

Appropriate measures of association for case-control studies

PHW250B - Jack Colford

Odds ratios - a mathematical convenience

- For many years, epidemiologists were taught that the only appropriate measure of association in a case-control study is an odds ratio.
 - This is because case-control studies sample from populations based on disease status.
- The mathematical equivalence of the odds of exposure and the odds of disease conveniently allowed epidemiologists to calculate a measure of association that did not require them to incorrectly pool across the diseased and non-diseased in a 2x2 table.

$$\text{Odds}_{\text{exp cases}} = \frac{\frac{a}{a+c}}{1 - \left(\frac{a}{a+c} \right)} = \frac{a}{c}$$

$$\text{Odds}_{\text{exp controls}} = \frac{\frac{b}{b+d}}{1 - \left(\frac{b}{b+d} \right)} = \frac{b}{d}$$

$$\text{OR}_{\text{exp}} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a \times d}{b \times c} = \frac{\frac{a}{b}}{\frac{c}{d}} = \text{OR}_{\text{dis}}$$

Odds ratios and the rare disease assumption

- Odds ratios are not naturally intuitive — usually what we're really interested in is the relative risk.
- When the disease of interest is rare, the odds ratio approximates the relative risk.
- $$RR = \frac{OR}{1 - [R_U - (OR \times R_U)]}$$
- R_U = risk or incidence in the unexposed
- As R_U approaches zero, $RR \approx OR$
- Often case-control studies are used to study rare diseases, so this conveniently meant that the OR could approximate the RR in most case-control studies.

What is a “rare” disease?
Typically a disease with incidence < 5% (or 5 per 1000) is considered rare. When incidence is > 10%, the OR and RR differ from each other substantially.

The modern view on odds ratios and case-control studies

- Today, epidemiologists learn that “in theory, every case-control study takes place within a cohort” (Wacholder et al., 1992).
- As a result, to determine the appropriate measure of association in a case-control study, it is necessary to consider the way in which cases and controls were sampled from the target population.
 - (We'll learn more about this in the case-control study unit.)
- In some case-control studies, it is not necessary to make the rare disease assumption when estimating an odds ratio.
 - Example: when the control group is a random sample of the target population

Case-control designs in which the OR approximates the RR

- The odds ratio directly estimates the relative risk (no rare disease assumption needed) when the control group is a sample of the total reference population.
 - This is true when the case-control study is defined within a cohort.
 - These are sometimes referred to as “population controls”.
- Two designs achieve this:
 - Case-cohort study
 - Nested case-control study

Case-control designs in which the OR approximates the RR

- **Case-cohort study**
 - Cases: individuals who develop disease in the cohort
 - Controls: study population (or random sample of study population) at baseline (including future cases)
- The division of the odds of exposure in cases by that in controls yields the relative risk (specifically, the cumulative incidence ratio):

$$\text{OR}_{\text{exp}} = \frac{\text{Odds}_{\text{exp cases}}}{\text{Odds}_{\text{exp total population}}} = \frac{\left(\frac{a}{c}\right)}{\left(\frac{a+b}{c+d}\right)} = \frac{\left(\frac{a}{a+b}\right)}{\left(\frac{c}{c+d}\right)} = \text{RR} \quad (\text{Eq. 3.14})$$

TABLE 3-11 Cross-tabulation of a defined population by exposure and disease development.

Exposure	Cases	Noncases	Total population (cases + noncases)
Present	a	b	a + b
Absent	c	d	c + d

Another advantage of this approach: you can calculate exposure prevalence, which allows you to calculate the **population attributable risk**

Example: case-cohort study

TABLE 3-12 Case-cohort study of the relationship of previous vaccination to local reaction.*

Previous vaccination	Cases of local reaction	Cohort sample
Yes	260	514
No	68	482
Total	328	996

$$OR_{exp} = \frac{\frac{260}{68}}{\frac{514}{482}} = 3.59 = RR$$

Case-control designs in which the OR approximates the RR

- **Nested case-control study**
 - Cases: incidence-density / risk set sampling
 - Controls: sampled at approximately the time when a case occurs
- The odds ratio of exposure represents an estimate of the relative rate or incidence density ratio.
- More on this later in the course.

Summary of what the OR estimates under different control sampling schemes

TABLE 3-13 Summary of the influence of control selection on the parameter estimated by the odds ratio of exposure in case-control studies within a defined cohort.

Design	Population frame for control selection	Exposure odds ratio estimates
Nested case-control	Population at approximate times when cases occur during follow-up	Rate (density) ratio
	(Population during follow-up minus cases)	(Density odds ratio)
Case-cohort	Total cohort at baseline	Cumulative incidence ratio (relative risk)
	(Total cohort at baseline minus cases that develop during follow-up)	(Probability odds ratio)

We'll return to this topic in more detail when we learn about case-control study designs.

