

Study Design 1: Overview & Cohort Studies

Stephen McCurdy, M.D.,
M.P.H.
Department of Public Health
Sciences
School of Medicine
University of California, Davis



Life's a Journey...



We each carry the burden of personal and group risk factors and exposures.

As health professionals, we hope to identify those characteristics causing disease.

Life's a Journey...



In individuals, the only way to know if a risk factor caused disease would be to find an exact double, living in a parallel universe, **identical in every way** to the exposed subject—except for the exposure.

Life's a Journey...



If only the exposed subject developed disease, we could be certain the exposure was causal.

Life's a Journey...



This is called the “counterfactual argument” because exact doubles and parallel universes do not exist.

(Stephen Hawking’s opinions notwithstanding...)

How to address this problem?

Life's a Journey...



The best we can do is compare **populations** that are similar (not identical) in everything except the risk factor.

If we see increased disease only in the group with the risk factor, we can suspect that the risk factor caused the disease.

Life's a Journey...



The “study base” is a population of individuals, each carrying the burden of personal and group risk factors.

(Rothman and Greenland, Modern Epidemiology, 1998)

Life's a Journey...



Epidemiologic studies are meant to examine the study base to determine if disease is more likely in the exposed group.

We will discuss three fundamental designs for this.

Behold!

Disease (Outcome)

