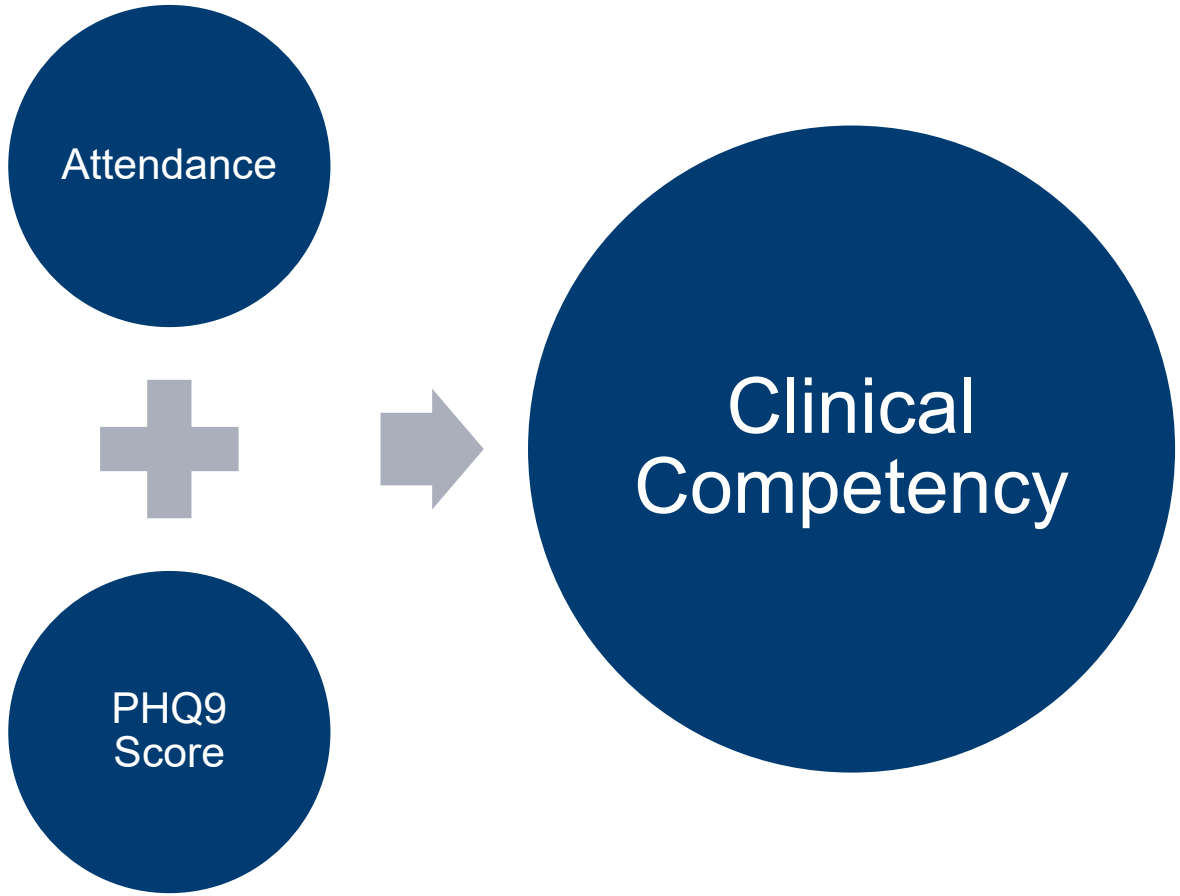
Multivariable Analysis

- The simultaneous analysis of several variables
- Creates predictive models that can help us understand causality
- Used to control for confounding factors and uncover interactions
- ANOVA, multiple linear/logistic/Cox Hazards models