Additive scale measures in case-control studies

- We can calculate the risk difference in case-control studies when the OR is a reasonable estimate of the RR.
- This is true when the disease is rare and/or when the control group is representative of non-cases in the study population (e.g., a case-cohort study).
- If either is true, we can substitute the OR for the RR in formulas for the attributable proportion among the exposed (APe%) and attributable proportion in the total population (APt%).

Estimating the attributable proportion among the exposed (APe%) in case-control studies

- First, let's derive the formula for the APe% that includes the RR. Then we can substitute the OR for the RR under certain conditions.

$$\begin{aligned} \text{APe\%} &= [(R_e - R_u) / R_e] \times 100\% \\ &= [1 - (1 / \text{RR})] \times 100\% \\ &= [(\text{RR} - 1) / \text{RR}] \times 100\% \end{aligned}$$

- If (1) the disease is rare and/or (2) the control group is representative of non-cases in the study population (e.g., a case-cohort study), we can substitute the OR for the RR.

$$\text{APe\%} = [(\text{OR} - 1) / \text{OR}] \times 100\%$$

Estimating the attributable proportion in the total population (APt%) in case-control studies

$$APt\% = [(R_t - R_u) / R_t] \times 100\%$$

P_e = prevalence of exposure

$$R_t = (R_e \times P_e) + R_u(1 - P_e)$$

Now we substitute this formula in for R_t and rearrange terms to obtain a version of the formula with the RR.

$$AP_t\% = \frac{(R_e \times P_e) + (R_u \times (1 - P_e)) - R_u}{(R_e \times P_e) + (R_u \times (1 - P_e))} \times 100$$

$$AP_t\% = \frac{(R_e \times P_e) + (R_u \times P_e)}{(R_e \times P_e) - (R_u \times P_e) + R_u} \times 100$$



Estimating the attributable proportion in the total population (APt%) in case-control studies (continued)

$$AP_t\% = \frac{(R_e \times P_e) + (R_u \times P_e)}{(R_e \times P_e) - (R_u \times P_e) + R_u} \times 100$$

Next, divide all terms by R_u

$$AP_t\% = \frac{\frac{R_e \times P_e}{R_u} + \frac{R_u P_e}{R_u}}{\frac{R_e \times P_e}{R_u} - \frac{R_u P_e}{R_u} + \frac{R_u}{R_u}} \times 100$$

Rearrange including RR in the formula

$$AP_t\% = \frac{(RR \times P_e) + P_e}{(RR \times P_e) - P_e + 1} \times 100 = \frac{(OR \times P_e) + P_e}{(OR \times P_e) - P_e + 1} \times 100$$

Berkeley



School of
Public Health

If (1) the disease is rare and/or
(2) the control group is
representative of non-cases in
the study population (e.g., a
case-cohort study), we can
substitute the OR for the RR.

Summary of key points

- Epidemiologists were previously taught that ORs were the only appropriate measure of association in case-control studies.
- The modern view of case-control studies is that they occur within a cohort.
- To assess what measure of association is estimated by the OR and whether the rare assumption is needed for the OR to approximate the RR, one must consider the method used to sample controls from the underlying cohort.

Use and misuse of population attributable fractions

PHW250B - Jack Colford

Based on the article by Rockhill et al. *Am J Public Health* 1998; 88(1).

Quick review: population attributable fractions (PAFs)

- **Most common definitions:**
 - Proportional reduction in average disease risk over a specific time interval that would be achieved by eliminating the exposure(s) of interest from the population while distributions of other risk factors in the population remain unchanged
 - Proportion of disease cases over a specified time that would be prevented following elimination of the exposures, assuming the exposures are causal
- **Example:** a cohort study found that air pollution was associated with the cumulative incidence of asthma (CIR=2.0). The cumulative incidence was 0.4 among the exposed and 0.2 among the unexposed. The total risk in the population was 0.9.
 - $\text{PAF} = (\text{Risk}_{\text{total}} - \text{Risk}_{\text{unexposed}}) / \text{Risk}_{\text{total}} \times 100\%$
 - $\text{PAF} = (0.9 - 0.2) / 0.9 = 0.78$
 - **Interpretation:** 78% of incident asthma cases in the study population would be prevented if no one in the study population was exposed to air pollution.