+

-

**Exposure
(Risk Factor)**

+

-

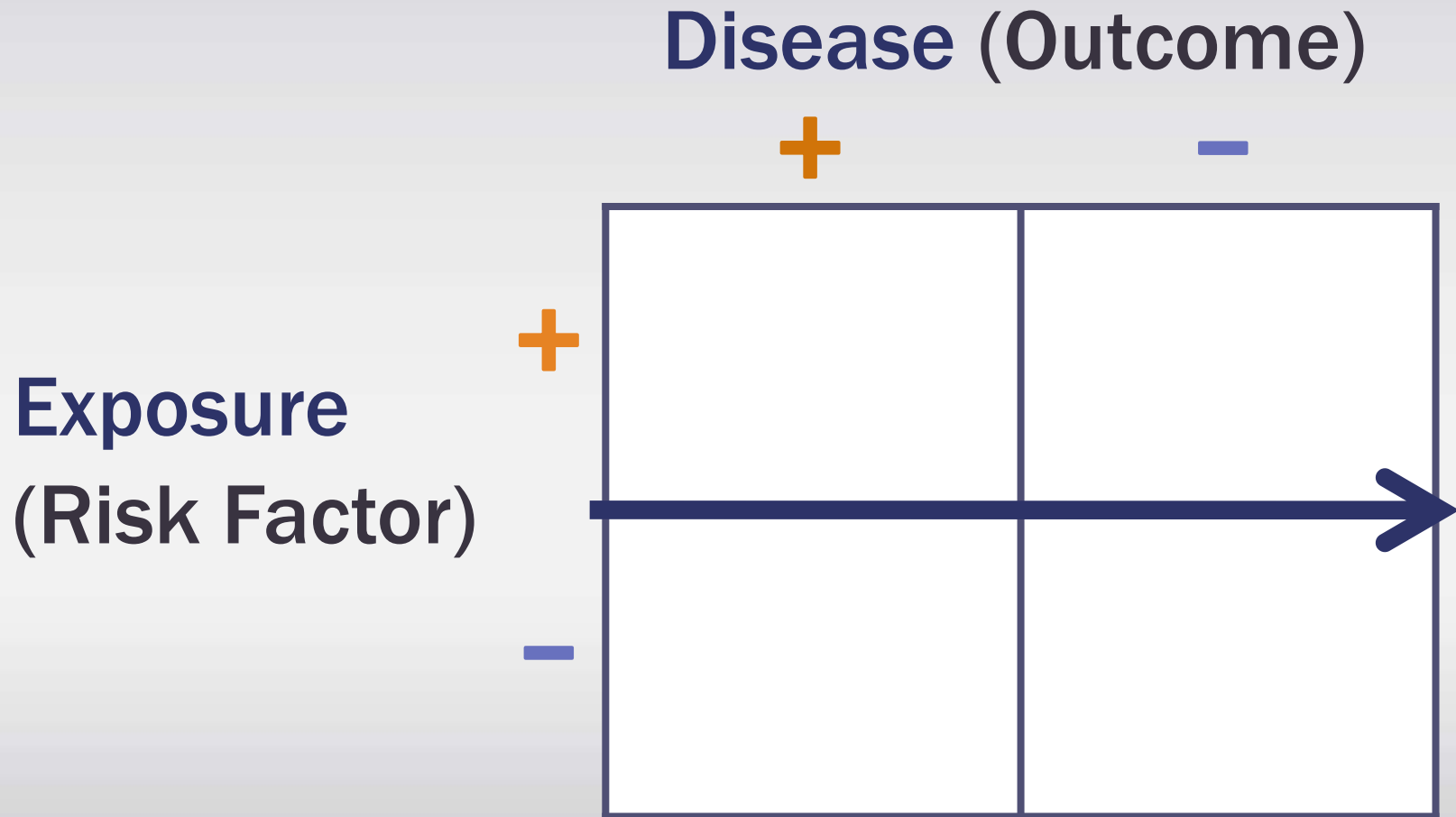
Objectives

1. Introduce concepts of “counterfactual argument” and “study base”
2. Review the three fundamental study designs
 - Cohort (including clinical trials)
 - Case-Control
 - Cross-Sectional survey
3. Discuss Cohort Studies
 - Uses
 - Strengths/weaknesses
 - Measure of effect (Relative Risk)

Cohort Studies

- Begin with sample “**Healthy Cohort**” (i.e., subjects without the outcome yet)
- Start with **Exposure** status, then compare **subsequent disease** experience in exposed vs. unexposed.

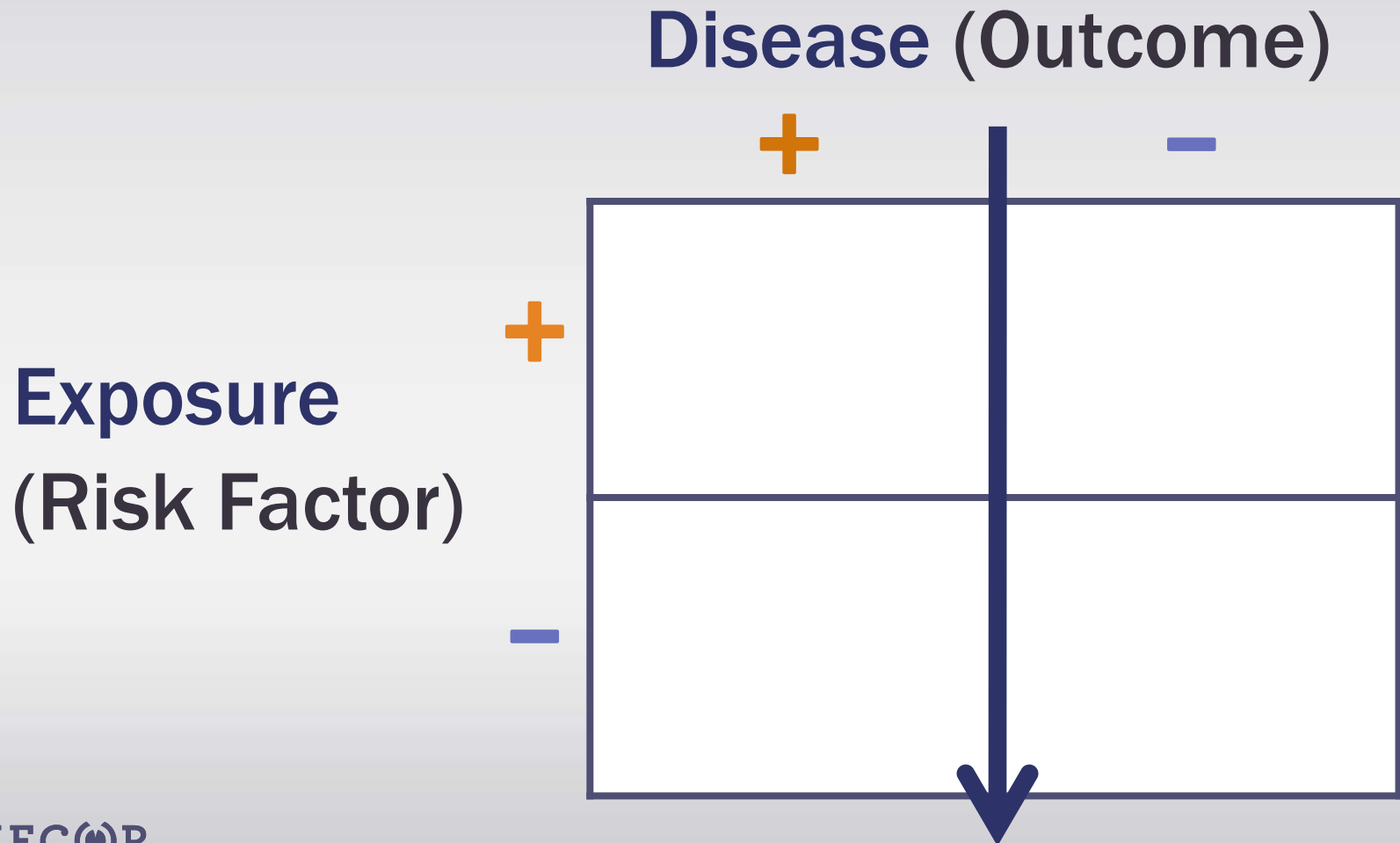
Study Design



Case-Control Studies

- Begin with sample of “**Cases and Controls**”
- Start with **Disease** status, then assess and compare **Exposures** in cases vs. controls.

