Berkeley School of Public Health

Introduction to matching

PHW250B

Overview of matching topics

- What is matching? Why match?
- Matching in different study designs
- Types of matching
- Matching in each study design
- Disadvantages of matching

This video

- Numerical examples of matching
- Overmatching
- Analysis of matched data

Upcoming videos

What is matching? Why match?

- Selection of a reference series (controls in a case-control study or unexposed in a cohort study) that is nearly identical to the index series with respect to one or more potential confounders.
- Generally the goal is to improve the quality of your counterfactual, whether that is the:
  - Control arm in a trial
  - Unexposed group in a cohort study
  - Controls in a case-control study
- Matching helps reduce bias (and approximate a counterfactual) by making study groups more comparable.
  - (There are other reasons to match too)

Matching in different study designs

Case-control studies

Match cases to controls

Example: Match women who experienced toxic shock syndrome to other women who did not experience it with similar ages who live in the same neighborhood

Cohort studies

Match exposed to unexposed

Match people receiving an intervention to people not receiving an intervention

Example: match students in a school vaccination program to students in similar schools not receiving the program

Trials

Match people in the intervention and the control group by a certain factor

Example: block randomization based on geographic area matches participants in each arm by geography

Types of matching

- **Individual:** match subjects together based on individual characteristics (e.g., age, sex, neighborhood) (also called “pair matching”)
- **Frequency:** match the distribution of characteristics between subjects
- **Distance:** match multiple characteristics at once using algorithms that find the distance between these characteristics

Frequency matching

- Match the distribution of characteristics between subjects
- The target population must be stratified into levels of the matching factor.
- The study is then conducted within each stratum of the matching factor.
    - E.g., in a case-control study this guarantees that there are both cases and controls within each stratum of the matching factor.
- In case-control studies, after frequency matching, it is no longer possible to estimate the association of the matched factor on the outcome.

Distance matching

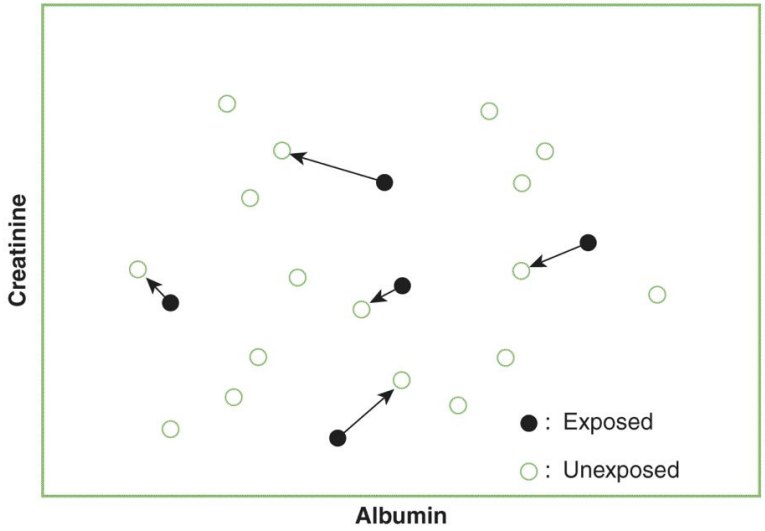


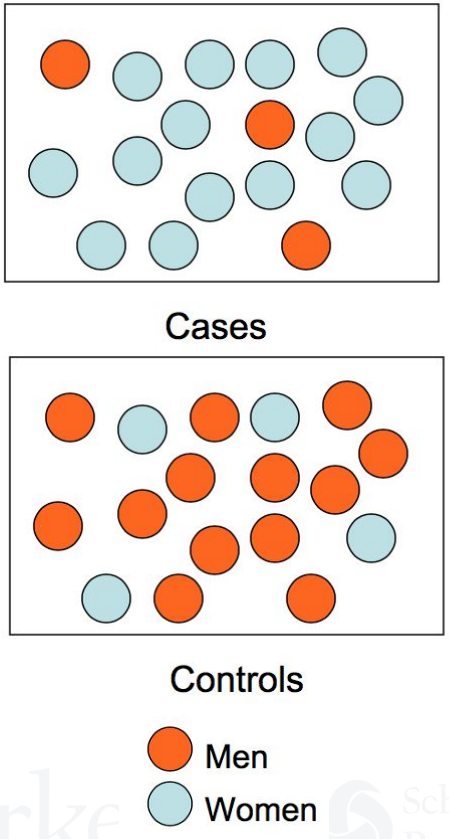
FIGURE 1-24 Matching according to minimal Euclidean distance measure method. Hypothetical example of a cohort study of survival after transplantation in multiple myeloma patients in which exposed individuals (e.g., older individuals) are matched to unexposed (younger) patients according to two prognostic factors: serum albumin and

- Distance measures identify people with the closest combination of multiple matching factors
- Particularly useful with continuous matching factors, and situations with many matching factors
- The figure shows only two potential matching variables (creatinine and albumin) but usually this method is used when there are multiple variables.
- In a case-control study for example, each case is matched to the control with the closest distance in bidimensional space.
- With multiple variables, individuals or groups would be matched in multidimensional space.

Szklo & Nieto, 3rd Ed.

Other reasons to match (besides reducing confounding)

- Increase **statistical efficiency** (precision)
  - Prevent confounder distributions that are dramatically different between cases and controls (or exposed and unexposed), as shown in the figure
- Increase **statistical power**
  - Most common in case-control studies - matching can help there are sufficient controls to estimate associations within important subgroups (e.g., gender)
- **Feasibility** during study enrollment
  - Example: in a study of a perinatal outcome, next birth could be the control (waiting to take a random sample would delay selection etc.



Matching in case-control studies

- Matching can introduce bias in case-control studies; thus, matched case-control studies must utilize a matched analysis. (See upcoming video on this topic)
- The primary reason to match in case-control studies is to improve statistical efficiency (i.e., precision).
- Case-control studies tend to be small so there is concern about overlap in the distribution of confounders if the cases of disease likely to differ from the study base dramatically on strong confounder(s) (e.g., age).
- In small studies, some strata may have few observations in them, making it difficult/impossible to adjust for confounding and inflating standard errors.
- In case-control studies, the main purpose of matching is to avoid strata with small cells. Matching ensures that after stratification by the matched factor, there will be cases and controls in each stratum.

What sparse data looks like

- Red cells in the table below are sparse - disease that is rare among people <18 years

Age < 18 years		
	Disease	No disease
Exposed	3	200
Unexposed	0	450

Age ≥ 18 years		
	Disease	No disease
Exposed	86	3824
Unexposed	351	8900

Example of matched case-control study

- Study of toxic shock syndrome and tampon use. Controls were friends of cases and other women matched on age and neighborhood. Main purpose was to reduce confounding.

Table 1. Characteristics of patients and controls enrolled in a multistate study of risk factors for menstrual toxic shock syndrome.

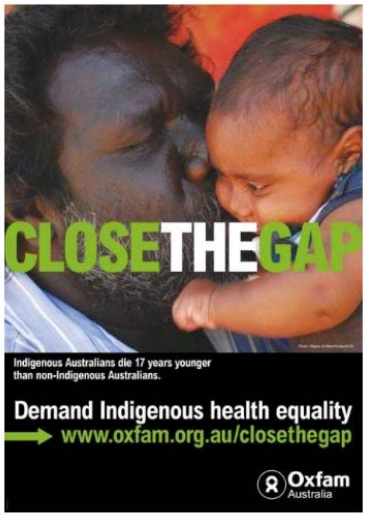
	Value for indicated group			
Characteristic (unit)	Patients	Friend controls	Neighborhood controls	Combined controls
Mean age (y)*	24.3 ± 8.1 (13–46)	24.8 ± 8.4 (11–48)	24.5 ± 8.1 (13–48)	24.6 ± 8.2 (11–48)
White (%)	94	94	89	91
Married (%)	44	39	36	37
Interval from onset of index menstrual period to interview (d)*	88 ± 50 (25–249)	. . .	. . .	87 ± 51 (17–281)
Interviews successfully completed with blinding to case/control status (%)	82	91	87	89

* Values given are mean ± SD (range).