Different terms are often used interchangeably for PAFs

Measure of association	Formula	Interpretation
Population attributable fraction	$(R_{\text{total}} - R_{\text{unexposed}}) / R_{\text{total}} \times 100\%$	Proportion of disease cases over a specified time that would be prevented following elimination of the exposures, assuming the exposures are causal
Population attributable risk	$R_{\text{total}} - R_{\text{unexposed}}$	Difference in risk over a specified time that would be prevented following elimination of the exposures, assuming the exposures are causal
Excess fraction / Etiologic fraction	$(R_{\text{exposed}} - R_{\text{unexposed}}) / R_{\text{exposed}}$ <i>(definition in Rothman)</i>	Proportion of exposed risk that is attributable to exposure Note: to the etiologic fraction is not estimable without strong biologic assumptions

Distributive property of PAFs

- The PAF can be partitioned into exposure category-specific attributable fractions, which then sum to the population attributable fraction.
- Category-specific PAF = fraction of total risk in the population that would be eliminated if persons in only that specific exposure category were shifted to the unexposed group

$$= p_i \frac{(RR_i - 1)}{RR_i}$$

Where p_i is the proportion of total cases in the population arising from the i th exposure category.

Distributive property of PAFs

Example: calculate category-specific PAFs for both smoking categories

	Lung cancer	No lung cancer	Total
E2: Smoked 2+ packs per day	50	50	100
E1: Smoked up to 1 pack per day	30	120	150
E0: Non smokers	20	180	200
Total	100	325	450

$$RR_{E2} = 0.5 / 0.1 = 5$$

$$RR_{E1} = 0.2 / 0.1 = 2$$

$$p_{E2} = 50/100 = 0.5$$

$$p_{E1} = 30/100 = 0.3$$

$$PAF_{E2} = 0.5 * (5-1) / 5 = 0.4$$

$$PAF_{E1} = 0.3 * (2-1) / 2 = 0.15$$

$$PAF = 0.4 + 0.15 = 0.55$$

$$= (R_{total} - R_{unexp}) / R_{total} =$$

$$= (0.22 - 0.1) / 0.22 = 0.55$$

Caution: PAFs for different risk factors **cannot be summed** to derive the total fraction of disease risk attributable to all risk factors

- Example:
 - PAF for smoking = 0.55
 - PAF for air pollution = 0.56
- Almost always total PAF for smoking and air pollution $\neq 0.55 + 0.56$
- This will almost always yield a value larger than correctly calculated summary population attributable fraction for all of the factors considered simultaneously.
- This is because a set of individual PAFs from a study will often sum to more than 1.0.
 - Most epidemiologists assume that a given case of disease can be prevented by eliminating any of the necessary causal factors present.

Defining exposure / intervention levels in the PAF

- From a public health perspective, the PAF is most useful when the exposure of interest is:
 - Clearly causally related to the outcome
 - Amenable to intervention
- It is still common practice to estimate PAFs for **attributes that are not modifiable**, such as ethnicity or family history of cancer.
 - Example: A PAF for ethnicity and breast cancer does not provide the public health community with actionable information. Relative risks stratified by ethnicity may be more appropriate.
- Using a **surrogate for more proximate exposure** may also be less helpful than estimating the PAF for the exposure itself.
 - Example: Marital status might be a surrogate for certain health behaviors. Estimating the PAF for breast cancer and marital status does not provide information about how to reduce breast cancer because ensuring all women marry is not an appropriate or feasible public health intervention.

Defining exposure / intervention levels in the PAF

- Ideally the unexposed category in a PAF is defined in a way that makes it realistic for the exposed people to become unexposed.
- This is another reason why non-modifiable exposures are not ideal choices for the PAF.
- Even for modifiable exposures, in some instances it is difficult for the exposed to become unexposed:
 - PAF for lung cancer and never smokers (**not realistic**)
 - PAF for lung cancer and current non smoker (**realistic**)

Common misinterpretations of the PAF

- A study estimated a PAF of 25% for 10 breast cancer risk factors.
- **Quiz:** Here are several interpretations of this finding. Read these and indicate whether each was interpreted correctly or incorrectly.
 - “Although various risk factors have been identified as causes of breast cancer, the fact remains that in 75% of all breast cancer no identifiable risk factor can be found.”
 - “Only 25% of cancers could be attributed to one or more risk factors, meaning that the majority of cancers occur in women with no risk factors.”
 - “25% of the population risk of breast cancer would be eliminated if all 10 risk factors were to be eliminated from the population.”

Common misinterpretations of the PAF

- “Although various risk factors have been identified as causes of breast cancer, the fact remains that in 75% of all breast cancer no identifiable risk factor can be found.” **Incorrect**
- This is a very frequent misinterpretation of the PAF.
- This interpretation equates the PAF with the proportion of cases having any risk factors. The PAF measures the proportion of disease that would be prevented following elimination of the exposures, not the proportion of the population with the risk factors.
- The study evaluated 10 risk factors, but there are many other potential risk factors that were not studied or measured.