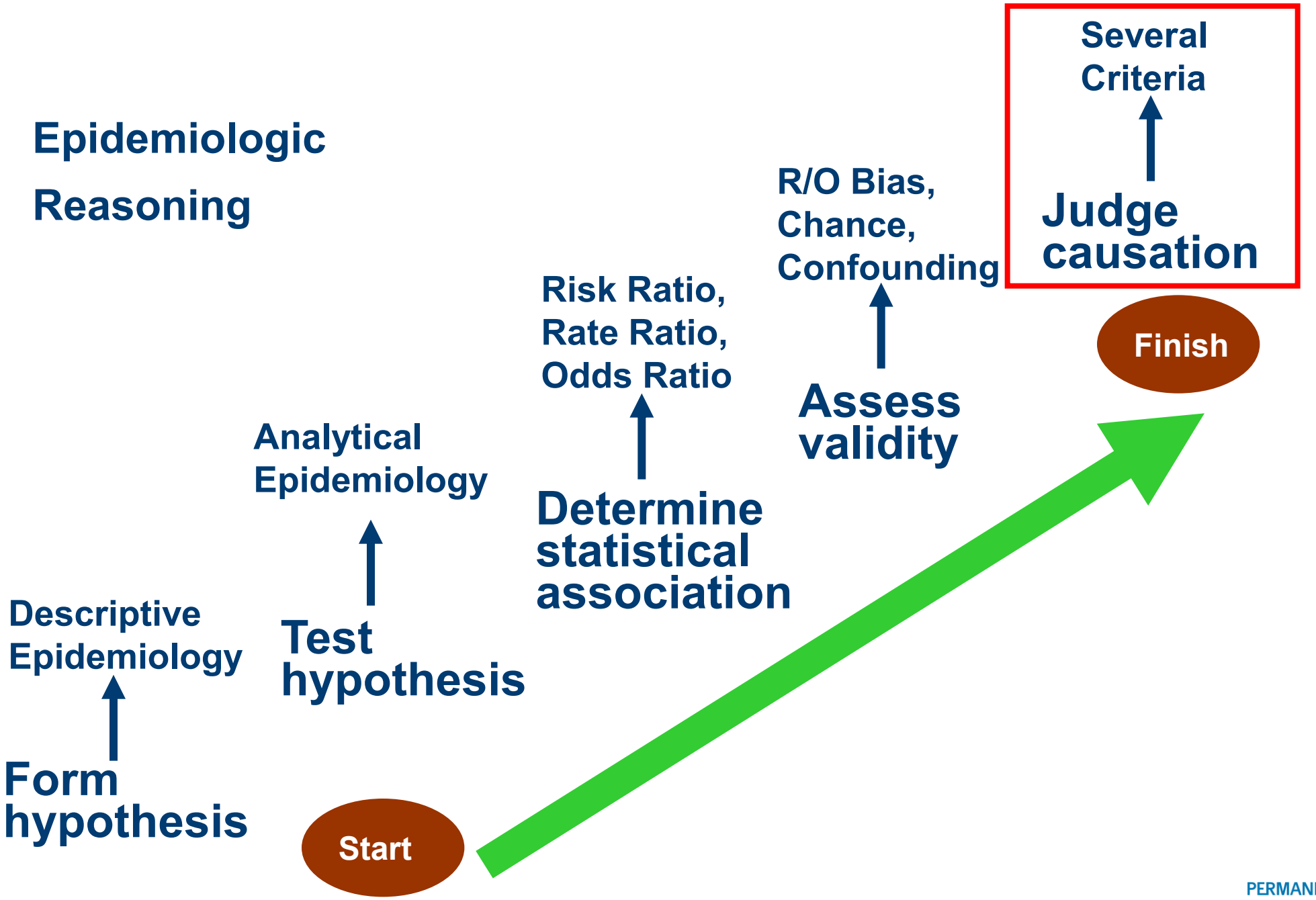


Clinical Learning Environment Example (multivariable)