Reingold et al., 1989

Matching in cohort studies

- Studies tend to be large and intended to examine multiple exposures and outcomes so other approaches to controlling confounding are usually preferable
- **Prognostic cohort studies** use matching more often because they typically have one exposure and are smaller, so there is concern about overlap in distribution of confounders
- Example: study comparing prognosis of Indigenous Australians with cancer to other Australians with cancer in Queensland
- Indigenous people diagnosed with cancer identified through the cancer registry (n~800)
- Compared with randomly selected non-Indigenous patients who were frequency-matched for age, sex, place of residence, cancer site, and year of diagnosis



Matching in cohort studies

- **Impact evaluations** use matching when randomization is not feasible. Goal is to evaluate an existing, non-randomized intervention.
- Use matching with census (or other data) to identify individuals / communities who can serve as a control group who can approximate a control group in a trial
- Example: evaluation of a sanitation mobilization, water supply, and hygiene intervention in rural India
- Intervention in 12 villages
- Identified 13 control villages using pre-intervention census data and matching

Arnold et al., 2010

Matched cohort study - control selection

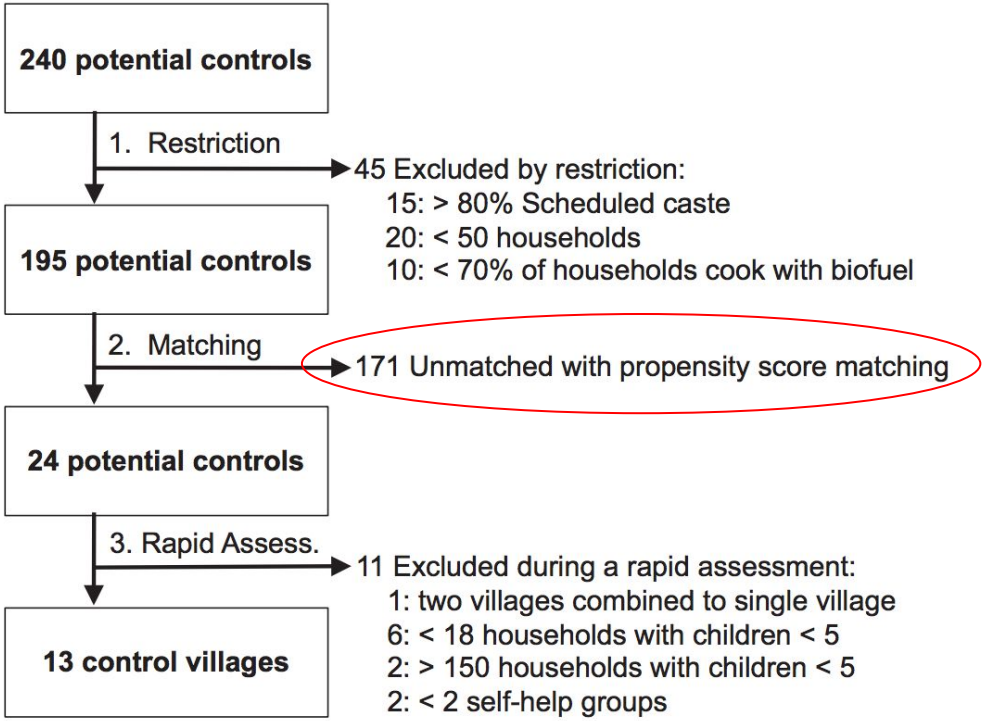


Fig. 2. Control village selection process in the Tamil Nadu study.

Arnold et al., 2010

Matched cohort study - comparison of study groups following matching

- The “All villages” columns compare intervention and control villages prior to matching.
- The “Study sample” columns compare intervention and control villages after matching.
- Characteristics were very similar between groups after matching.

Table 2. Summary of preintervention characteristics before and after village selection

	All villages		Study sample	
Mean	Control	Intervention	Control	Intervention
Demographic				
Total households	170	161	181	161
Persons per household	5	5	5	5
Scheduled caste, %	19	12	15	12
Children ≤5 y old, %	12	12*	12	12
Female literacy, %	52	48	49	48
Socioeconomic				
Employment rate, %	81	78	79	78
Cultivators, %	27	28	31	28
Agricultural laborers, %	24	33	21	33
Marginal workers, %	19	22	21	22
Females work, %	74	69	71	69
Panchayat income (Rp/person)	12,255	7,470***	7,143	7,470
Per-capita cattle ownership	4	4	5	4
Use banking services, %	29	25	25	25
Use biofuel for cooking, %	91	97**	96	97
Own radio, %	43	43	38	43**
Own television, %	21	16	17	16
Own scooter/moped, %	10	10	9	10
Sanitation and water				
Private toilet/latrine, %	15	8**	9	8*
Open defecation, %	85	92**	91	92*
Tap water (private/public), %	75	76	75	76
Hand pump, %	12	14**	18	14
Other water source, %	13	10*	7	10
Persons per hand pump	260	302	240	302
Persons per deep bore well	437	679**	510	679
Water supply level (lpcd)	12	15**	14	15
No. of villages	240	12	13	12

Arnold et al., 2010

Matching in trials

- Though less common in trials, blocked randomization implies matching by block.
- Increases comparability of participants in different randomized arms within the same block.
    - This increases statistical efficiency (reduces the width of confidence intervals) when the blocking variable is strongly correlated with outcomes.
- Example: in the WASH Benefits trials, village clusters were randomized within geographic blocks.
    - Ensured that the study arms were balanced with respect to characteristics and outcomes that were clustered within space.
    - One of the primary outcomes of the trials was diarrhea. Enteric infections are known to be highly spatially clustered.

Disadvantages of matching

- Cost – it can complicate the sampling scheme
- Exclusion of cases when no match found - reduces N (pair matching)
- Longer study duration if matching increases the length of the enrollment period
- Reduced flexibility in analysis
    - Cannot estimate association of the matched variable
    - Requires use of matched pairs analysis and conditional logistic regression (pair matching) (more on this in a future video)
- Improper matching can prevent estimation of effects of interest—over-matching (more on this in a future video)

Summary of key points

- Using a matched design can help reduce confounding or increase statistical efficiency.
- Different types of matching exist that allow for investigators to match on one or multiple variables and to match pairs or groups of individuals.
- Matching can be done in any type of epidemiologic study. In a future video we will discuss the implications on matching in the analysis of different study designs.
- Matching can increase the cost and complexity of a trial, so usually it is best to match when the benefits in increasing validity or increasing statistical efficiency outweigh any such implications.

Berkeley School of Public Health

Numerical examples of matching

PHW250B

Overview of matching topics

- What is matching? Why match?
- Matching in different study designs
- Types of matching
- Matching in each study design
- Disadvantages of matching
- Numerical examples of matching } **This video**
- Overmatching
- Analysis of matched data

Numerical examples of matching

Examples from Rothman et al. *Modern Epidemiology*

- Cohort study
  - Does not necessitate control of the matched factor in the analysis
- Case-control study
  - Necessitates control of the matched factor in the analysis

Matching in a cohort study

Example of data from target population

	Men		Women	
	Disease	**Total**	**Disease**	**Total**
Exposed	4,500	900,000	100	100,000
Unexposed	50	100,000	90	900,000

Crude CIR pooling across gender: $\frac{(4500 + 100) / (900{,}000 + 100{,}000)}{(50 + 90) / (100{,}000 + 900{,}000)} = 33$

CIR for Men: $\frac{4500 / 900{,}000}{50 / 100{,}000} = 10$

CIR for Women: $\frac{100 / 100{,}000}{90 / 900{,}000} = 10$

There is strong confounding by gender.

Rothman et al. *Modern Epidemiology.* 3rd Ed.

