P-VALUES AND CONFIDENCE INTERVALS; CONFOUNDING

SAPIENZA
UNIVERSITÀ DI ROMA

annarita.vestri@uniroma1.it

P-Values and Confidence Intervals

OVERVIEW

- Chance is always an explanation for our data, because we are trying to draw a conclusion about all people with an exposure and/or an outcome based on a sample.
- Chance or sampling variability must be taken into account when we describe our data, as well as when we make comparisons between groups.
- Overriding principle: size of the sample on which we are basing conclusions will play a major role in the likelihood of chance being an explanation for our findings.

- One common way to measure the effect of chance is by conducting a test of statistical significance.
- Set up a null hypothesis (Ho): nothing is going on, no difference, no association.
- Test the alternative hypothesis (H_1): something is happening, there is a difference, there is an association.
- Perform the appropriate test of statistical significance.

- All tests of significance lead to some measure of the effect of chance on the results of a study.
- One measure is the resultant p-value: the probability of obtaining a result as extreme as or more extreme than the actual sample value obtained given that the null hypothesis (H_0) is true.
- On basis of p-value ($p \leq 0.05$, $p > 0.05$), either will reject H_0 or will not reject H_0.

PROBLEM

- The p-value reflects both the size of the association and the sample size of the study (i.e., the variability)
- Even a small difference will achieve statistical significance (i.e., be judged unlikely to be due to chance) if the sample size is big enough
- Even a big difference will not achieve statistical significance (i.e., chance cannot be ruled out as a possible explanation) if the sample size is too small
- Problem is when you have a small to moderate-sized difference which is not statistically significant - can you conclude that nothing is going on (no effect) or is it that the sample size wasn't large enough to detect an effect that size statistically even if truly there

- **To separate out these two components of the p-value, the confidence interval should always be reported**

- **The range of values within which the true magnitude of effect (eg. RR or absolute difference) lies with a certain degree(eg. 95%) of confidence.**

- **The confidence interval can provide the information of the p-value, in deciding whether an association is statistically significant at a specified level.**

- **But far more importantly, the width of the confidence interval reflects the precision of the estimate, i.e., what the true value is likely to be.**

- **The interpretation of the confidence interval, then, will depend on the scientific question you are trying to address.**

EXAMPLE

Your relative is trying to decide whether to take postmenopausal hormones (PMH). You look up the literature and first tell her about the clear benefits of such therapy, include amelioration of postmenopausal symptoms and reduction in risk of osteoporosis. She then asks you about possible adverse effects of long-term use of PMH, and you inform her that there is clear evidence of a strong increased risk of endometrial cancer and a likely small increased risk of breast cancer. She has a great fear of breast cancer (given her numerous risk factors including late age at first birth, strong family history and personal history of benign breast disease), and wants to understand whether there truly is an increased risk; if so, how large; and how does this compare in magnitude to the increased risk of endometrial cancer.

PMH AND ENDOMETRIAL CANCER

STUDY 1: **RR = 7.5** **$p \leq 0.05$**

STUDY 2: **RR = 7.5** **$p \leq 0.05$**

PMH AND ENDOMETRIAL CANCER

STUDY 1: RR = 7.5 p ≤0.05 95% CI* = (1.1, 32.1)

STUDY 2: RR = 7.5 p ≤0.05

* Note: If p ≤0.05, then H_0 is rejected. This means that the null value (ex. RR=1) cannot be in the CI

PMH AND ENDOMETRIAL CANCER

STUDY 1: RR = 7.5 p ≤0.05 95% CI = (1.1, 32.1)

STUDY 2: RR = 7.5 p ≤0.05 95% CI = (7.2, 8.3)

PMH AND BREAST CANCER

STUDY 1: **RR = 1.13** **p >0.05**

STUDY 2: **RR = 1.13** **p >0.05**

PMH AND BREAST CANCER

STUDY 1: RR = 1.13 p >0.05 95% CI* = (0.2,13.0)

STUDY 2: RR = 1.13 p >0.05

* Note: If p >0.05, then H_0 cannot be rejected. Thus the null value (ex. RR=1) must be in CI.