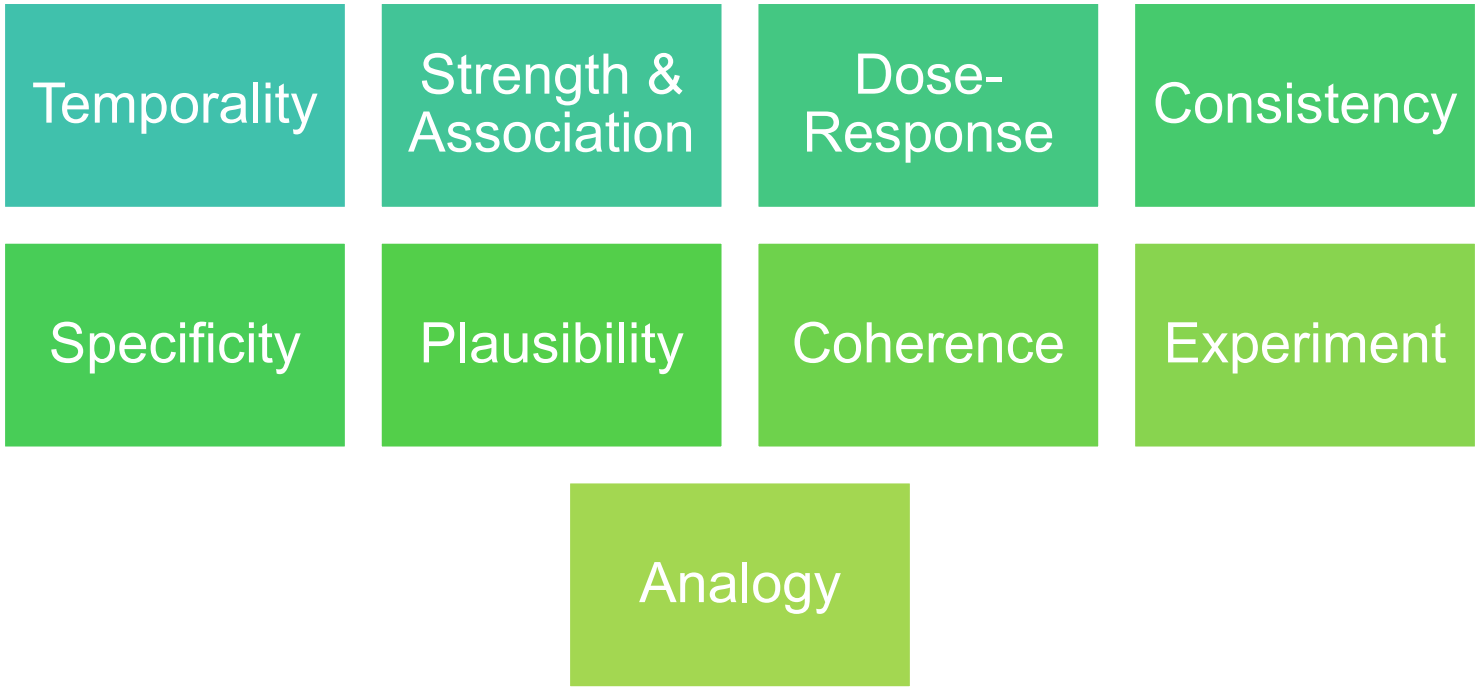
- Question: Is there *a relationship* between residents who have attended at least three professional development sessions and having a high clinical competency score, *after adjusting for resident well-being*?
  - Three variables:
    - Attended at least 3 session (y/n) – primary predictor
    - PHQ9 score (low/high) - covariate
    - High medical knowledge competency score (y/n)
  - Analysis: Multivariable logistic regression