Matching in a cohort study

Designing a cohort study to control for confounding by gender

	Men		Women	
	Disease	**Total**	**Disease**	**Total**
Exposed	4,500	900,000	100	100,000
Unexposed	50	100,000	90	900,000

- Notice that 90% of those exposed are male and 10% are female.
- We want to conduct a cohort study using a 10% sample of the exposed.
- If we draw a random sample of the unexposed, the same confounding will occur.
- If we draw a sample so that the proportion of men matches that in the exposed cohort, how much confounding remains?

Rothman et al. *Modern Epidemiology.* 3rd Ed.

Matching in a cohort study

blue: sampling fraction
black: number in cohort study

Cohort study matching on gender

	Men		Women	
	Disease	**Total**	**Disease**	**Total**
Exposed	4,500***10%** **= 450**	900,000***10%** **= 90,000**	100***10%** **= 10**	100,000***10%** **= 10,000**
Unexposed	50***90%** **= 45**	100,000***90%** **= 90,000**	90***1.1%** **= 1**	900,000***1.1%** **= 10,000**

10% sampling fraction for exposed

Sampling fraction for unexposed that makes the totals equal for each gender

Rothman et al. *Modern Epidemiology.* 3rd Ed.

Matching in a cohort study

blue: sampling fraction
black: number in cohort study

Cohort study matching on gender

	Men		Women	
	Disease	**Total**	**Disease**	**Total**
Exposed	4,500***10%** **= 450**	900,000***10%** **= 90,000**	100***10%** **= 10**	100,000***10%** **= 10,000**
Unexposed	50***90%** **= 45**	100,000***90%** **= 90,000**	90***1.1%** **= 1**	900,000***1.1%** **= 10,000**

10% sampling fraction for exposed

Sampling fraction for unexposed that makes the totals equal for each gender

Crude CIR pooling across gender: $\frac{(450 + 10) / (90{,}000 + 10{,}000)}{(45 + 1) / (90{,}000 + 10{,}000)} = 10$

CIR for Men: $\frac{450 / 90{,}000}{45 / 10{,}000} = 10$

CIR for Women: $\frac{10 / 10{,}000}{1 / 10{,}000} = 10$

Confounding by gender was removed by matching on gender.

Rothman et al. *Modern Epidemiology.* 3rd Ed.

Take home message: matching in cohort studies

- Matching on a confounder in a cohort study can remove confounding.
- When a cohort study is matched in the design phase, it is not necessary to adjust for the confounder that was matched on in the analysis because the matching already removed confounding.

Matching in a case-control study

Example of data from target population

	Men		Women	
	Disease	**Total**	**Disease**	**Total**
Exposed	4,500	900,000	100	100,000
Unexposed	50	100,000	90	900,000

	Cases	**Controls**
Exposed	4,600	4,114
Unexposed	140	626

Suppose that we sample 4,740 controls from the target population that are matched on sex.

Crude OR = 4,600*626 / 4,114*140 = 5.0

This OR is lower than the true RR of 10.

Rothman et al. *Modern Epidemiology.* 3rd Ed.

Matching in a case-control study

Example of sampling the target population for the case-control study

	Men		Women	
	Cases	**Controls**	**Cases**	**Controls**
Exposed	4,500		100	
Unexposed	50		90	
Total	4,550	4,550	190	190

- All 4740 individuals with disease included as cases.
- Equal number of controls matched on gender are sampled from the target population.

Rothman et al. *Modern Epidemiology.* 3rd Ed.

Matching in a case-control study

Example of sampling the target population for the case-control study

	Men		Women	
	Cases	**Controls**	**Cases**	**Controls**
Exposed	4,500	4,095	100	
Unexposed	50	455	90	
Total	4,550	4,550	190	190

- All 4740 individuals with disease included as cases.
- Equal number of controls matched on gender are sampled from the target population.
- Of the 4550 male controls, assume 90% are exposed (N=4095) and 10% unexposed (N=455).

Rothman et al. *Modern Epidemiology.* 3rd Ed.

Matching in a case-control study

Example of sampling the target population for the case-control study

	Men		Women	
	Cases	**Controls**	**Cases**	**Controls**
Exposed	4,500	4,095	100	19
Unexposed	50	455	90	171
Total	4,550	4,550	190	190

- All 4740 individuals with disease included as cases.
- Equal number of controls matched on gender are sampled from the target population.
- Of the 4550 male controls, assume 90% are exposed (N=4095) and 10% unexposed (N=455).
- Of the 190 female controls, assume 10% are exposed (N=19) and 90% unexposed (N=171).

Rothman et al. *Modern Epidemiology.* 3rd Ed.

Matching in a case-control study

Example of sampling the target population for the case-control study

	Men		Women	
	Cases	**Controls**	**Cases**	**Controls**
Exposed	4,500	4,095	100	19
Unexposed	50	455	90	171
Total	4,550	4,550	190	190

Crude OR = ((4,500 + 100) * (455 + 171)) / ((4,095 + 19) * (50 + 90))= 5.0 (Much lower than true RR=10)

Male OR = 4500 * 455 / 4,095* 50 = 10

Female OR = 100 * 171 / 90* 19 = 10

Confounding of the pooled OR was not removed by matching on gender.

Rothman et al. *Modern Epidemiology.* 3rd Ed.

Take home message: matching in case-control studies

- Matching on a confounder in a case-control study can remove confounding but only if you stratify on the confounder in the analysis.
- When a case-control study is matched in the design phase, it is necessary to adjust for the confounder that was matched on in the analysis because matching alone has not removed confounding.
- If you do not stratify on the confounder in the analysis, there will be bias towards the null.

Summary of key points

- In a matched cohort study, **it is not necessary** to adjust for the confounder that was matched on in the analysis.
- In a matched case-control study, **it is necessary** to adjust for the confounder that was matched on in the analysis.

Berkeley School of Public Health

Diagnosing overmatching with DAGs

PHW250B

Overview of matching topics

- What is matching? Why match?
- Matching in different study designs
- Types of matching
- Matching in each study design
- Disadvantages of matching
- Numerical examples of matching
- Overmatching } **This video**
- Analysis of matched data

What variables should be matched on?

- Strong confounders
  - Variables that strongly affect the outcome that you expect to have very different distributions between
    - Cases and controls
    - Exposed and unexposed
- Variables whose effects on disease are not of scientific interest:
  - Age, race, sex
- If the variable has a weak association with disease, concerns about cost efficiency and potential misclassification may justify matching on that variable.

What is overmatching?

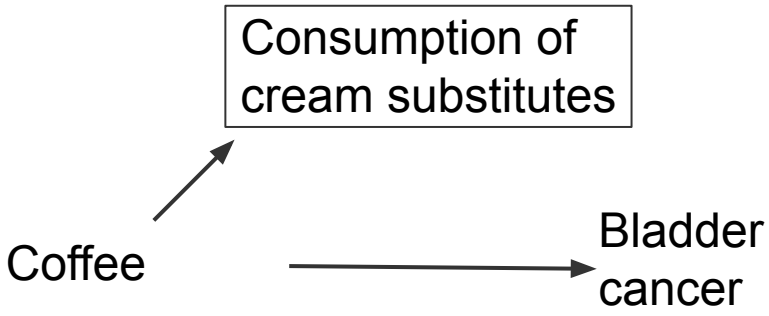
- Overmatching occurs when matching on a non-confounder.
- It can occur when matching on an intermediate between exposure and disease, or a factor that is affected by both exposure and disease can lead to bias
- **Types of overmatching:**

1) Overmatching that harms statistical efficiency (precision)
2) Overmatching that harms validity (accuracy)

Rothman et al., 3rd Ed.

Overmatching that harms statistical efficiency

- Results from matching on a non-confounder associated with exposure but not disease
- Causes a loss of information in the analysis because the stratified analysis would have been unnecessary without matching.
- Worst candidate for matching: variable strongly correlated with exposure but not disease.

