Case-Control Study

Duanping Liao, MD, Ph.D

Email: Dliao@psu.edu

Phone: 531-4149

Cohort Study

Cumulative Incidence (CI) of a Disease is defined as the total number of NEWLY diagnosed cases of the disease in the study population who were disease-free at the baseline in a specific period of time. It is used as an estimate of likely a disease is to occur in the population over a certain period of time. Thus, the "risk" of suffering from a disease in the population. n / N (%).

Incident Density: Number of new cases **/** Total person-time at risk. (n / 1000 person-year).

Measures of association from a cohort study:

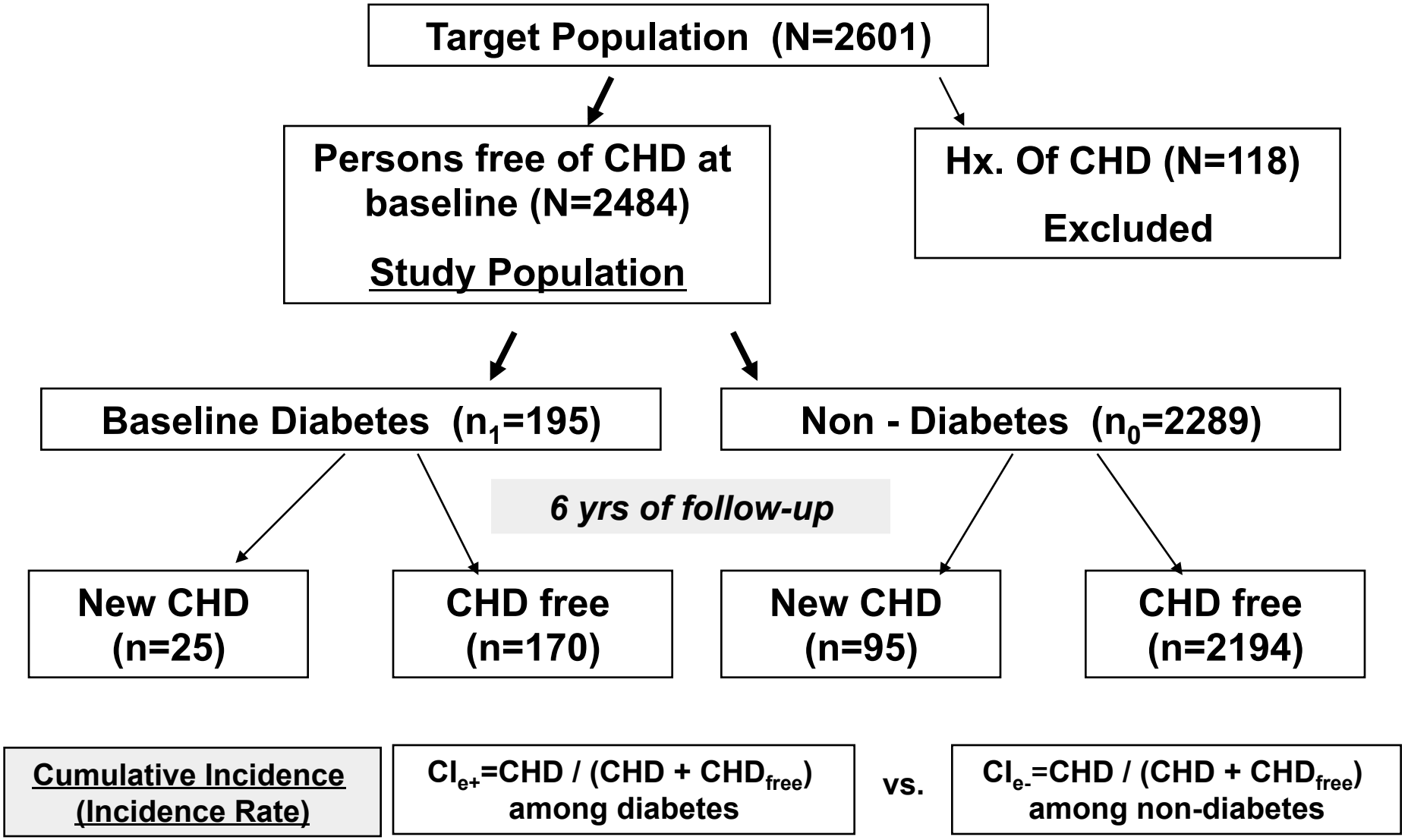
CI of a disease in population that are "exposed" to a risk factor (X%). vs. Incidence of a disease in population that are "not-exposed" to a risk factor (Y%).
X^2 test, with p value

Cumulative Incidence Ratio: with 95% Confidence interval and p-value

Incidence Density Ratio: with 95% Confidence interval and p-value

Odds ratio, with 95% Confidence interval and p-value

Example of a cohort study.
Study question: Is type 2 diabetes (exposure) associated with the development of coronary heart disease (outcome)?