PMH AND BREAST CANCER

STUDY 1: RR = 1.13 p >0.05 95% CI = (0.2,13.0)

STUDY 2: RR = 1.13 p >0.05 95% CI = (0.96,1.2)

PMH AND BREAST CANCER

STUDY 1:	**RR = 1.13**	**p >0.05**	**95% CI = (0.2,13.0)**
STUDY 2:	**RR = 1.13**	**p >0.05**	**95% CI = (0.96,1.2)**
STUDY 3:	**RR = 1.13**	**p $\leq$0.05**	**95% CI = (1.01,1.2)**

When is a confidence interval narrow enough? - this depends on the question being asked

PMH AND ENDOMETRIAL CANCER

STUDY 1:	**RR = 7.5**	**p ≤0.05**	**95% CI = (1.1, 32.1)**
STUDY 2:	**RR = 7.5**	**p ≤0.05**	**95% CI = (7.2, 8.3)**

1. **QUESTION:** **Is there something going on? Is the observed association unlikely to be due to chance?**

   ANSWER: **Both study 1 and 2 would tell you "yes". Both p ≤0.05, both informative about this question.**

2. **QUESTION:** **How sure are we about the precision of the observed magnitude of this association?**

   ANSWER: **From Study 1, not very sure (uninformative). From Study 2, precise estimate (informative).**

PMH AND BREAST CANCER

STUDY 1:	RR = 1.13	p >0.05	95% CI = (0.2,13.0)
STUDY 2:	RR = 1.13	p >0.05	95% CI = (0.96,1.2)

1. QUESTION: Is there something going on? Is the observed association unlikely to be due to chance?

   ANSWER: Both studies would tell you "no" - both are "null studies" - not statistically significant - chance cannot be ruled out as an explanation for the findings.

2. QUESTION: But does this mean there is truly no association between PMH and breast cancer or was the sample size too small to detect this effect even if present.

   ANSWER: Study 1 is an uninformative null result - you cannot distinguish between these two alternatives; Study 2 is an informative null (you can tell the magnitude of effect)

EVALUATION OF THE ROLE OF CHANCE INVOLVES 3 STEPS:

1. Estimation of magnitude of effect or association (ex. RR)

2. Hypothesis testing:association due to chance? Is this a reasonable alternative explanation?

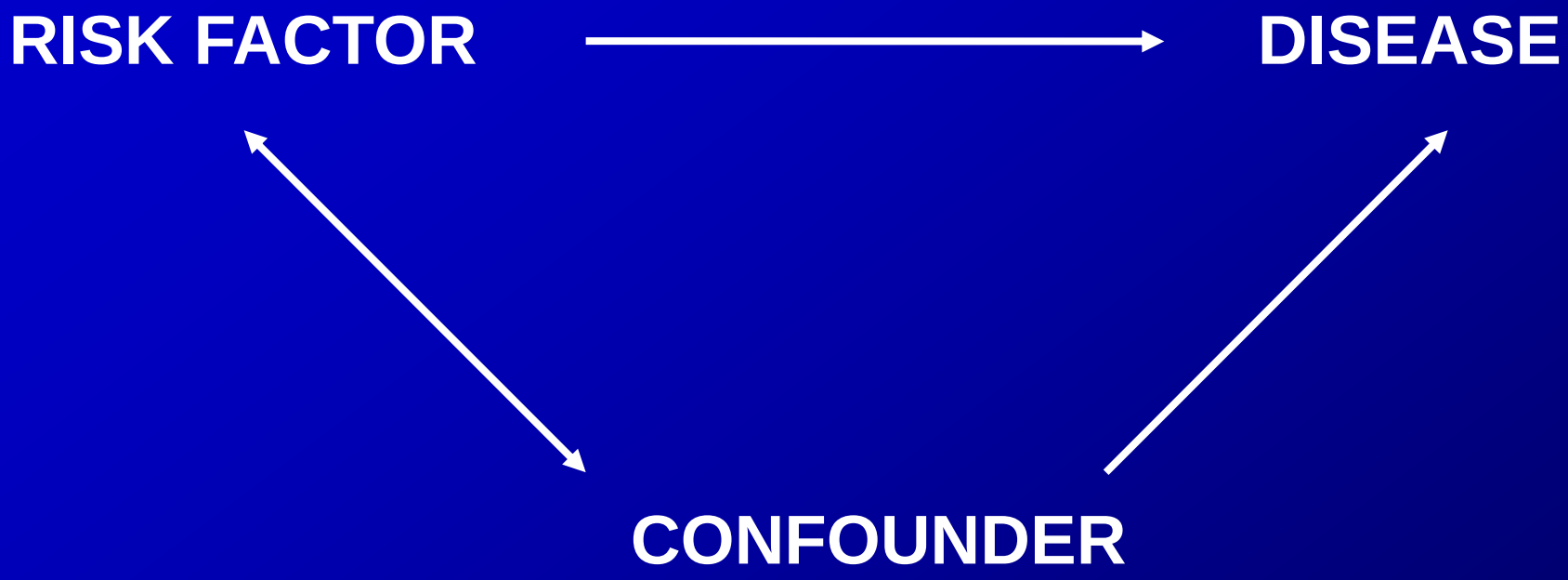
   p-value:probability that the observed association or one more extreme is due to chance alone, given that there is truly no association between the exposure and disease (i.e., H_0 is true)