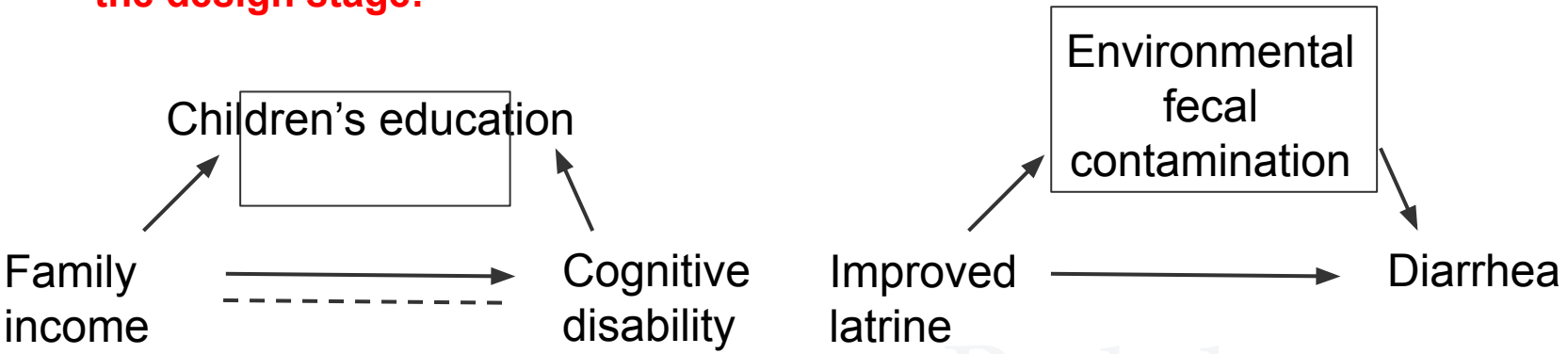


In this example, many controls matched to cases will be classified identically to the case with regard to coffee drinking merely because they also consume cream substitutes.

Rothman et al., 3rd Ed.

Overmatching that harms validity

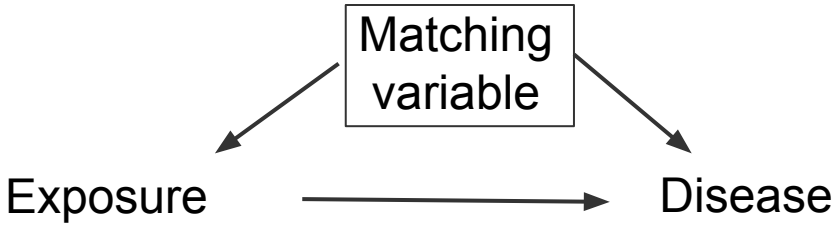
- Results from matching on a variable that is affected by the exposure or disease or both (a collider)
- Matching on (i.e. conditioning on) a collider of the exposure and outcome opens a backdoor pathway between them, introducing bias.
- Matching on an intermediate also introduces bias.
- **This bias cannot be fixed in the analysis! It is very important to avoid in the design stage!**

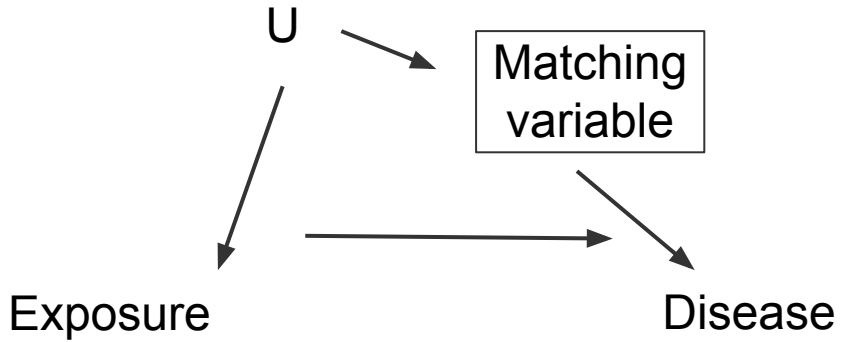


Rothman et al., 3rd Ed.

Summary of key points

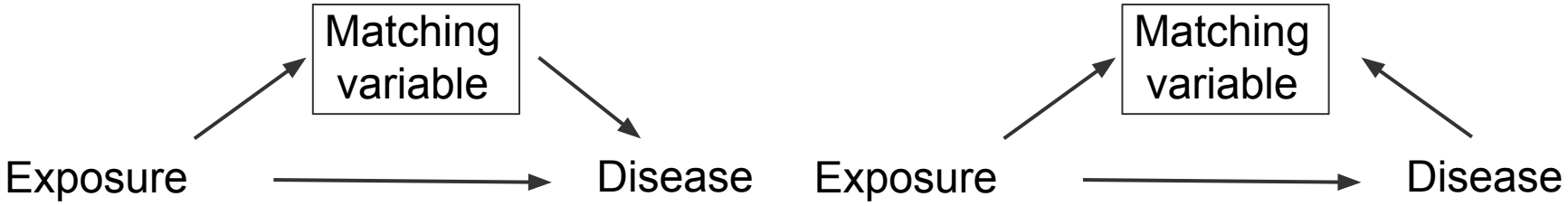
DAGs in which matching reduces confounding:





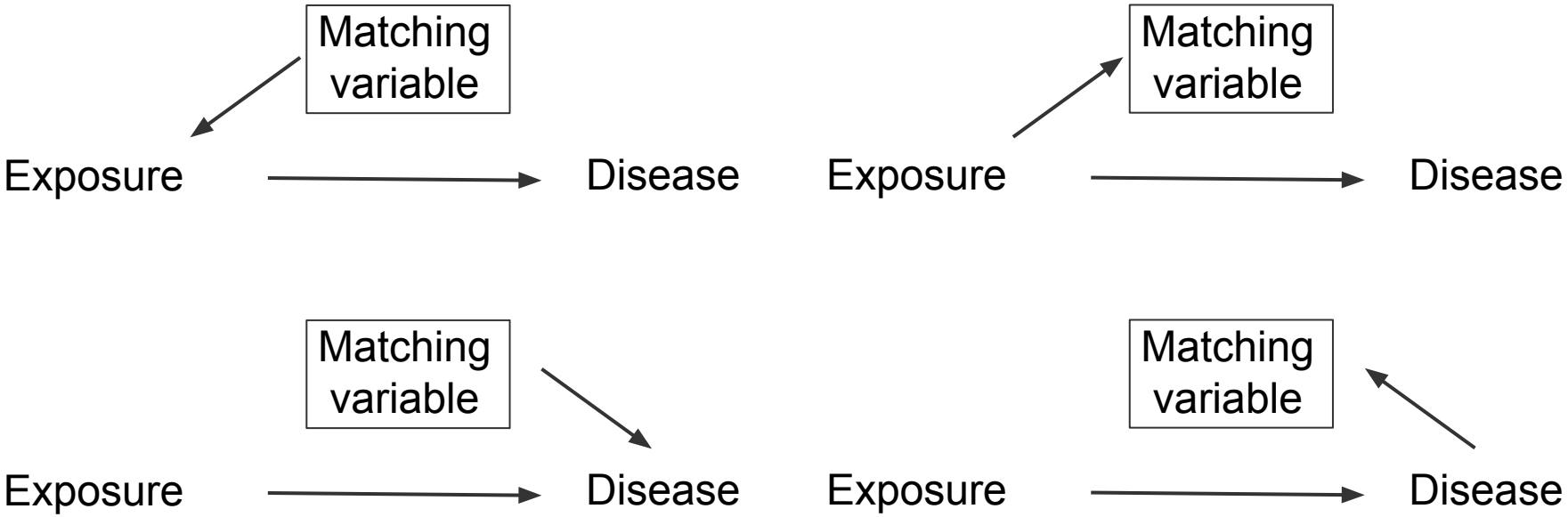
Summary of key points

DAGs in which matching introduces bias:



Summary of key points

DAGs in which matching reduces statistical efficiency:



Berkeley School of Public Health

Analysis of matched data

PHW250B

Overview of matching topics

- What is matching? Why match?
- Matching in different study designs
- Types of matching
- Matching in each study design
- Disadvantages of matching
- Numerical examples of matching
- Overmatching
- Analysis of matched data } **This video**

Analysis of matched data topics

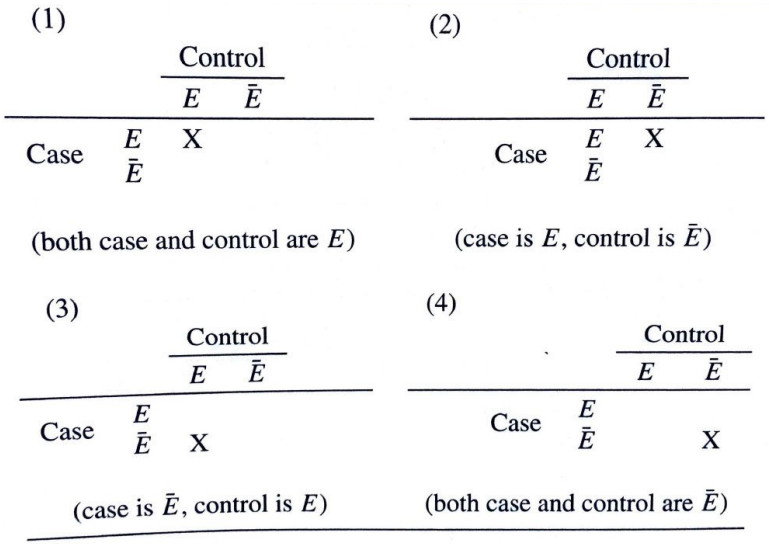
- Analyzing **frequency matched data**
  - Can use typical methods for stratification
    - Usual OR and RR formulas
- Analyzing **pair matched data**
  - Must use special methods that account for matching
    - Alternative OR and RR formulas
- **Breaking the match** (ignoring the matching in the analysis)
  - Cohort studies
  - Case-control studies

Jewell. *Statistics for Epidemiology.* 2004.

Analyzing pair matched data

This example focuses on case-control data, but these points also apply to cohort data.

- In pair-matched studies, there are only 2 observations in each stratum of the matched factor — a case and a control.
  - Type 1: Both case and control are exposed
  - Type 2: Case is exposed, control is unexposed
  - Type 3: Case is unexposed, control is exposed
  - Type 4: Both case and control are unexposed
- We can use this classification to summarize pair matched data.



Analyzing pair matched data

- The lower table shows a new type of 2x2 table with the number of exposed vs. unexposed cases on the left and the same layout for controls at the top.
- The counts A, B, C, and D are equivalent to the counts of each “type” of pair.
- N is the number of pairs.
- The total number of people sampled = 2*N

(1)

		Control	
		E	$\bar{E}$
Case	E	X	
	$\bar{E}$		

(both case and control are E)

(2)

		Control	
		E	$\bar{E}$
Case	E	X	
	$\bar{E}$		

(case is E, control is $\bar{E}$)

(3)

		Control	
		E	$\bar{E}$
Case	E		
	$\bar{E}$	X	

(case is $\bar{E}$, control is E)

(4)

		Control	
		E	$\bar{E}$
Case	E		
	$\bar{E}$		X

(both case and control are $\bar{E}$)

Organization of matched pair data

		Control	
		E	$\bar{E}$
Case	E	A	B
	$\bar{E}$	C	D

N

Jewell. *Statistics for Epidemiology.* 2004.

Odds ratio formula for matched pair data - case-control study

- When case and control both either exposed or unexposed we get no information about the exposure disease relation.
- The only information is in the discordant pairs.
- **Odds Ratio = B / C**
  - (See Jewell pg 261 for the derivation)
- **Intuition**: B and C are the pairs in which we have variation in the exposure – if no variation in exposure, cannot look at relation between exposure and disease

Concordant pairs
Discordant pairs

Table 16.3 *Exposure patterns in the four types of matched pairs*

(1)

	D	$\bar{D}$	
E	1	1	2
$\bar{E}$	0	0	0
	1	1	

(2)

	D	$\bar{D}$	
E	1	0	1
$\bar{E}$	0	1	1
	1	1	

(3)

	D	$\bar{D}$	
E	0	1	1
$\bar{E}$	1	0	1
	1	1	

(4)

	D	$\bar{D}$	
E	0	0	0
$\bar{E}$	1	1	2
	1	1	

Organization of matched pair data

		Control	
		E	$\bar{E}$
Case	E	A	B
	$\bar{E}$	C	D

N

Example

- Study of whether spontaneous abortion history is related to coronary heart disease, possibly due to endocrine effects.
- Matched case-control study
  - **Cases**: Had coronary heart disease (CHD)
  - **Controls**: Did not have CHD
  - **Exposure**: At least one spontaneous abortion
  - **Matching factors**: age and location of residence

Jewell. *Statistics for Epidemiology.* 2004.

Example

		Control		
		≥1 SA	No SA	
Case	≥1 SA	7	18	
	No SA	5	20	
				50

- "SA": spontaneous abortions
- Odds Ratio = B/C = 18/5 = 3.6
- This finding suggests that a history of at least one spontaneous abortion increases the odds of coronary heart disease.

Breaking the match - case-control studies

- In a case-control study, breaking the matching **will produce a biased estimate** of the measure of association. The direction of this bias will be toward the null value (OR = 1.0)
  - (Except in a special case when the population OR = 1.0 or when exposure probabilities are constant for all cases and all controls.)
  - Since these special cases are rare, we strongly advise against breaking the match in case-control studies.
- However, pair matching can legitimately be converted to frequency matching as long as all pairs do not have completely unique sets of matched factors
  - Example: individual matching only on sex can easily be converted to frequency matching on sex – then sex has to be controlled in the analysis

Jewell. *Statistics for Epidemiology.* 2004.

Risk ratio formula for matched pair data - cohort study

Note the different 2x2 table layout than for case-control studies.

Risk Ratio:

$$\frac{\text{Proportion of exposed with disease}}{\text{Proportion of unexposed with disease}}$$

$$= \frac{(A + B) / (A + B + C + D)}{(A + C) / (A + B + C + D)}$$

$$= \mathbf{(A + B) / (A + C)}$$

		Unexposed	
		Disease	No disease
Exposed	Disease	A	B
	No disease	C	D

Breaking the match - cohort studies

- Sometimes we conduct a matched study and then are tempted to ignore the matching during statistical analyses.
- In a cohort study, breaking the matching does not prevent us from estimating **a valid estimate** of the measure of association.
    - However, the variance estimates for the measure of association will be incorrect if they ignore the matching.
    - If there is differential misclassification of the outcome, we need to control for the matched factor to obtain a valid effect estimate.

Jewell. *Statistics for Epidemiology.* 2004.

Summary of key points

- Analyzing **frequency matched data**
  - Can use typical methods for stratification
    - Usual OR and RR formulas
- Analyzing **pair matched data**
  - Must use special methods that account for matching
    - Alternative OR and RR formulas
- **Breaking the match** (ignoring the matching in the analysis)
  - Cohort studies - can still obtain valid estimate
  - Case-control studies - produces a biased estimate!

Berkeley School of Public Health

Screening measures in depth

PHW250B

Screening measures

- Sensitivity
- Specificity
- Positive predictive value
  - (also called Predictive value positive)
- Negative predictive value
  - (also called Predictive value negative)
- Diagnostic accuracy
- Likelihood ratios
- ROC Curves

Introduced in a prior video. In this video, understand the relationships between these measures and how they are affected by prevalence.

Introduced in this video

Recap: sensitivity and specificity

TABLE 8-4 Schematic representation of the calculation of sensitivity and specificity for a binary variable.