Summary of data -- 2 x 2 Tables

	CHD +	CHD -
E +	**a**	**b**
E -	**c**	**d**

	CHD +	CHD -
DIAB +	**25**	**170**
DIAB -	**95**	**2194**

Cumulative incidences – count data : 12.8% vs. 4.2%
Odds of exposure - in incident CHD group: a/c = 25/95 vs. in CHD free group: b/d =170/2194

CIR: incidence in E+ / incidence in E- –- count data

$$\{a / (a + b)\} \; / \{c / (c + d)\} : \; 12.8 / 4.2 = 3.05$$

OR: Odds ratio of disease given exposure: (a/c) / (b/d) =

$$(a \times d) / (b \times c) = (25 \times 2194) / (170 \times 95) = 3.40$$

Case-Control Study

An epidemiological study in which a group of persons with the disease of interest (case group) and a group of persons similar to the case group but not having the disease (control group) are selected to compare the proportion of persons exposed to a risk factor of interest in order to elucidate the causal relationship of the risk factor of interest and the disease.

From "case series", personal experience, others colleagues, -- almost 99% of patients suffered from condition Y had a history /evidence of exposure to X.

What is the problem in this study?

-- It will never prove a causal relationship, lack of control group.

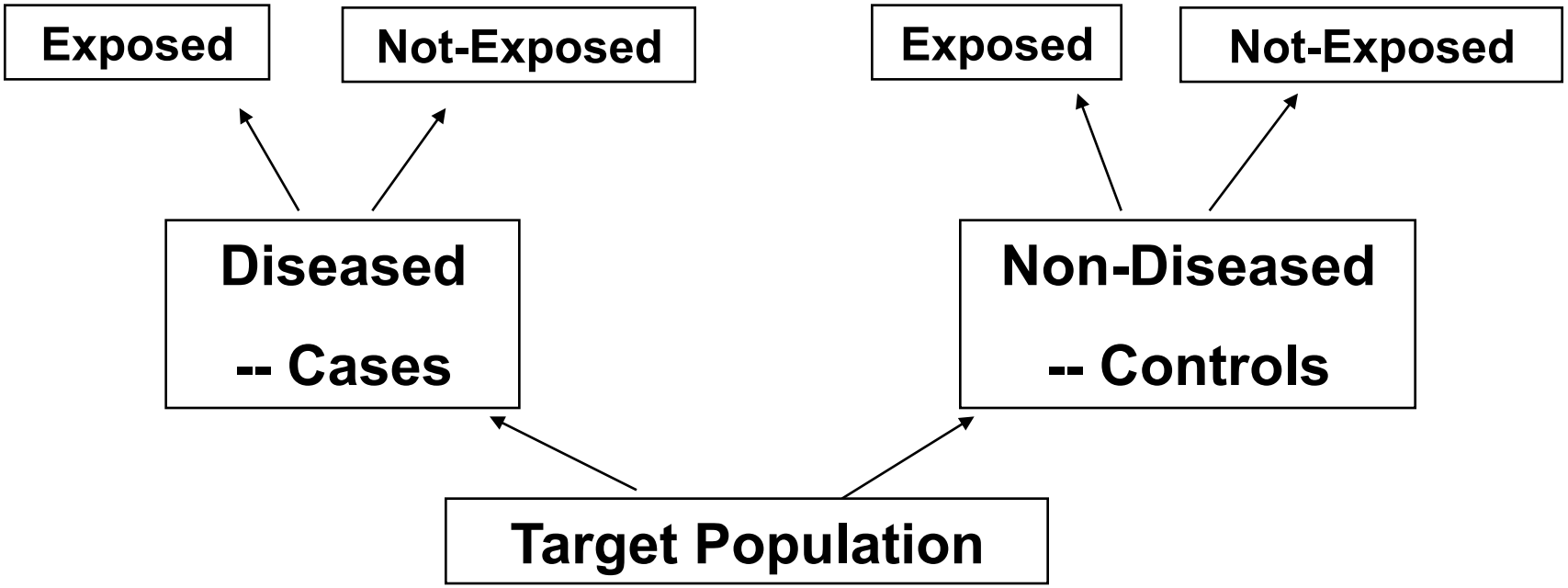
It will generate a hypothesis testable using epidemiological methods:

Persons with disease Y were more likely to have been exposed to factor X comparing to persons without the disease, but similar in other aspects.