Study Design



Cross-Sectional Studies

- Begin with “**Cross-sectional**” sample
- Determine **Exposure** and **Disease** at same time

Study Design

Disease (Outcome)

+

-

Exposure
(Risk Factor)

+

-



Cohort Study

Key Point

Presence or absence of risk factor is determined **before** outcome occurs.

Basic Idea

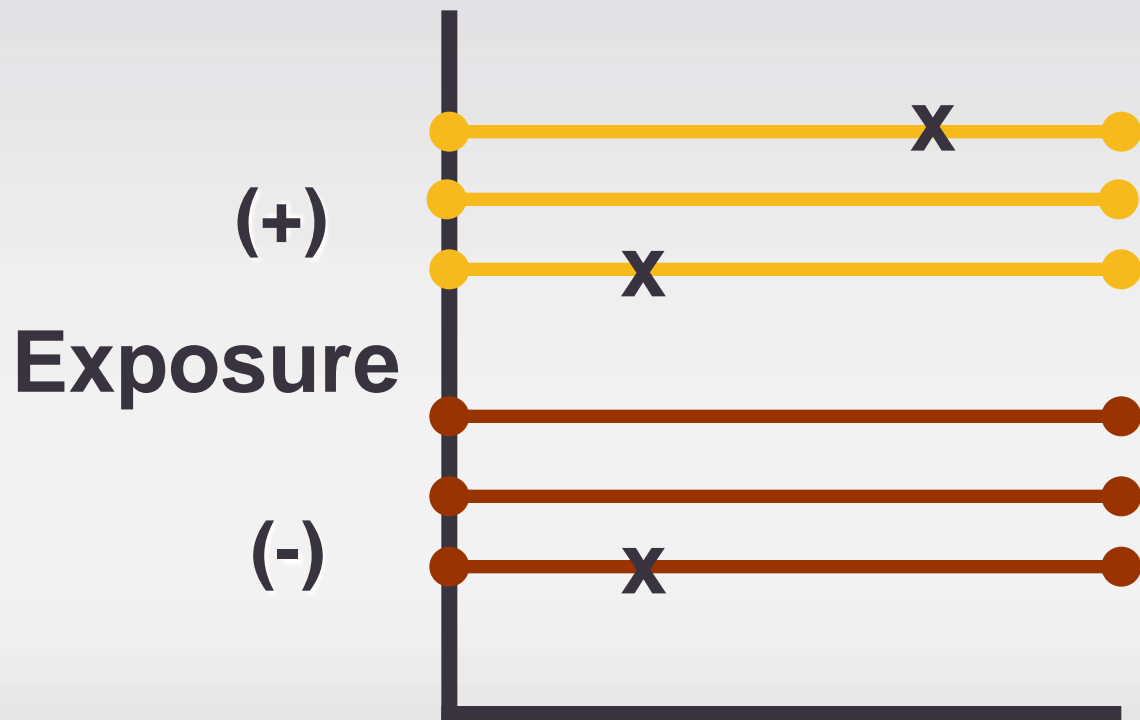
		Disease	
		+	-
R.F.	+	50%	50%
	-	10%	90%

See if those **with** the risk factor develop **more** disease than those without the risk factor

Basic Approach: Cohort Studies

- Identify Cohort(s)
- Measure exposure and outcome variables
- Follow for development of outcomes