Common misinterpretations of the PAF

- “Only 25% of cancers could be attributed to one or more risk factors, meaning that the majority of cancers occur in women with no risk factors.” **Incorrect**
- The proportion of the population exposed to the considered risk factors is different from the PAF.
- The PAF itself does not provide information about the proportion of the population that was exposed.

Common misinterpretations of the PAF

- “25% of the population risk of breast cancer would be eliminated if all 10 risk factors were to be eliminated from the population” **Correct**
- This interpretation is consistent with the definition of the PAF.

Other sources of confusion about PAFs

- In interpreting PAFs, authors often use the words “cause”, “explain”, and “attribute”
 - Example: Study by Madigan et al., PAF = 0.40
 - “A substantial proportion of breast cancer cases in the US are explained by well-established risk factors”.
- Their data suggest that nearly all women in the US have at least one of the risk factors included in the PAF
- Yet, the vast majority of “exposed” women will never experience breast cancer.
 - Extreme example: PAF for age > 15 years and breast cancer. To imply that being over age 15 years causes breast cancer is not valuable.

Summary of key points

- The population attributable fraction provides information about how the population level risk would change under different hypothetical interventions:
 - Remove harmful exposure
 - Introduce beneficial intervention
- The PAF can be partitioned into exposure category-specific attributable fractions, which then sum to the population attributable fraction.
- PAFs for different risk factors cannot be summed to derive the total fraction of disease risk attributable to all risk factors
- From a public health perspective, the PAF is most useful when the exposure of interest is (1) Clearly causally related to the outcome and (2) Amenable to intervention

Hazard ratios

PHW250B

Quick review: hazard

- Hazard is defined as the **instantaneous potential for change in disease status** per unit of time at time t relative to the size of a disease-free population at time t
 - Synonyms for hazard: “force of morbidity”, “force of mortality”
- The **incidence density** averages over the hazard.
- In a cohort study, the hazard at time t is defined as follows:
 - $$h(t) = \frac{P(\text{event in interval between } t \text{ and } [t + \Delta t] \text{ | at risk at } t)}{\Delta t}$$
- In practice, this means that it is calculated among survivors who did not develop the disease at earlier time points in a study.
- The hazard cannot be calculated directly from data because it is defined for an infinitely small period of time. Instead, it is estimated using a statistical model.

Quick review: hazard ratios

It is difficult to make causal inferences about hazard ratios even when:

1. There is no unmeasured confounding
2. There is no measurement error
3. The appropriate statistical modeling approach is used

This is because:

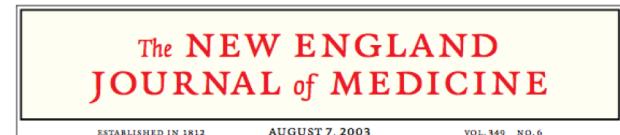
- Hazard ratios may change over time
- Hazard ratios have built in selection bias
 - Note: this can also be true of other measures of association (e.g. incidence density ratio) in cohort studies

Example: Women's health initiative

- Randomized trial
- ~16,000 women
- Treatment: hormone replacement therapy
- Primary outcome: heart disease
- Discontinued early due to concerns about the safety of the treatment

Women's health initiative main findings

- Overall HR = 1.24 ← Averages over time specific HRs
- Follow-up time 1 HR = 1.81
- Follow-up time 2 HR = 1.34
- Follow-up time 3 HR = 1.27
- Follow-up time 4 HR = 1.25
- Follow-up time 5 HR = 1.45
- Follow-up time 6+ HR = 0.70



Estrogen plus Progestin and the Risk of Coronary Heart Disease

JoAnn E. Manson, M.D., Dr.P.H., Judith Hsia, M.D., Karen C. Johnson, M.D., M.P.H., Jacques E. Rossouw, M.D., Ann Louise R. Assaf, Ph.D., Norman L. Lasser, M.D., Ph.D., Maurizio Trevisan, M.D., Henry R. Black, M.D., Susan R. Heckbert, M.D., Ph.D., Robert Detrano, M.D., Ph.D., Ora L. Strickland, Ph.D., Nathan D. Wong, Ph.D., John R. Crouse, M.D., Evan Stein, M.D., and Mary Cushman, M.D., for the Women's Health Initiative Investigators*

ABSTRACT

BACKGROUND

Recent randomized clinical trials have suggested that estrogen plus progestin does not confer cardiac protection and may increase the risk of coronary heart disease (CHD). In this report, we provide the final results with regard to estrogen plus progestin and CHD from the Women's Health Initiative (WHI).