3. Estimation of the precision of the effect measure, i.e., calculation of the confidence interval, or the range of values within which the true RR lies with a specified degree of confidence

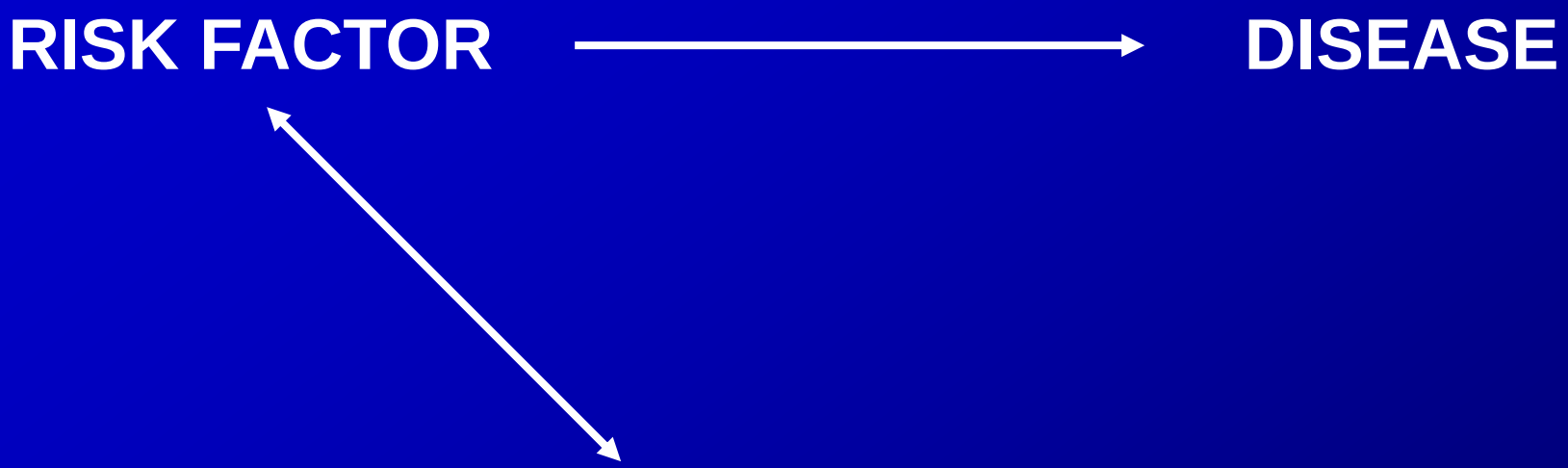
CONFOUNDING

CONFOUNDING

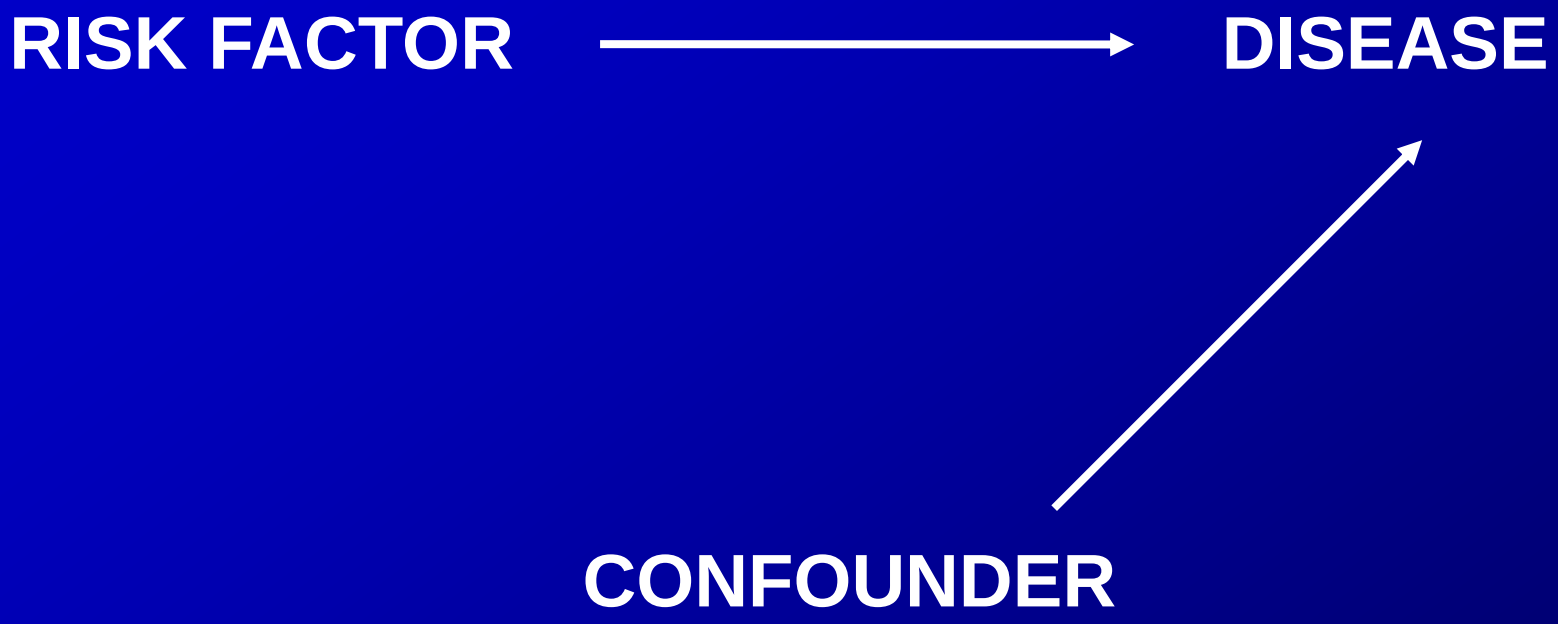
- A mixture of effects between the association under study and a third variable.
- This third factor (the confounder) must be BOTH associated with the exposure under study and, independently of the exposure, be a cause or correlate of the cause of the disease.
- The confounder may be responsible in part or totally for the association seen in the data.



NOTE: A confounder of an association in one population may not be so in another population.