Fixed Cohort



X = outcome

Relative risk
= (2/3) / (1/3)
= 2.0

Fixed Cohort

Disease: Hepatitis A

		+	-
Salad	+	30 a	70 b
	-	3 c	57 d

$$\text{Risk} = a/(a+b) = 0.3$$

$$\text{Risk} = c/(c+d) = 0.05$$

$$\text{Rel. risk} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

$$= 0.3/0.05$$

$$= \mathbf{6}$$

Fixed Cohort

Disease: Hepatitis A

	+	-
Salad +	30 a	70 b
Salad -	3 c	57 d

$$\text{Rel. risk} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

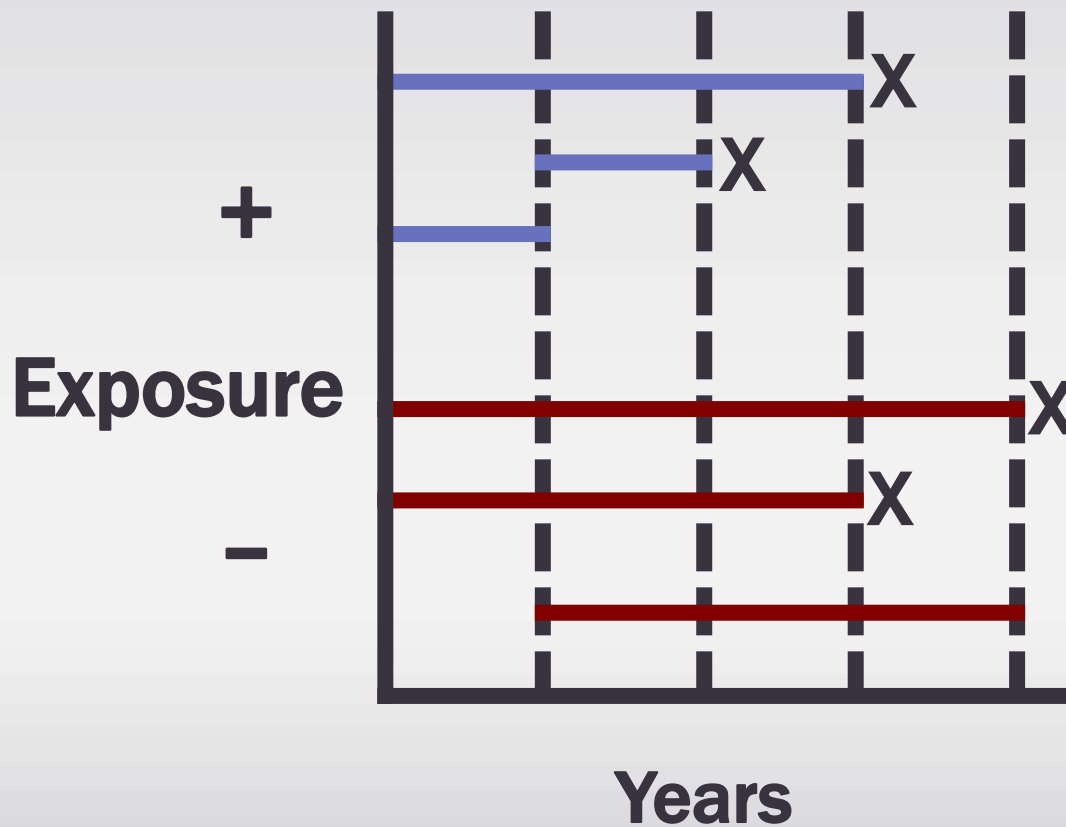
$$= 0.3/0.05 = \mathbf{6}$$

Odds Ratio:

$$(a/c)/(b/d) = (a/b)/(c/d)$$

$$(30/3)/(70/57) = \mathbf{8.14}$$

Dynamic Cohort



Rel. Risk =

$$\frac{2/3}{2/3} = 1$$

or

$$\frac{2/5_{\text{py}}}{2/10_{\text{py}}} = 2.0$$

Cohort Studies

- **Cohort:** 16, 936 Harvard grads
- **Measure:** Question re: activity level
- **Follow:** **Sedentary:** 24 CHD deaths per 10,000 person-years
vs.
Active: 16 CHD deaths per 10,000 person-years
- **Relative risk** = $24/16 = 1.5$

Questions

- Findings due to confounding?
- Could subclinical disease have affected the risk factor (activity)?

Take-Home Message

- The best measure of effect is the “relative risk.” For a fixed cohort, this will be the ratio of the cumulative incidences. For a dynamic cohort, this will be the ratio of the incidence rates.
- The odds ratio can be used for fixed cohorts comparing cumulative incidences. It will be close to the relative risk for rare diseases.

Variations on a Theme

Retrospective (Historical) Cohort

Cohort Studies

- **Prospective:** Outcomes have not yet occurred as study begins. Example: Women's Health Study.
- **Retrospective:** Outcomes have already occurred as the study begins. Example: finding a trove of medical records allowing you to follow a cohort born in 1880 to death.

Utility and Strengths

- Incidence and natural history
- Temporal sequence
- Avoid survivor bias
- Avoid reporting bias
- Look at multiple outcomes

Limitations

- Inefficient for rare diseases
- Confounding may occur
- Sub-clinical disease may affect risk factor levels
- Loss to follow-up