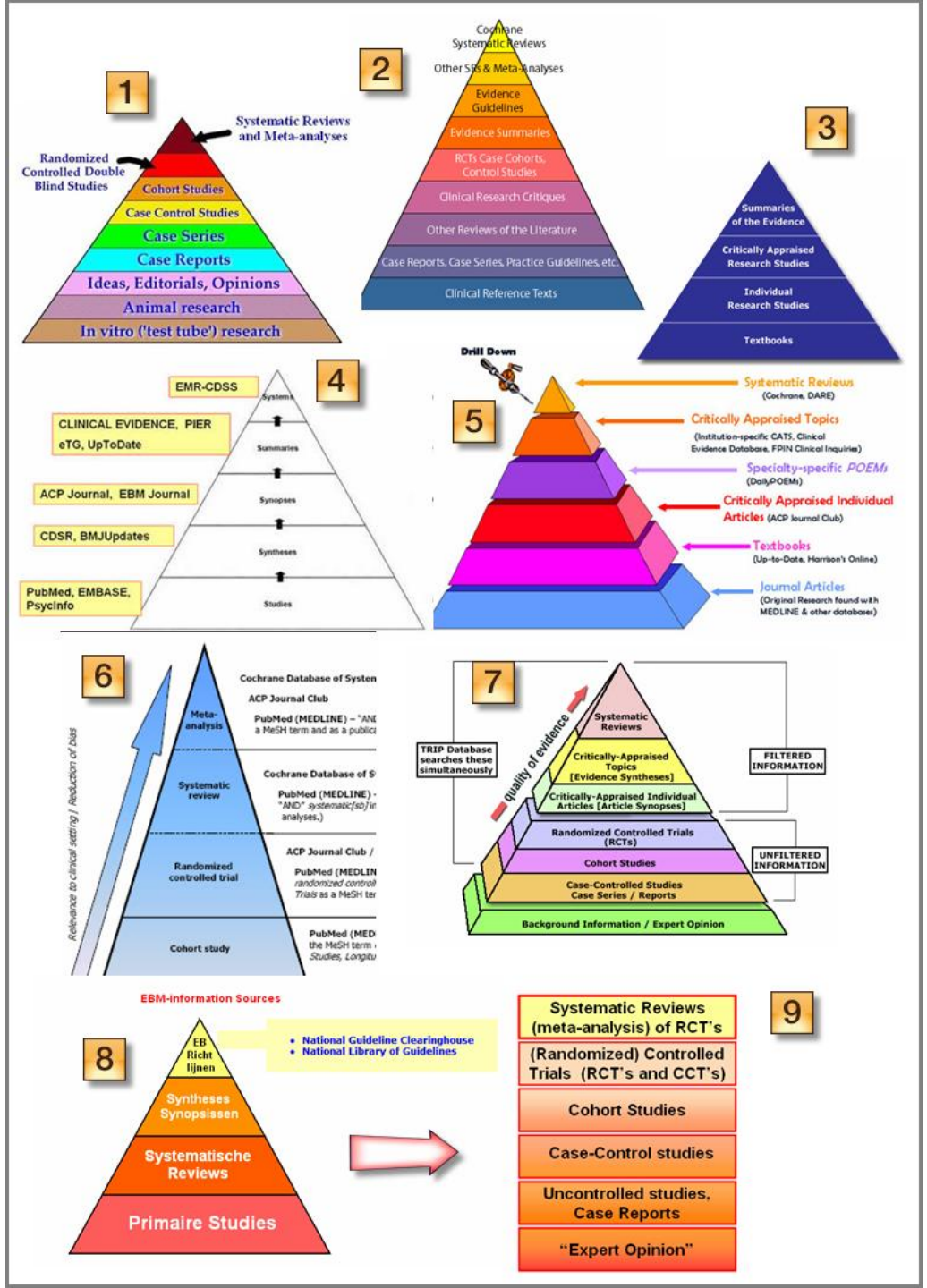


Example interpretation: We found that residents who attended at least 3 professional development sessions had 1.5 greater odds of having high medical knowledge competency, *after adjusting for mental health score.*


Epidemiologic
Reasoning
Several
Criteria
Judge
causation
R/O Bias,
Chance,
Confounding
Assess
validity
Risk Ratio,
Rate Ratio,
Odds Ratio
Determine
statistical
association
Analytical
Epidemiology
Test
hypothesis
Descriptive
Epidemiology
Form
hypothesis
Start
Finish

Hill's Criteria for Causality (1965)





Evidence Pyramids:

hierarchy of the quality/utility of evidence from top to bottom of each pyramid

http://laikaspoetnik.wordpress.com/2008/09/

Choosing Appropriate Statistical Tests

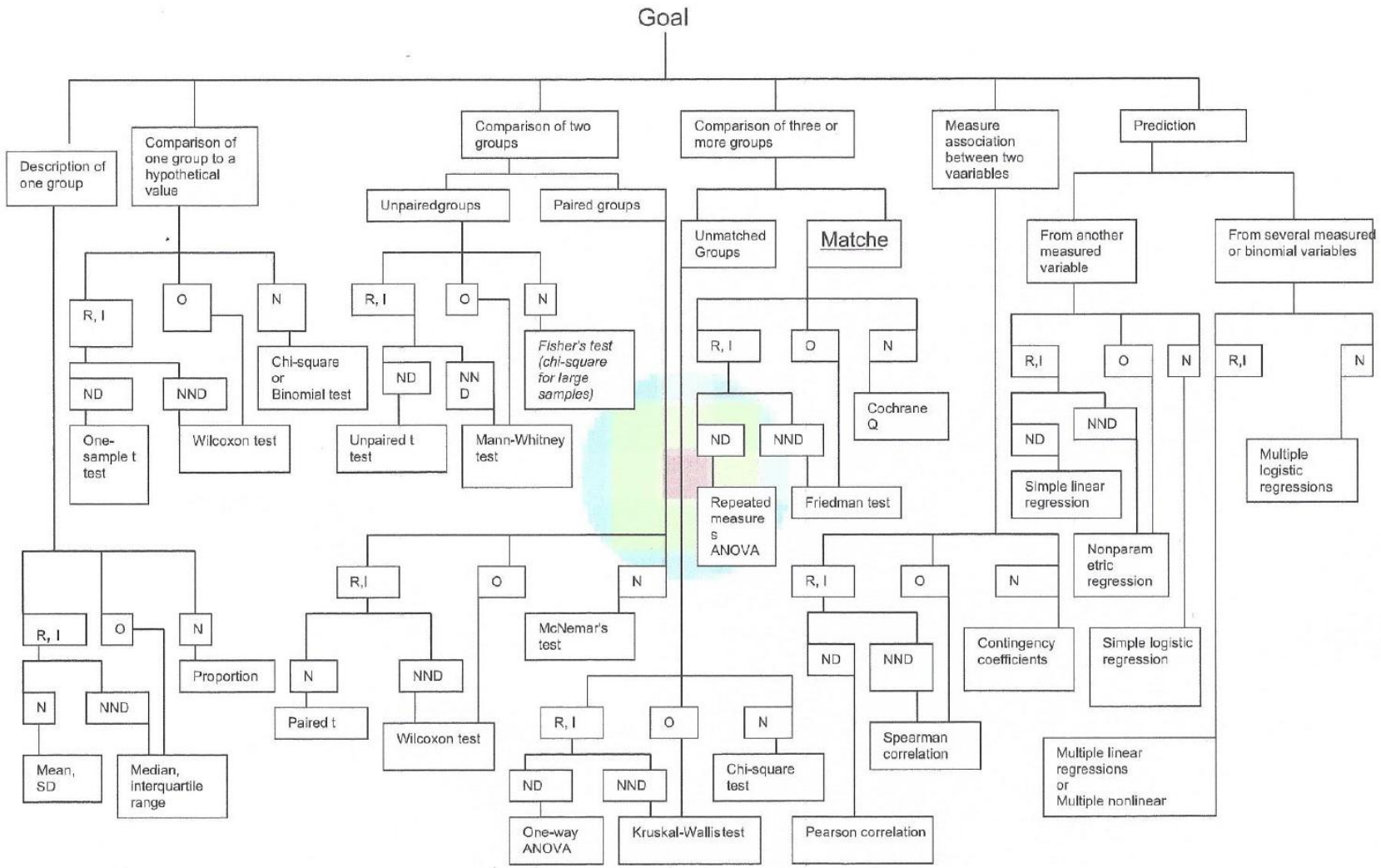


Figure 1: Flow chart for selection of appropriate statistical tests

Karan J. How to select appropriate statistical test.? *J Pharm Negative Results.* 2010;1:61–3. [Google Scholar]

Resources for conducting analysis in Excel (step by step)

- Descriptive analysis: https://www.excel-easy.com/examples/descriptive-statistics.html
- Bivariate analysis:
  - 2 categorical variables (chi-square): https://researchguides.library.vanderbilt.edu/c.php?g=156859&p=1171766
  - 2 continuous variables (correlation test): https://www.excel-easy.com/examples/correlation.html
  - 1 categorical and 1 continuous (t-test): https://statisticsbyjim.com/hypothesis-testing/t-tests-excel/
- Multivariable analysis: suggest working with someone with at least masters-level statistics/epidemiology background

Feasibility & Resources

Support

- **Funding/Institutional Support**
  - Protected time for scholarly activity
  - Research staff funding
  - Conference fees and travel
  - Paper fees and medical editor
- **Clinicians**
  - Subject-matter knowledge
  - Can chart review, if needed
- **Methods experts**
  - Research (investigator) vs quality improvement (improvement advisor)
  - Quantitative (biostatistician) vs Qualitative (focus group researcher)
  - Data extraction (programmer)
  - Regulatory oversight (compliance/IRB)