METHODS

The WHI included a randomized primary-prevention trial of estrogen plus progestin in 16,608 postmenopausal women who were 50 to 79 years of age at baseline. Participants were randomly assigned to receive conjugated equine estrogens (0.625 mg per day) plus medroxyprogesterone acetate (2.5 mg per day) or placebo. The primary efficacy outcome of the trial was CHD (nonfatal myocardial infarction or death due to CHD).

RESULTS

After a mean follow-up of 5.2 years (planned duration, 8.5 years), the data and safety monitoring board recommended terminating the estrogen-plus-progestin trial because **the overall risks exceeded the benefits.** Combined hormone therapy was associated with a hazard ratio for CHD of 1.24 (nominal 95 percent confidence interval, 1.00 to 1.54; 95 percent confidence interval after adjustment for sequential monitoring, 0.97 to 1.60). The elevation in risk was most apparent at one year (hazard ratio, 1.81 [95 percent confidence interval, 1.09 to 3.01]). Although higher base-line levels of low-density lipoprotein cholesterol were associated with an excess risk of CHD among women who received hormone therapy, higher base-line levels of C-reactive protein, other biomarkers, and other clinical characteristics did not significantly modify the treatment-related risk of CHD.

CONCLUSIONS

Estrogen plus progestin does not confer cardiac protection and may increase the risk of CHD among generally healthy postmenopausal women, especially during the first year after the initiation of hormone use. This treatment should not be prescribed for the prevention of cardiovascular disease.

From the Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston (J.E.M.); the Department of Medicine, George Washington University, Washington, D.C. (J.H.); the Department of Preventive Medicine, University of Michigan, Ann Arbor (K.C.J.); the Preventive Cardiology Center, Mayo Clinic (M.T.); the Program in National Heart, Lung, and Blood Institute, Bethesda, Md. (J.E.R.); Memorial Hospital, Brown Medical School, Pawtucket, R.I. (A.R.A.); the Preventive Cardiology Program, New Jersey Medical School, Newark (R.L.L.); the Division of Social and Preventive Medicine, University at Buffalo, Buffalo, N.Y. (M.C.); the Department of Preventive Medicine, Rush-Presbyterian-St. Luke's Medical Center, Chicago (H.R.B.); the Department of Preventive Medicine, University of Washington, Seattle (S.H.-I.); the Division of Cardiology, Harbor-UCLA Research and Education Institute, Torrance, Calif. (R.D.); the Woodruff School of Nursing, Emory University, Atlanta (O.L.S.); the Heart and Stroke Prevention Program, University of California, Irvine (N.Z.); the Department of Medicine, Wake Forest University, Winston-Salem, N.C. (J.R.C.); Medical Research Laboratories International, Highland Heights, Ky. (E.S.); and the Departments of Medicine and Pathology, University of Vermont, Burlington (M.C.). Address reprint requests to Dr. Manson at the Division of Preventive Medicine, Brigham and Women's Hospital, 900 Commonwealth Ave., Boston, MA 02215, or at manson@rics.bwh.harvard.edu.

*The Women's Health Initiative (WHI) investigators are listed in the Appendix.

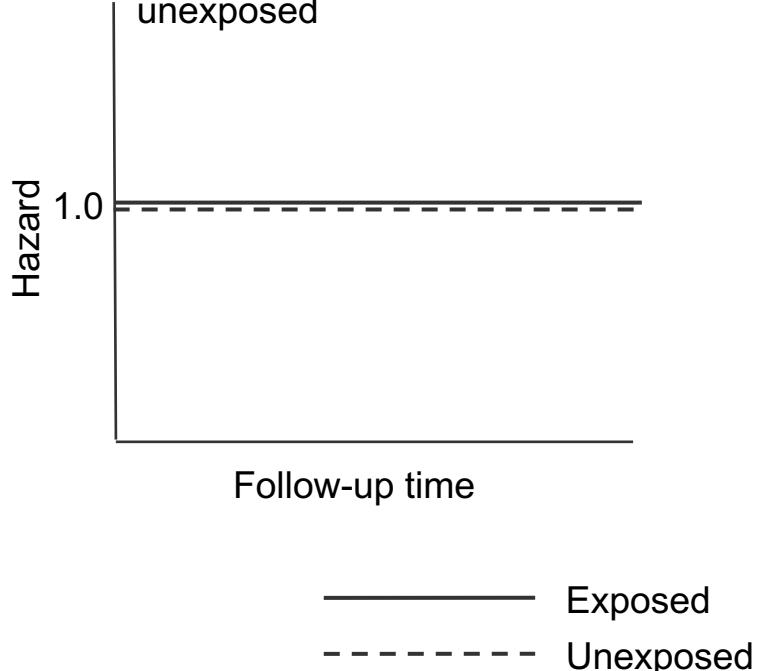
N Engl J Med 2003;349:523-34.
Copyright © 2003 Massachusetts Medical Society.