Study's result	Gold standard's result		Total
	Positive	Negative	
Positive	a	b	$a + b$
Negative	c	d	$c + d$
Total	$a + c$	$b + d$	N

$$\text{Sensitivity} = a/(a + c)$$
$$\text{Specificity} = d/(b + d)$$

Sensitivity measures the probability of a **true positive**.

Specificity measures the probability of a **true negative**.

Szklo & Nieto, 3rd Ed.

Relationship between sensitivity and specificity

- When a binary cutoff is used to classify a disease as present/absent based on a continuous measure of disease:
  - Increasing sensitivity decreases specificity.
  - Increasing specificity decreases sensitivity.
- Example: abnormal serum cholesterol is classified as serum cholesterol ≥ 200 mg/dL
- How does changing this cutoff affect sensitivity and specificity?

TABLE 8-5 Comparison of screening values of serum cholesterol under field conditions and values done in a standard laboratory.

Screening values	Standard laboratory values		Total
	Abnormal*	Normal	
Abnormal*	18	19	37
Normal	1	11	12
Total	19	30	49

Sensitivity = 18/19 = 0.95
Specificity = 11/30 = 0.37

Szklo & Nieto, 3rd Ed.

Relationship between sensitivity and specificity

- Example: change “abnormal” cutoff from serum cholesterol ≥ 200 mg/dL to ≥ 300 mg/dL
- How does changing this cutoff affect sensitivity and specificity?

TABLE 8-5 Comparison of screening values of serum cholesterol under field conditions and values done in a standard laboratory.

Screening values	Standard laboratory values		Total	
	Abnormal*	Normal		
Abnormal*	10	1	37	11
Normal	9	29	12	38
Total	19	30	49	

Using ≥ 200 mg/dL
Sensitivity: 18/19 = 0.95
Specificity: 11/30 = 0.37

Using ≥ 300 mg/dL
Sensitivity: 10/19 = 0.53
Specificity: 29/30 = 0.97

When it makes sense to increase sensitivity at the expense of specificity

- You want to minimize false negatives and you are less concerned about false positives
- Example: Pap smears, which are used to screen for cervical cancer
- Want to detect all possible cases of cervical cancer since it is a serious but treatable disease.
- If the test produces a false positive, the patient will undergo additional testing to confirm they do not have cervical cancer.

Sensitivity and specificity depend on the distribution of severity of a disease

- When using one test and the same cutoff point, the amount of potential misclassification in a study is larger if the distribution of values is closer to a true negative.

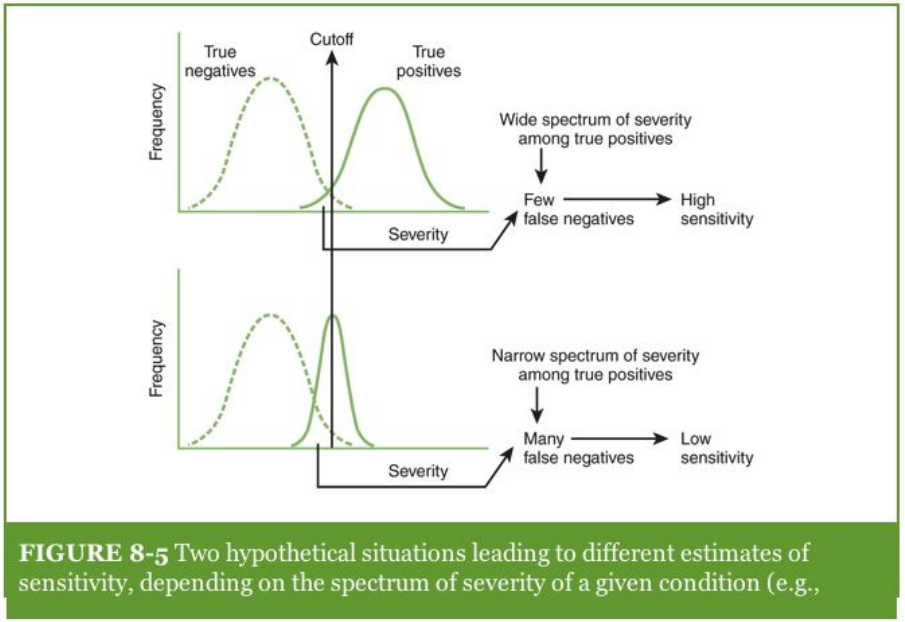


FIGURE 8-5 Two hypothetical situations leading to different estimates of sensitivity, depending on the spectrum of severity of a given condition (e.g.,

Szklo & Nieto, 3rd Ed.

Sensitivity and specificity estimated with self-reported data

- When disease status is measured through questionnaires, individuals' characteristics may affect their disease classification, introducing bias.
- **Example**:
  - A study of self-reported weight and height found that the sensitivity and specificity of BMI classification varied substantially by age and gender.
- When this is true, external validity of sensitivity and specificity is reduced.

Szklo & Nieto, 3rd Ed.

Recap: PPV and NPV

TABLE 8-4 Schematic representation of the calculation of sensitivity and specificity for a binary variable.

Study's result	Gold standard's result		Total
	Positive	Negative	
Positive	*a*	*b*	*a* + *b*
Negative	*c*	*d*	*c* + *d*
Total	*a* + *c*	*b* + *d*	*N*

- **Positive predictive value:** proportion of individuals with a positive test who have preclinical disease
  - PPV = (a / a + b) x 100% = 18 / 37 = 49%
- **Negative predictive value:** proportion of individuals without preclinical disease who test negative
  - NPV = (d / c + d) x 100% = 11 / 12 = 92%

Szklo & Nieto, 3rd Ed.

PPV and NPV are affected by prevalence. Sensitivity and specificity are not.

TABLE 8-4 Schematic representation of the calculation of sensitivity and specificity for a binary variable.

Study's result	Gold standard's result		Total
	Positive	Negative	
Positive	a	b	$a + b$
Negative	c	d	$c + d$
Total	$a + c$	$b + d$	N

Prevalence of disease
= a + c / N

Sens = a/a+c

- PPV and NPV condition on test results.
- The proportion who test positive depends on how prevalent the disease is.
  - **a** and **c** will be larger if the prevalence is high
  - **b** and **d** will be larger if the prevalence is low
- Prevalence doesn't affect sensitivity and specificity because they condition on true results.

Mathematical relationship between PPV and prevalence

$$PPV = \frac{a}{a+b}$$

1) Multiply by a+b+c+d/(a+b+c+d) and rearrange
2) Multiply by (a+c)/(a+c) and rearrange
3) Multiply by (b+d)/(b+d) and rearrange

$$= \frac{\textcircled{1}\ \frac{a}{a+b+c+d}}{\frac{a}{a+b+c+d} + \frac{b}{a+b+c+d}}$$

$$= \frac{\textcircled{2}\ \frac{a}{a+c} \times \frac{a+c}{a+b+c+d}}{\textcircled{2}\left(\frac{a}{a+c} \times \frac{a+c}{a+b+c+d}\right) + \left(\frac{b}{b+d} \times \frac{b+d}{a+b+c+d}\right)\textcircled{3}}$$

Sensitivity Prevalence

Sensitivity Prevalence 1 - Specificity 1 - Prevalence

Mathematical relationship between NPV and prevalence

$$NPV = \frac{d}{c+d}$$

1) Multiply by a+b+c+d/(a+b+c+d) and rearrange
2) Multiply by (b+d)/(b+d) and rearrange
3) Multiply by (a+c)/(a+c) and rearrange

$$= \frac{\frac{d}{a+b+c+d}}{\frac{d}{a+b+c+d} + \frac{c}{a+b+c+d}} \quad (1)$$

$$= \frac{\frac{d}{b+d} \times \frac{b+d}{a+b+c+d}}{\left(\frac{d}{b+d} \times \frac{b+d}{a+b+c+d}\right) + \left(\frac{c}{a+c} \times \frac{a+c}{a+b+c+d}\right)} \quad (2)\ (3)$$

(2) Specificity: $\frac{d}{b+d}$; 1 - Prevalence: $\frac{b+d}{a+b+c+d}$

(2) Specificity: $\frac{d}{b+d}$; 1 - Prevalence: $\frac{b+d}{a+b+c+d}$; (3) 1 - Sensitivity: $\frac{c}{a+c}$; Prevalence: $\frac{a+c}{a+b+c+d}$

Comparing screening measures

- **Sensitivity and specificity** assess how well a test classifies disease status compared to a gold standard.
- **PPV and NPV**
  - Population level: assess how well a test will perform in populations with different prevalence of disease.
  - Clinically: Answers the question “given that I tested positive for a disease, how likely is it that I truly have the disease?”