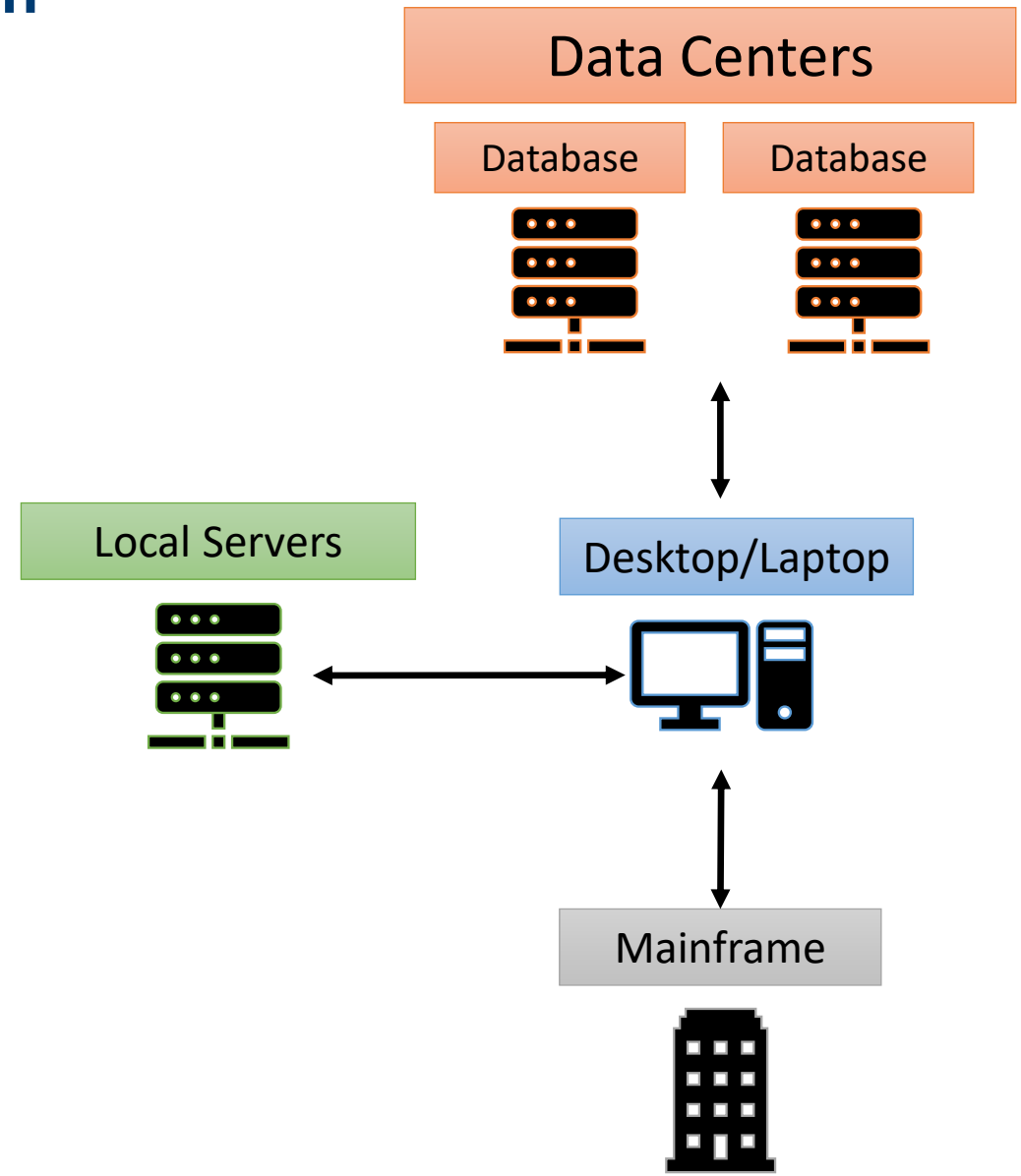
Consider Level of Methodology Support Needed

Past graduates of DOR's Programmer Analyst training program

	Research	Quality Improvement
Design	• Traditional study designs (cross sectional, cohort, case-control), pre/post • Typically, observational	• LEAN, Six Sigma • PDSA cycles, pre/post intervention • Typically, observational and prospective
Analytical Scope	• Descriptive • Bivariate and multivariate modeling • Analytical trends over time • Likely several power calculations	• Descriptive • Bivariate comparison • Descriptive trends over time (run charts) • Likely no power calculations • Process/root cause maps