Values of HRs depend upon the follow-up time that they average over

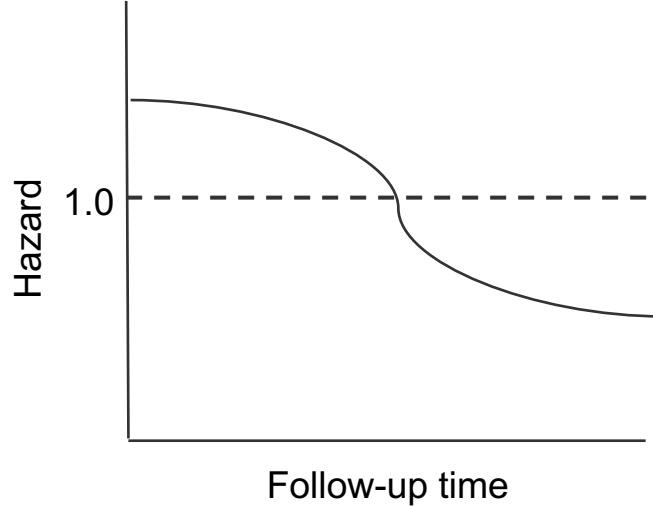
- Overall HR = 1.24
 - Follow-up time 1 HR = 1.81
 - Follow-up time 2 HR = 1.34
 - Follow-up time 3 HR = 1.27
 - Follow-up time 4 HR = 1.25
 - Follow-up time 5 HR = 1.45
 - Follow-up time 6+ HR = 0.70
- If there had only been 2 years of follow-up, the overall HR would have been ~1.58 instead of 1.24

In both plots, the HR = 1

The hazard is constant over time for the exposed and the unexposed



The hazard starts >1 and then falls <1 among the exposed and is constant among the unexposed



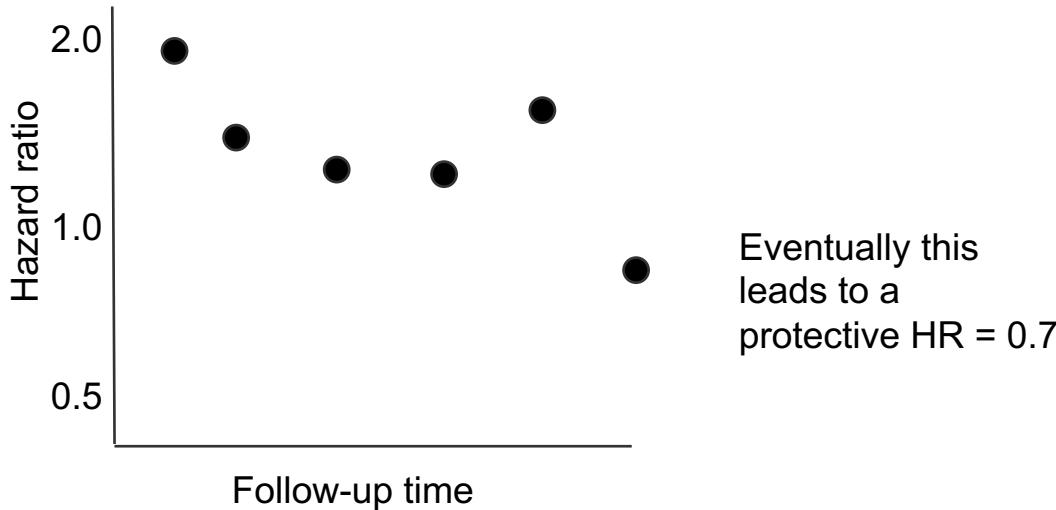
Take home message: HRs that average over long periods of time may obscure important time patterns in the hazard by exposure group

Can we just report period-specific HRs?

- No — they have selection bias built in.
- **Definition of hazard:** the instantaneous potential for change in disease status per unit of time at time t relative to the size of a disease-free population at time t
- Overall HR = 1.24
- Follow-up time 1 HR = 1.81
- Follow-up time 2 HR = 1.34
- Follow-up time 3 HR = 1.27
- Follow-up time 4 HR = 1.25
- Follow-up time 5 HR = 1.45
- Follow-up time 6+ HR = 0.70 ← Is this truly a protective effect of the intervention?

HRs have selection bias built in, so just reporting period-specific HRs is not enough to account for this bias

Each year there are fewer women left who are susceptible to heart disease and who are harmed by the treatment



What is the proper way to analyze hazard in a cohort study?

