Harvard Pilgrim Health Care Institute

CLINICAL EPIDEMIOLOGY AND POPULATION HEALTH

Key Points – Study Design, Causality, and Types of Variables

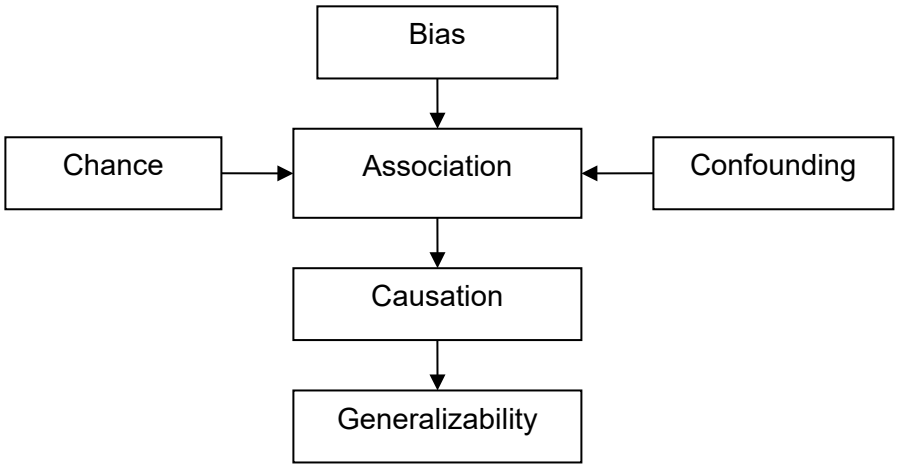
Interpreting Results from Scientific Studies

Studies usually present us with associations observed in a sample (e.g. vitamin Q intake associated with decreased mortality). Then…

We assess threats to **internal validity** by examining whether the result could be due to:

Chance - Random Error. Statistics helps us to quantify our uncertainty
Bias - Misleading associations introduced by research methods
Confounding - Factors related to both predictor and outcome which could obscure their true relationship

Even if internally valid, we have to decide if the association is **causal** and whether the findings are **generalizable** to individuals or populations not studied (e.g. the patients we will treat). This is sometimes called **external validity**.



Study Designs:

Most clinical decisions are based on evidence that comes from a handful of study designs. These designs vary in the timing of exposures and outcomes relative to the collection of data. We characterize studies based both on their timing and their structure.

Prospective studies follow individuals forward in time and, at some point in the future, collect data on outcomes. These include prospective cohort studies and randomized controlled trials.

Retrospective studies use existing data collected at some point in the past. These include case control studies (participants selected on the basis of outcome present / absent, and then their exposures are compared), and retrospective cohort studies (participants selected based on exposure or a common characteristic, measured in the past, and then outcomes are assessed either in the present or the future, relative to when the study begins).

Some studies measure everything at a single point in time (**cross-sectional**).

Here, we use the terms "exposure" and "outcome" generically. Exposures can include anything from asbestos to a new drug to treat arthritis; outcomes can include disease states or other events or attributes (e.g. satisfaction with medical care).

Cross-sectional studies – observational studies in which exposures and outcomes (e.g. a disease state) are measured simultaneously in a population.

Cohort studies – observational studies in which we follow a population forward in time collecting data on exposures and the development of outcomes. Most cohort studies are prospective: we assemble a cohort and start collecting data on exposure, then follow the cohort for development of outcome. Cohort participants are all free of the outcome at the time they are included in the cohort. Sometimes we can go back in time and find a cohort of people (e.g. a group of teens who were 9th graders in 2012), and use data that were collected at the time (e.g. their grades). Some call this a "retrospective cohort study," but people use this term variably. In these studies, outcomes may have occurred in the past and be already recorded (e.g. SAT scores in 11th grade), or can be assessed in the present/future by researchers themselves (e.g. career choice after college). The main point is that cohorts are identified by baseline characteristics, not according to outcome status.

Randomized Controlled Trials – experiments in which we randomly assign participants to a treatment group (the exposure) or a comparison (often a placebo) group and follow them forward in time for an outcome. Some trials have no control arm, they are called "single-arm trials". Because we are assigning exposure, trials are always prospective.

Case-Control studies– observational studies in which we identify a group with a particular outcome and another group without the outcome, and compare their past exposures. Because we choose based on outcome status, we always know the outcome status when we begin the study, so therefore case-control studies are always retrospective.

There are variations of each of these that we will be learning about in the next several weeks.

Causation

Asserting that an association is causal usually requires more than the assessment of a single study. It requires us to consider the totality of the evidence and our accepted understanding of biology and other sciences. The list of considerations put forward by Hill may be useful, but are not a "checklist" for determining cause. They include:

- Temporality
- Strength
- Consistency
- Specificity
- Dose-response (biological gradient)
- Plausibility
- Coherence
- Reversibility (or Experimental Evidence)
- Analogy

Even if a cause can be identified, it may only result in the outcome when other component causes are present. And, there may be many causal pathways to the same outcome (e.g. stroke may result from hypertension, dyslipidemia, atrial fibrillation with thromboembolism, etc.)

Types of Variables can depend on:

1. How we pose our research question:

Independent Variable – A variable we measure or manipulate (as an intervention or treatment**)** that may be associated with an outcome of interest. If it occurs before development of an outcome, it is sometimes called an **exposure** or **predictor variable.**

Dependent Variable (Outcome) – Measurable outcome of interest (e.g. a disease state, cure, or death) which we'd like to predict or explain with independent variables/predictors.

Note: depending on the question, the same variable may be an exposure in one analysis and the outcome in another (e.g. does smoking predict lung CA? does neighborhood of residence predict smoking?)

2. How the variable itself is structured:

Nominal or Categorical – Named but not necessarily ordered (may be dichotomous) - (e.g. sex, death, ethnic or cultural background).

Ordinal – Necessarily ordered categories where the distance between each unit is not defined (e.g. military ranks; or, a satisfaction scale of poor, fair, good, very good, excellent).

Interval - Discrete – Take on discrete (e.g. integer) values with equal magnitude between points (e.g. number of medications, days hospitalized).

Interval - Continuous – May take on any value over a continuum (e.g. height or weight).

Note: we can take a continuous variable and split it into two or more categories, e.g. we use body mass index (BMI), a continuous variable, to define the ordinal categories underweight, normal weight, overweight, and obese.