Accessing Electronic Medical Records can be Complex

- May require sophisticated programming skills to pull and manage data
  - KPNC analyst training program – 6 months to initially learn databases and how to pull it
- Software commonly used to communicate with our databases:
  - Oracle SQL Developer
  - SAS SQL
  - R studio
- No programmer to pull data?
  - Chart review
  - Prospective



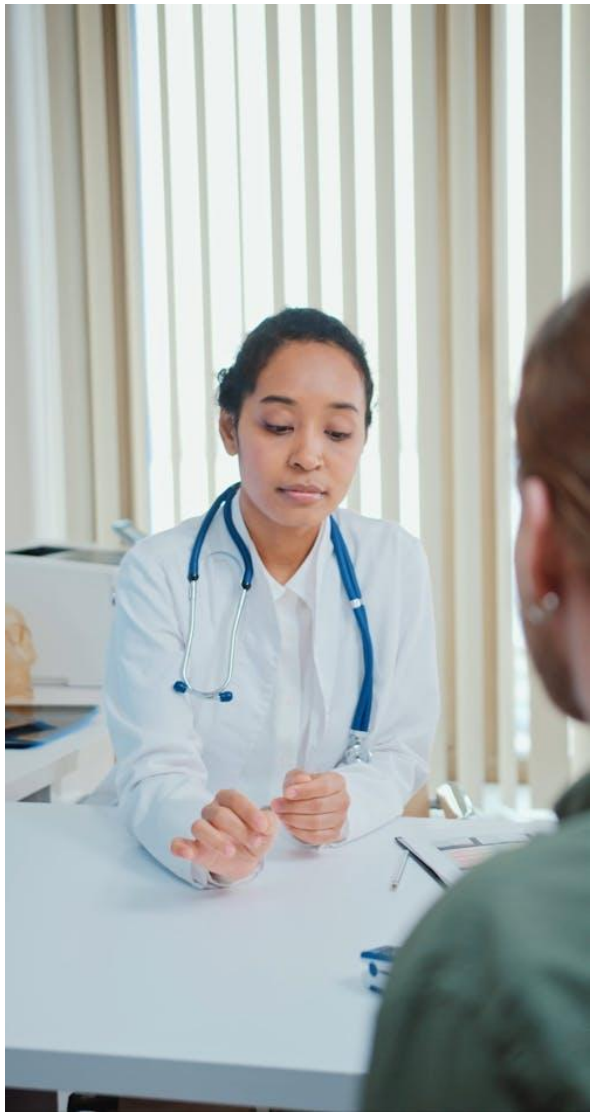
*Slide adapted from DOR Programmer Analyst Group training slides (July 2020)

Chart Reviewing can be Complex

- If no programmer to identify cohort, may require large # of charts for inclusion/exclusion
- If timelines are short, consider time for chart review and feasibility (resident rotations)
- If multiple chart reviewers, need to consider interrater reliability, typically need an analyst to conduct, takes time
- Keep chart review simple
- Use a structured data collection tool (DCT)

Prospective Methods

- Need to consider:
  - Recruitment (how you recruit can bias sample)
  - Informed consent (if research)
  - Incentives to increase response rate
  - Data collectors (surveyors, focus group facilitators)



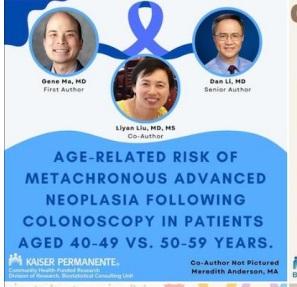
Additional Support & Considerations

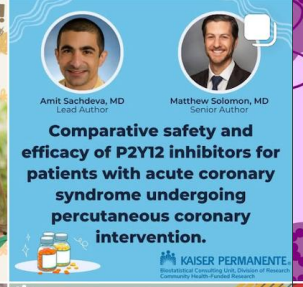
- **Local Support Staff**
  - Research project managers, research faculty
- **Compliance and Approvals**
  - Institutional Review Board (IRB) or Research Determination Office (RDO)
- **Leadership**
  - Local research chair, chief of dept, physician-in-chief, area manager, executives, legal
- **Presentation/Paper**
  - Medical editor if writing needs extra support
  - Programming/analysis in response to peer reviewers