Diagnostic accuracy

TABLE 8-4 Schematic representation of the calculation of sensitivity and specificity for a binary variable.

Study's result	Gold standard's result		Total
	Positive	Negative	
Positive	a	b	$a + b$
Negative	c	d	$c + d$
Total	$a + c$	$b + d$	N

- The proportion of results that are correct:
  - a + d / (a + b + c + d)
- Measures overall accuracy combining sensitivity and specificity.

Likelihood ratios

TABLE 8-4 Schematic representation of the calculation of sensitivity and specificity for a binary variable.

Study's result	Gold standard's result		Total
	Positive	Negative	
Positive	a	b	$a+b$
Negative	c	d	$c+d$
Total	$a+c$	$b+d$	N

A likelihood ratio contrasts the proportions of patients with and without a disease for a given positivity criterion.

- Likelihood ratios can be used to connect pre-test and post-test probability of disease.
- **Positive likelihood ratio (LR^+)**: is the odds that a positive test result would be seen in a patient with the disease

$$LR^+ = \frac{sensitivity}{(1-specificity)} = \frac{A/(A+C)}{B/(B+D)}$$

can also be written as:

$$LR^+ = \frac{p(test+|D+)}{p(test+|D-)} = \frac{p(test+, D+)/p(D+)}{p(test+, D-)/p(D-)}$$

- **Negative likelihood ratio (LR^-):** the odds that a negative test result would be seen in a patient with the disease

$$LR^- = \frac{(1-sensitivity)}{specificity} = \frac{C/(A+C)}{D/(B+D)}$$

can also be written as:

$$LR^- = \frac{p(test-|D+)}{p(test-|D-)} = \frac{p(test-, D+)/p(D+)}{p(test-, D-)/p(D-)}$$

LRs, pre- and post-test probability of disease

- The pre-test odds of disease x LR^+ = post-test odds of disease
- Derivation:

$$\underbrace{\frac{P(D+)}{P(D-)}}_{\text{Pre-test odds of disease}} \times \underbrace{\frac{P(\text{Test}+, D+)/P(D+)}{P(\text{Test}+, D-)/P(D-)}}_{LR^+} = \frac{P(\text{Test}+, D+)}{P(\text{Test}+, D-)}$$

$$\underbrace{\frac{P(\text{Test}+|D+)}{P(\text{Test}+|D-)}}_{\text{Post-test odds of disease}} = \underbrace{\frac{P(\text{Test}+, D+)/P(\text{Test}+)}{P(\text{Test}+, D-)/P(\text{Test}+)}}_{\text{Convert conditional to joint probability}} = \underbrace{\frac{P(\text{Test}+, D+)}{P(\text{Test}+, D-)}}_{\text{Post-test odds of disease given that the test is positive}}$$

Example of clinical application

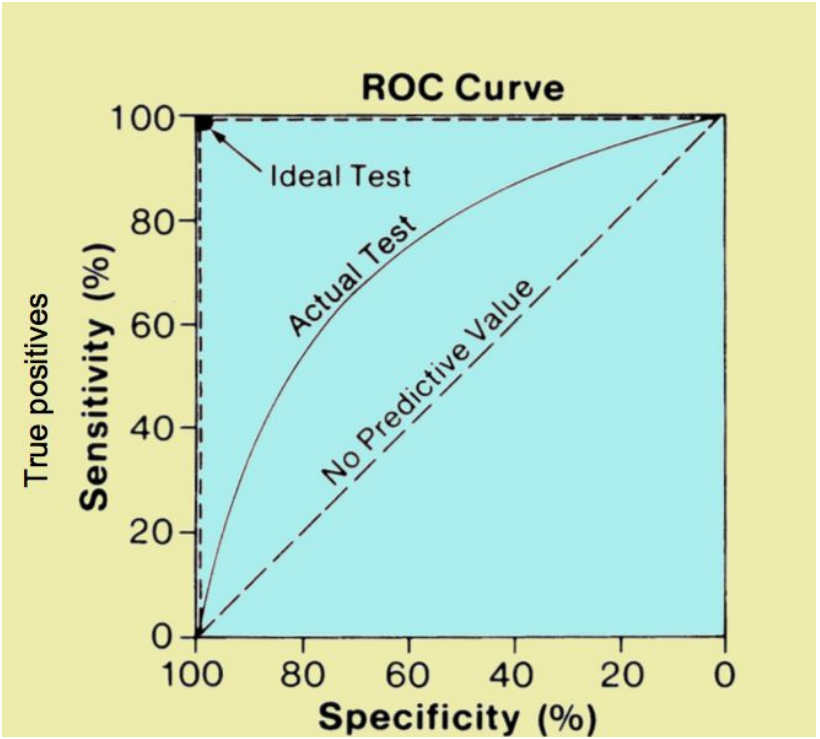
- You are working in an emergency room when a 50 year old woman presents with a complaint of “chest pain”. She is a smoker with a history of heart disease in her father.
- She does not have “classical” symptoms of a heart attack.
- You estimate, based on her story, that there is only a 10% chance that she having a heart attack.
- You decide to apply a new blood test to decide whether to admit her or not. The test has:
  - Sensitivity = 0.86
  - Specificity = 0.95
- What is the post-test probability that she is having a heart attack if the test result is 150 units/liter?

Example of clinical application

- **Pre-test odds of disease:** P(D+) / P(D-)
  - From prior slide, we estimate probability of disease is 0.1
  - Odds of disease = 0.1 / (1-0.1) = 0.11
- **Likelihood ratio + :** sensitivity / 1 - specificity
  - LR^+ = 0.86 / (1 - 0.95) = 17.2
- **Post-test odds of disease:** pre-test odds of disease x LR^+
  - Post-test odds = 0.11 x 17.2 = 1.89
- **Post-test odds of probability:** Probability = odds / (1+odds)
  - Post-test probability = 1.89 / (1+1.89) = 0.65
- We conclude that there is a 65% probability she is having a heart attack. Since this is higher than the pre-test probability, we choose to admit her to the hospital.

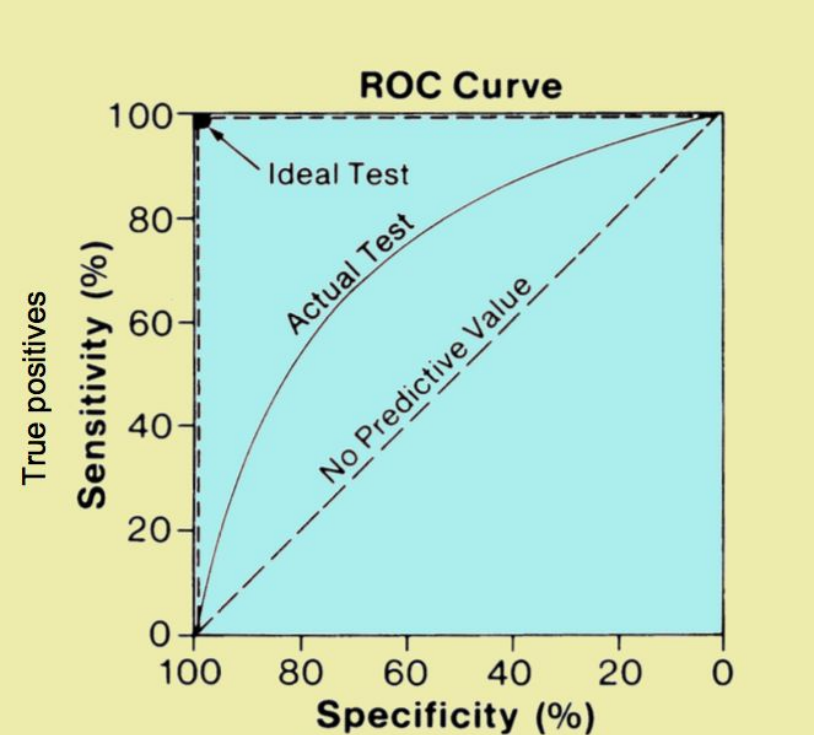
ROC curve

- Allow us to directly compare sensitivity and specificity when assessing how changing a test's cutoff (positivity criterion) affects sensitivity and specificity.
- Plots the sensitivity against specificity (note descending order of values for specificity on x-axis). Alternately, some texts plot the x-axis as (1-specificity) with values 0 to 100
- An ideal test has perfect sensitivity and specificity and has values in the upper left hand corner.