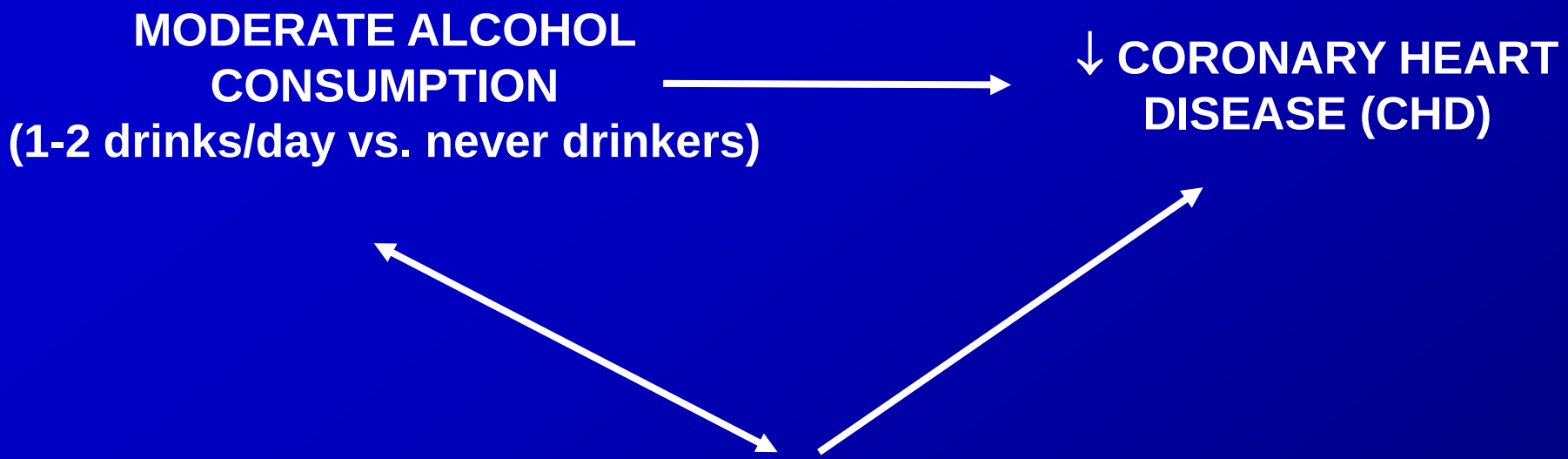


NOTE: If associated with exposure but not the disease, then not a confounder (eg., for smoking and lung cancer, alcohol is not a potential confounder because alcohol is not an independent risk factor for lung cancer).



NOTE: If associated with disease but not the exposure, then not a confounder (eg., moderate alcohol drinking and CHD; if exercise level not associated with alcohol drinking, not a confounder).

NOTE: This is an intermediate marker, not a confounder. If controlled for, then would be looking at relationship of risk factor and disease over and above the effect of this mechanism. (eg., obesity and coronary heart disease, diabetes not a confounder).



POTENTIAL CONFOUNDERS:

- Age (↓ age → ↓ CHD, overestimate benefit)
- Gender (male → ↑ CHD, underestimate benefit)
- Exercise (↑ exercise → ↓ CHD, overestimate benefit)
- Smoking (↑ smoking → ↑ CHD, underestimate benefit)

NOT POTENTIAL CONFOUNDERS:

- Eye color (not associated with CHD)
- HDL cholesterol level
  (intermediate factor, link in causal chain)

WHAT ARE POTENTIAL CONFOUNDERS?

- **Factors known to be related to the exposure of interest and disease of interest, but are not the mechanisms by which the exposure is postulated to act.**
- **If all this is unknown, then suspect all known risk factors for the disease to be potential confounders, and collect information on them in the design of the study.**

METHODS FOR CONTROLLING POTENTIAL CONFOUNDERS

1. **IN THE DESIGN OF THE STUDY:**
   - **Restriction: (restrict study subjects to one stratum of the confounding factor)**
   - **Matching: (match study groups so identical levels of the confounding factor)**
   - **Randomization: (if trial)**

2. IN THE ANALYSIS OF THE STUDY:

- **Matched analysis: (if matched in design)**
- **Stratification: (analyze association separately for each level of the confounding factor - special type: standardization)**
- **Multivariate analysis: (mathematical modeling to control for many confounders simultaneously)**

HOW DO YOU KNOW IF A POTENTIAL CONFOUNDER WAS A REAL CONFOUNDER?

- **Compare the overall (crude) RR and the adjusted RR's: the difference between these values is due to confounding. Report and use the adjusted RR.**

Example:

- **Low fat diet and CHD, RR = 0.60 compared to usual fat diet.**
- **Adjusted for BMI, RR of low fat diet vs. usual fat diet = 0.80.**

EFFECT MODIFICATION (INTERACTION)

When the magnitude of the relationship between the exposure and disease differs in size (is modified) by the level of a third variable (called the effect modifier).

- **EXAMPLE: Oral contraceptives and myocardial infarction in women of childbearing age.**

  Crude RR = 2.0

  RR (OC and MI, among nonsmokers) = 1.9
  RR (OC and MI, among smokers) = 41.0
  RR (OC and MI, among drinkers) = 2.0
  RR (OC and MI, among nondrinkers) = 2.0

- **Smoking modifies the effect of OC's on MI - the association is different in smokers and nonsmokers. Alcohol drinking is not an effect modifier of this association.**