1. Plot Kaplan-Meier curve of cumulative risk, hazard, or survival
 - Plots the proportion of the study population that is disease-free by follow-up time
 - Provides information about the absolute risk that the HR does not
 - Typically unadjusted, but you can also adjust for confounders (see Hernan 2010)

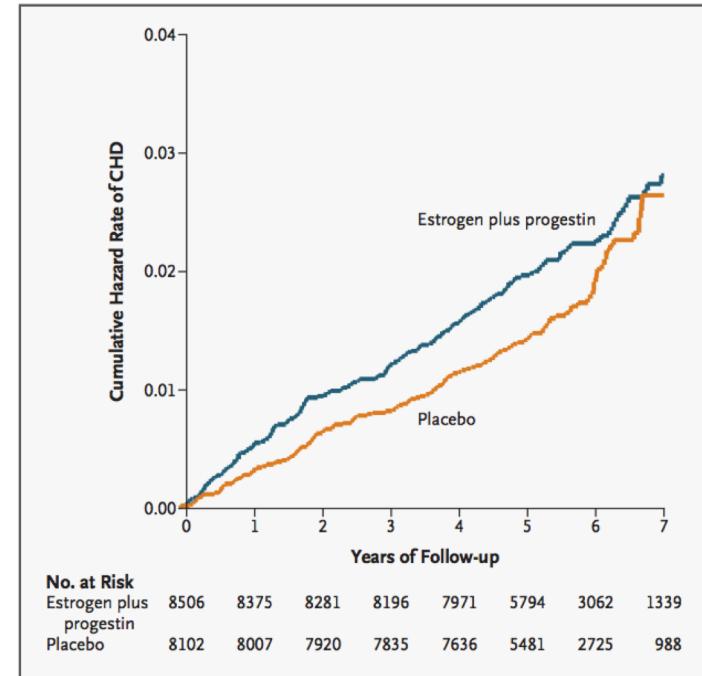


Figure 2. Kaplan-Meier Estimates of Cumulative Hazard Rates of CHD.

CHD included nonfatal myocardial infarction and death due to CHD. The overall hazard ratio for CHD was 1.24 (nominal 95 percent confidence interval, 1.00 to 1.54; 95 percent confidence interval with adjustment for sequential monitoring, 0.97 to 1.60).

Manson et al., 2003. N Engl J Med 2003;349:523-34.

What is the proper way to analyze hazard in a cohort study?

2. Estimate and report a set of HRs that having an increasing period of follow-up

- HR for 1 year = 1.8
- HR for 1-2 years = 1.7
- HR for 1-5 years = 1.2
- (But it's still a good practice to complement this with a survival or hazard curve)

Summary of key points

- Hazard ratios are commonly the primary measure of association in cohort studies.
- They are calculated among survivors who did not develop the disease at earlier time points in a study. As a result, they have selection bias built in.
- As a result, we must be careful about making causal inferences about hazard ratios even when 1) there is no unmeasured confounding, 2) there is no measurement error, and 3) the appropriate statistical modeling approach is used.
- Pooled HRs must be reported and interpreted with caution because they can obscure meaningful variation in hazard over time.
- Kaplan-Meier curves provide helpful complementary information to HRs related to the absolute risk / survival of events over time.

Causal perspective on measures of association

PHW250B

Traditional vs. modern perspectives on measures of association and causal inference

- In the traditional perspective:
 - Relative measures (RR, OR) provide information about etiology and causation
 - Absolute measures (RD) provide information about public health impact
 - These “epidemiologic traditions” have been “handed down from one generation to the next, without citation or critical reflection, as though their truth were self-evident.” — Poole, 2010. *Epidemiology* 21(1).
 - This assumption was tested in a heated controversy about smoking and lung cancer in the early 1950’s and 1960’s.
- In the modern perspective, epidemiologists use a simple model of the different “types” of people in the population to distinguish between the utility of the RR vs. the RD in making causal inferences.

Simple model to distinguish causation from association

Type	Response ^a under		Description	Proportion of types
	Exposure	Nonexposure		
1	1	1	Doomed	p_1
2	1	0	Exposure is causal	p_2
3	0	1	Exposure is preventive	p_3
4	0	0	Immune	p_4

- “Doomed”: disease occurs no matter what
- “Causal”: disease occurs if exposed
- “Preventive”: disease occurs if unexposed
- “Immune”: disease never occurs