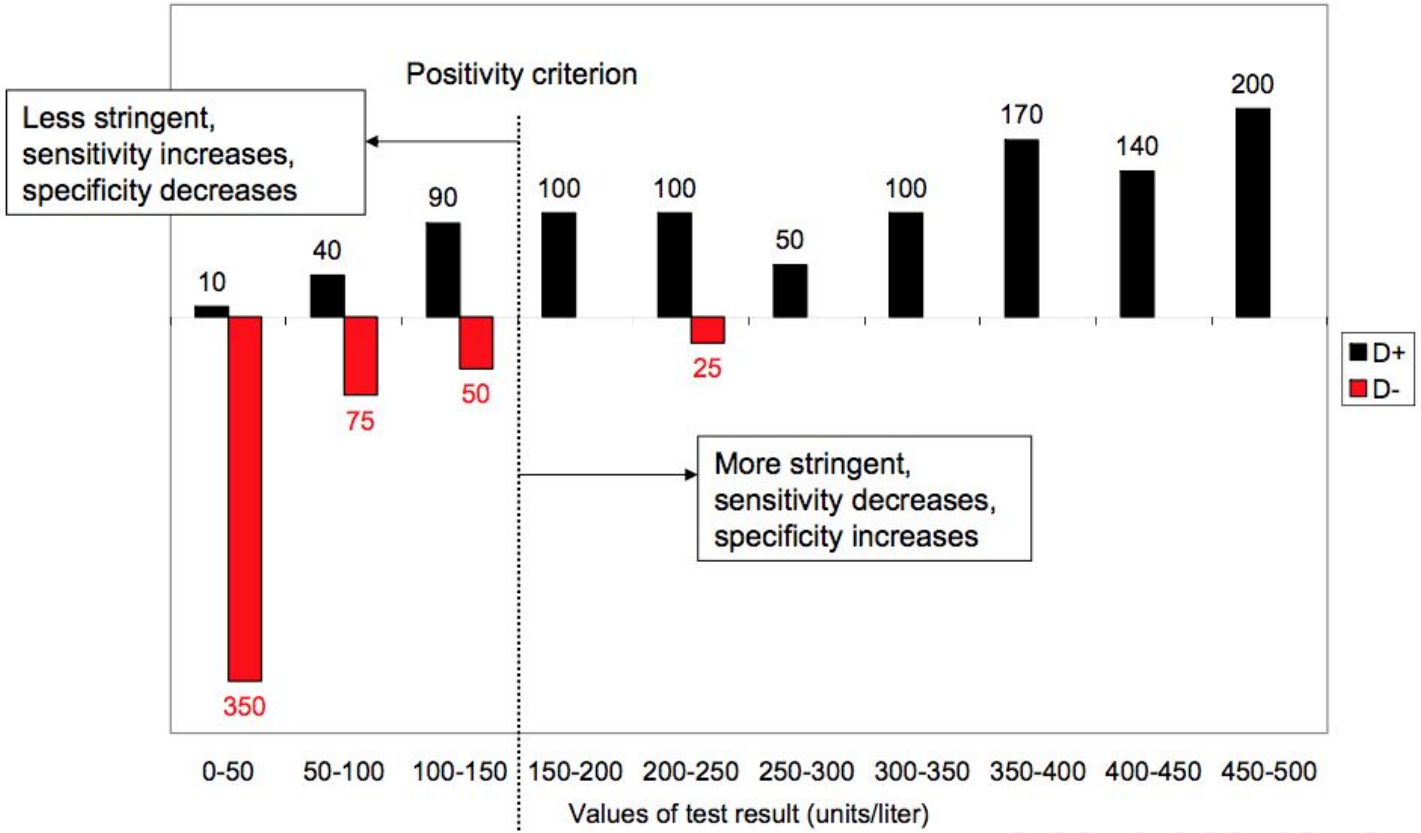


ROC curve

- As the positivity criterion for a test becomes more stringent (criterion has a larger value), the point on the curve corresponding to sensitivity and specificity moves down and to the left (lower sensitivity, higher specificity)
- As the criterion for a test becomes less stringent (criterion has a smaller value), the point on the curve corresponding to sensitivity and specificity moves up and to the right (higher sensitivity and lower specificity)



ROC curves and positivity criteria



Comparing two tests

- ROC curves are useful devices for comparing two or more screening tests
- Statistical procedures exist to allow determination of whether two ROC curves differ significantly from each other
- Usual method involves a determination of the area under the curve for each ROC curve and a modification of the Wilcoxon rank-sum procedure to compare them

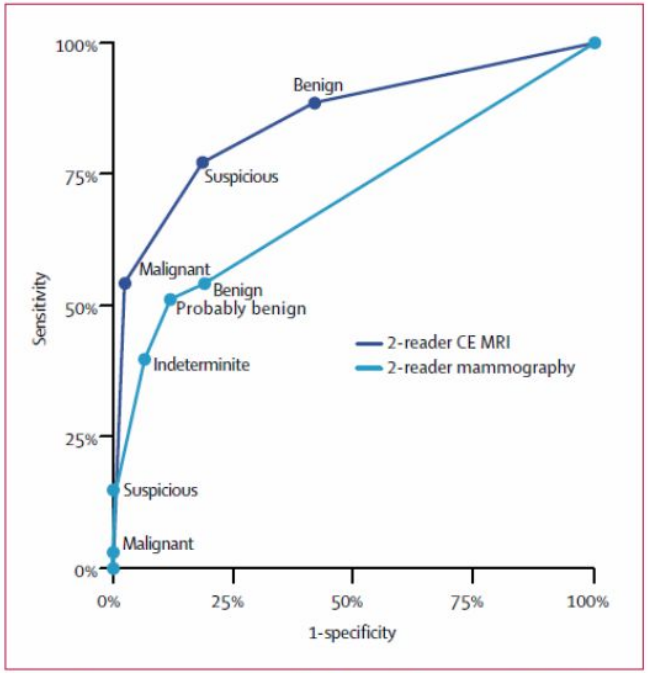


Figure 2: Receiver operator characteristic curves for two-reader CE MRI and mammography

Non-screening applications of these concepts

- So far this video has focused on using these measures to assess screening tools, but there are many other common applications of these concepts in epidemiology.
- **Diagnostic test accuracy**
  - Example: Compare microscopic examination of stool to a molecular assay for the detection of intestinal worm infections
- **Information bias**
  - Example: Compare the accuracy of self-reported vaccination data to medical records of vaccination

Summary of key points

- Sensitivity and specificity depend on the cutoff value for disease presence/absence when disease status is classified based on a continuous measure.
- Sensitivity and specificity depend on the distribution of severity of a disease
- PPV and NPV depend on the prevalence of disease in the population.
- Likelihood ratios can be used to connect pre-test and post-test probability of disease.
- ROC curves can be used to compare two tests or assess how sensitivity and specificity change when the positivity criterion changes.