Design a case-control study:

Case-control study contrast cases with controls for the exposure status.

		1st Step: Select cases and controls	
2nd step:	Exposure	Diseased (Cases)	Non-diseased (controls)
Measure past exposure and co-factors	Yes	a	b
	No	c	d
	Total	a + c	b + d
	Proportions E+	a / (a + c)	b / (b + d)



What we need to know about a case-control study

Because we started with cases and controls, we **cannot** estimate:

(1) the prevalence of disease in exposed and not-exposed.
(2) the incidences of disease in exposed and not-exposed.
(3) the relative risk to determine if there is an association between the exposure and the disease.

		1st Step: Select cases and controls	
2nd step:	Exposure	Diseased (Cases)	Non-diseased (controls)
Measure past exposure and co-factors	Yes	a	b
	No	c	d
	Total	a + c	b + d
	Proportions E+	a / (a + c)	b / (b + d)

Odds ratio is used as a measure of exposure-disease association in case-control study by asking slightly different questions:

(1) what are the odds that a case was exposed? =a:c or (a/c)
(2) what are the odds that a control was exposed? =b:d or (b/d)
(3) what is the odds ratio - the ratio of the odds that the cases were exposed to the odds that the controls were exposed. =(a/c)/(b/d)=(ad)/(bc)
(4) ***Odds ratio calculated differently for pair-matched case-control study***

What we need to know about a case-control study

Interpretation of Odds Ratio - same as relative risk:

OR = 1: exposure is not related to disease.

OR > 1: a positive association (E is associated with increased risk of D).

OR < 1: a negative association (E is associated with a lower risk of D).

Note: If follow-up is short, and outcome is rare, relative risk estimations from CIR, IDR, and OR are very close. Thus, logistic regression, Poisson regression, and Proportional Hazard model will produce similar estimates.

Case-control study: If the outcome is rare in the general population, OR from a case-control study is a good estimation of the true relative risk for the exposure and outcome relationship.

What we need to know about a case-control study

Odds Ratio is a good estimation of Relative Risk when:

- Cases are selected as representative sample of all people with such disease, regardless of the history of exposure.
- Controls are selected as a representative sample of all people without the disease in the population from which the cases were selected, regardless of the history of exposure.
- The disease is relatively rare.

What we need to know about a case-control study

Costello: Parkinson's Disease and Residential Exposure to Maneb and Paraquat From Agricultural Applications in the Central Valley of California. AJE. 2009;169:919–926.

Example of a case-control study:

		1st Step: Select cases and controls	
2nd step:	Hx of Exp.	PD Cases	Non-PD Controls
Measure past	Yes	173 - a	195 - b
exposure to	No	195 - c	146 - d
Cigarette	Total	368 - a + c	341 - b + d
	Proportions E+	47% - a / (a + c)	57% - b / (b + d)

- The PD incidence or prevalence ≠ 368 / (368 + 341), which is 52%. This parameter cannot be estimated from case-control study. Since they begin with cases, if they had selected 500 cases or 100 cases, the "prevalence" would have been 59% or 23% respectively. - The number of cases and controls are under control by the investigators.

- Odds ratio is used as a measure of exposure-disease association in case-control study by asking slightly different questions:
  (1) the odds that a case was exposed? =(a/c)=173/195
  (2) what are the odds that a control was exposed? =(b/d)=195/146
  (3) what is the odds ratio - the ratio of the odds that the cases were exposed to the odds that the controls were exposed.
  OR=(ad)/(bc)=(173*146)/(195*195) = 0.66

What we need to know about a case-control study

From Table 2 of the AJE Paper (Costello et al)

1974 – 1989 Exposed to both agents vs. not exposed

Odds ratio =(148*113)/(137*93)=1.31

		1st Step: Select cases and controls	
2nd step:	Exposure	Diseased (Cases)	Non-diseased (controls)
Past Exp to Either Pesticides	Yes	148	137
	No	93	113
	Total	241	250

Odds ratio =(74*113)/(39*93)=2.31

		1st Step: Select cases and controls	
2nd step:	Exposure	Diseased (Cases)	Non-diseased (controls)
Past Exp to both Pesticides	Yes	74	39
	No	93	113
	Total	167	152

What we need to know about a case-control study

Conclusion from this case-control study:

Smokers (Past or current) have a lower risk for PD than Never Smokers.