Calculate the incidence for the exposed and unexposed

Type	Response ^a under		Description	Proportion of types
	Exposure	Nonexposure		
1	1	1	Doomed	p_1
2	1	0	Exposure is causal	p_2
3	0	1	Exposure is preventive	p_3
4	0	0	Immune	p_4

$$R_{\text{exp}} = p_1 + p_2 + p_3 + p_4$$

$$= p_1 + p_2$$

Calculate the incidence for the exposed and unexposed

Type	Response ^a under		Description	Proportion of types
	Exposure	Nonexposure		
1	1	1	Doomed	p_1
2	1	0	Exposure is causal	p_2
3	0	1	Exposure is preventive	p_3
4	0	0	Immune	p_4

$$R_{\text{exp}} = p_1 + p_2 + p_3 + p_4$$

$$= p_1 + p_2$$

$$R_{\text{unexp}} = p_1 + p_2 + p_3 + p_4$$

$$= p_1 + p_3$$

Calculate the incidence for the exposed and unexposed

Type	Response ^a under		Description	Proportion of types
	Exposure	Nonexposure		
1	1	1	Doomed	p_1
2	1	0	Exposure is causal	p_2
3	0	1	Exposure is preventive	p_3
4	0	0	Immune	p_4

$$R_{\text{exp}} = p_1 + p_2 + p_3 + p_4$$

$$= p_1 + p_2$$

$$R_{\text{unexp}} = p_1 + p_2 + p_3 + p_4$$

$$= p_1 + p_3$$

$$RR = (p_1 + p_2) / (p_1 + p_3)$$

$$RD = (p_1 + p_2) - (p_1 + p_3) = p_2 - p_3$$

Simple model to distinguish causation from association

Type	Response ^a under		Description	Proportion of types
	Exposure	Nonexposure		
1	1	1	Doomed	$p_1 = 10\%$
2	1	0	Exposure is causal	$p_2 = 50\%$
3	0	1	Exposure is preventive	$p_3 = 10\%$
4	0	0	Immune	$p_4 = 30\%$

If all exposed (N=100)

	Disease	No disease
Exposed	10+50	30+10
Unexposed	0	0

Summing up causal types:

$$RD = (p_1 + p_2) - (p_1 + p_3) = p_2 - p_3 = 0.5 - 0.1 = 0.4$$

Using 2x2 tables:

$$RD = 60/100 - 20/100 = 0.4$$

If all unexposed (N=100)

	Disease	No disease
Exposed	0	0
Unexposed	10+10	30+50

Interpreting “no effect” under this model

Type	Response ^a under		Description	Proportion of types
	Exposure	Nonexposure		
1	1	1	Doomed	p_1
2	1	0	Exposure is causal	p_2
3	0	1	Exposure is preventive	p_3
4	0	0	Immune	p_4

$$RR = (p_1 + p_2)/(p_1 + p_3)$$

$$RD = p_2 - p_3$$

- If there are equal causal and preventive types, $p_2 = p_3$ and $RR = 1$ and $RD=0$
- Thus $RR=1$ and $RD=0$ does not necessarily mean no effect but rather equality of causal types and preventive types.
- No effect would require $p_2 = p_3 = 0$.
- $RR=1$ or $RD=0$ are more accurately described as “no net effect” than “no effect”.

Influence of the “doomed” category on the RR

$$RR = (p_1 + p_2) / (p_1 + p_3)$$

$$RD = p_2 - p_3$$

Type	Response ^a under		Description	Proportion of types
	Exposure	Nonexposure		
1	1	1	Doomed	p_1
2	1	0	Exposure is causal	p_2
3	0	1	Exposure is preventive	p_3
4	0	0	Immune	p_4

Types amenable to intervention

Example 1: $p_1 = 0.75$, $p_2 = 0.19$, $p_3 = 0.01$, $p_4 = 0.05$

- $RR = (0.75 + 0.19) / (0.75 + 0.01) = 1.24$
- $RD = 0.19 - 0.01 = 0.18$

Example 2: $p_1 = 0.05$, $p_2 = 0.19$, $p_3 = 0.01$, $p_4 = 0.75$

- $RR = (0.05 + 0.19) / (0.05 + 0.01) = 4$
- $RD = 0.19 - 0.01 = 0.18$

- The RR depends not only on the people amenable to intervention (types 2 and 3) but also on the people who are doomed.
- The RD only depends on people amenable to intervention.
- If the proportion of the doomed in the population is high, the causal effect is diluted compared to when it is lower. In both cases, the RD stays the same.
- For this reason, some have argued that the RD is a better measure of causal impact than the RR.

Summary of key points

- In the traditional perspective: relative scale measures provide information about etiology and causation and absolute scale measures provide information about public health impact
- Some have argued that the RD is a better measure of a causal effect than the RR because the RR depends on the proportion of people who are doomed to get disease (and cannot be affected by intervention), whereas the RD isolates the impact on those who are amenable to intervention.