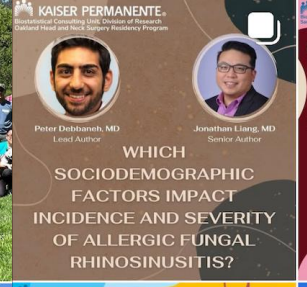
Key Takeaways

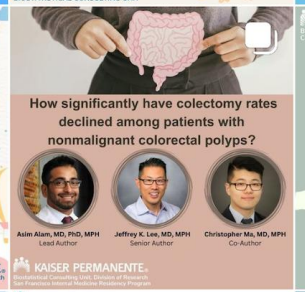
Key Takeaways

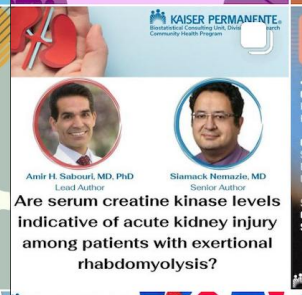
- Operationalize big ideas into analyzable variables
- Understanding the type of data collected will inform appropriate approaches to analysis
- Your type of analysis will inform your investigation's level of evidence
- Be realistic with the resources and support available to you

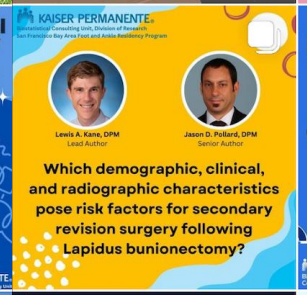


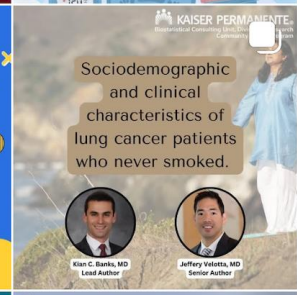


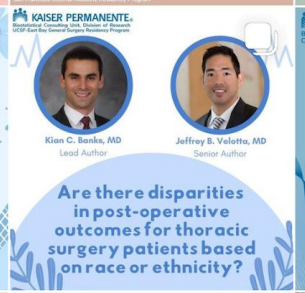


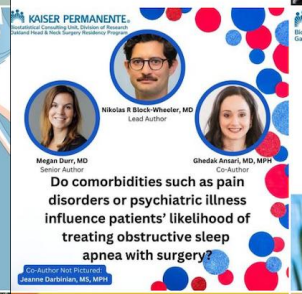


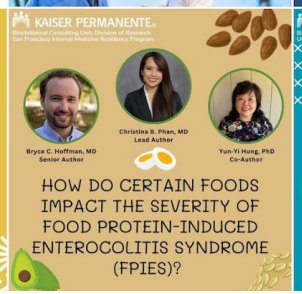


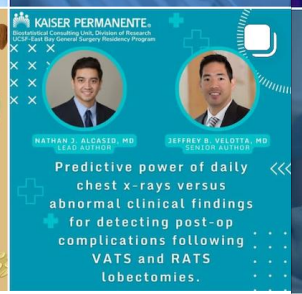


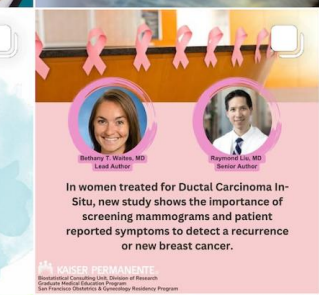


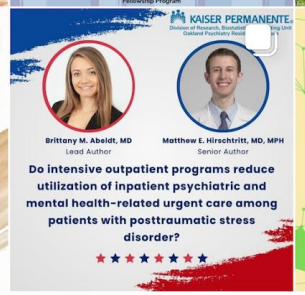


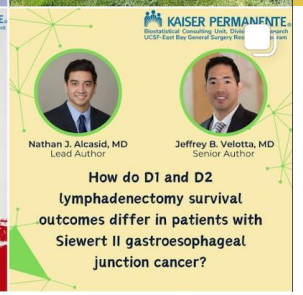












Follow us @kpbscu