Persons exposed to both paraquat and maneb 1974-1989 had almost 2 and half times higher risk of having PD than those not exposed to neither agent.

What might be the problems with the above conclusions:

All of the above associations could have been introduced by our subject selection process, the problems of data collection, and/or the other factor(s) that are related to both exposure and the outcome under study. Epidemiological studies cannot derive a causal relationship before addressing these concerns -- **Bias and Confounding**.

The reason for the adjustment for multiple confounding factors in **Table 2.**

What we need to know about a case-control study

Bias - Any systematic error in the design / subject recruitment, data collection, and/or data analysis that results in a mistaken estimation of the true exposure and disease association.

Selection bias: Error due to systematic differences in characteristics between those selected and those not selected into a study, or systematic differences in which cases and control, exposed and non-exposed subjects are selected so that distorted association is observed. Distortion of association due to the selection of study population.

Oral Contraceptive (OC) users --> more Dr.'s attention --> more likely to be diagnosed with certain conditions, when compared to non-users. OC - Disease association can be overestimated.

Smoking and MI Relationship: If hospitalized cases of AMI are selected as cases, and heavy/long-time smokers are more likely to die outside of hospital (massive AMI), the smoking-AMI association would be underestimated.

Smoking and MI Relationship: If hospitalized AMI and hospitalized non-MI patients are selected as cases and controls, and smokers are more likely to be hospitalized for other conditions (pulmonary), the smoking-AMI association would be underestimated.

Impact of selection bias: Distortion of association, limited generalizability.

What we need to know about a case-control study

Selection bias in case-control study:

- Exposure has influence on the process of case assessment: the exposure prevalence in cases is biased.
- Exposure has influence on the process of control selection: the exposure prevalence in controls is biased, (e.g., use chronic bronchitis patients as controls for study of smoking and CHD, smoking and lung cancer.)
- Selective survival or selective migration: the exposure prevalence in prevalent cases may be biased compared to incident cases.

Case-control study is only valid:

- Cases are selected as representative sample of all people with such disease.
- Controls are selected as a representative sample of all people without the disease in the population from which the cases were selected.

What we need to know about a case-control study

Information bias (Misclassification bias): systematic error due to inaccurate measurement and / or classification of study variables. It can occur in the classification of outcomes, exposures, and co-variables. (also called information bias)

Differential misclassification: misclassification of variables differs in different study groups, particularly, conditional on case – control or exposure-nonexposure status.

Lung cancer cases are more likely to recall their smoking history (quantitatively and qualitatively). -- overestimation of smoking-cancer relationship.

Overweighed persons are more likely to have missing data in ultrasound examination of carotid arteries.

Females are more likely to under report of their weight, and males over report.

Impact of differential misclassification: distorts the association in often unpredictable direction. Most difficult to handle and should be minimized.

What we need to know about a case-control study

Non-differential misclassification: misclassification of some variables systematically, regardless of study groups.

Often a inherited problem due to the data collection methods / instruments. Such as BP under estimated by 5 mmHg in all subjects, both cases and control, males and females, and smokers and non-smokers.

Impact of non-differential misclassification: dilutes the association towards null (toward no association if the truth is a significant association).

Bias is often a result of an error in the design and conduct of a study, thus, efforts should be made to reduce or eliminate bias. At lease, it should be recognized and taken into account when making interpretation.

Bias can also be introduced when performing analysis on existing data, especially when the investigators set the exclusion / inclusion criteria to the data.

What we need to know about a case-control study

Confounder (confounding)

Confounding - A situation in which a measure of the association between exposure and outcome is distorted because of the relationship between the exposure and the other factor (the confounder) that influences the outcome under study.

Confounding is the most important concept.

It impacts the validity of an observation study.

A factor (Factor X) is a confounder only if it meets all of the three criteria:

1. It is an risk factor of the outcome under study, independent of the risk factor under study.
2. It is associated with the exposure under study.
3. It is not in the causal pathway between the exposure and the outcome. (it is not an intermediate link between the exposure and the outcome).

Account for Confounding: Confounded association is not acceptable!

In the design and conduct of a study:

1. Matching (individual or group) by potential confounders – in case-control studies.
2. Collect information on potential confounders.

In the analysis of a study:

1. Conceptualize your potential confounders
2. Stratification by potential confounders to identify major confounders
3. Stratification to derive stratum specific E-D association
4. Statistical adjustment of confounders

Confounding arises from unequal distribution of risk factor in the data.