

# 病例对照研究

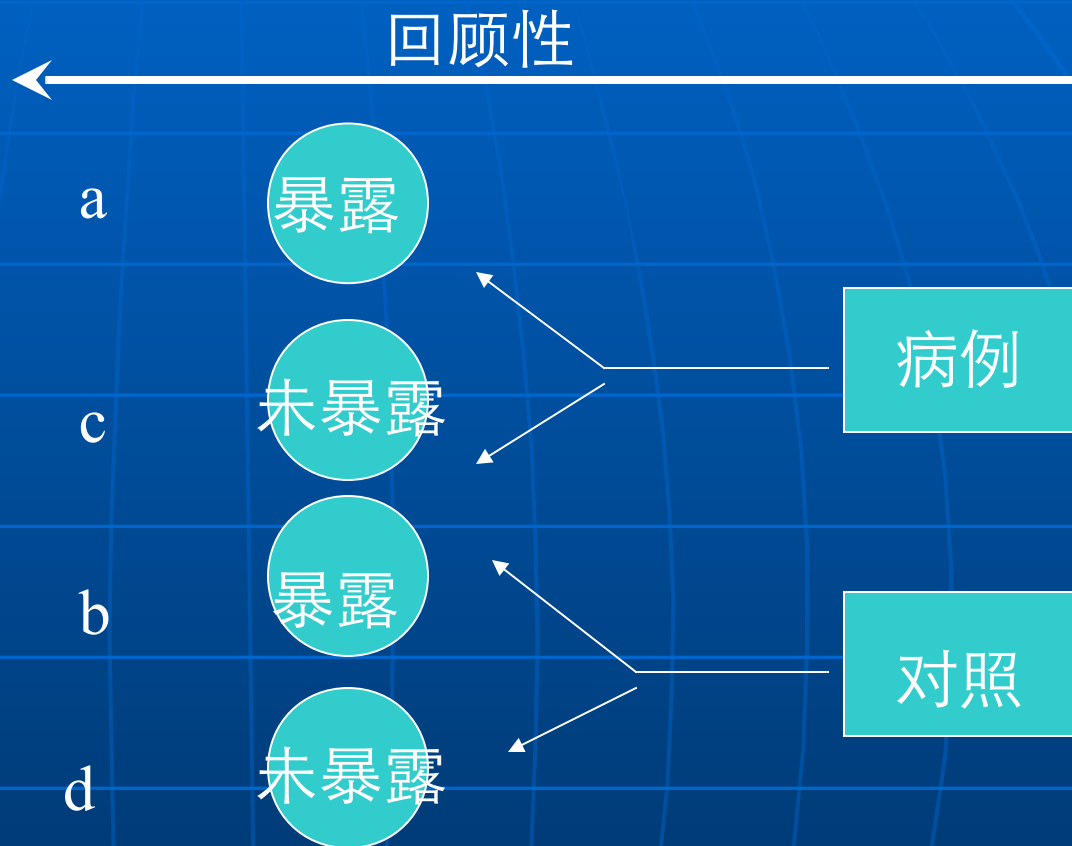
## case control study

# 第一节 概述

- 一、定义、基本原理及有关概念
- 二、用途
- 三、特点

# 一、定义、基本原理及有关概念

# 1. 基本原理



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

暴露因素与疾病无关

(2)  $a/a+c>b/b+d$

暴露因素为疾病危险因素

(3)  $a/a+c<b/b+d$

暴露因素为疾病保护因素

## 2. 定义:

选定患某病病例和未患该病的对照，分别调查其既往暴露于某个（或某些）危险因子的情况及程度，以判断暴露危险因子与某病有无关联及其关联程度大小的一种观察研究方法。

### 3. 有关概念

- (1) 暴露：指曾经接触过某种研究因素或具备某种特征。
- (2) 危险因素：能使人群发病率升高的因素。
- (3) 保护因素：能使人群发病率降低的因素。

## **二、用途：**

- 1. 广泛探索疾病的危险因素。**
- 2. 深入检验某个或某几个病因假说。**

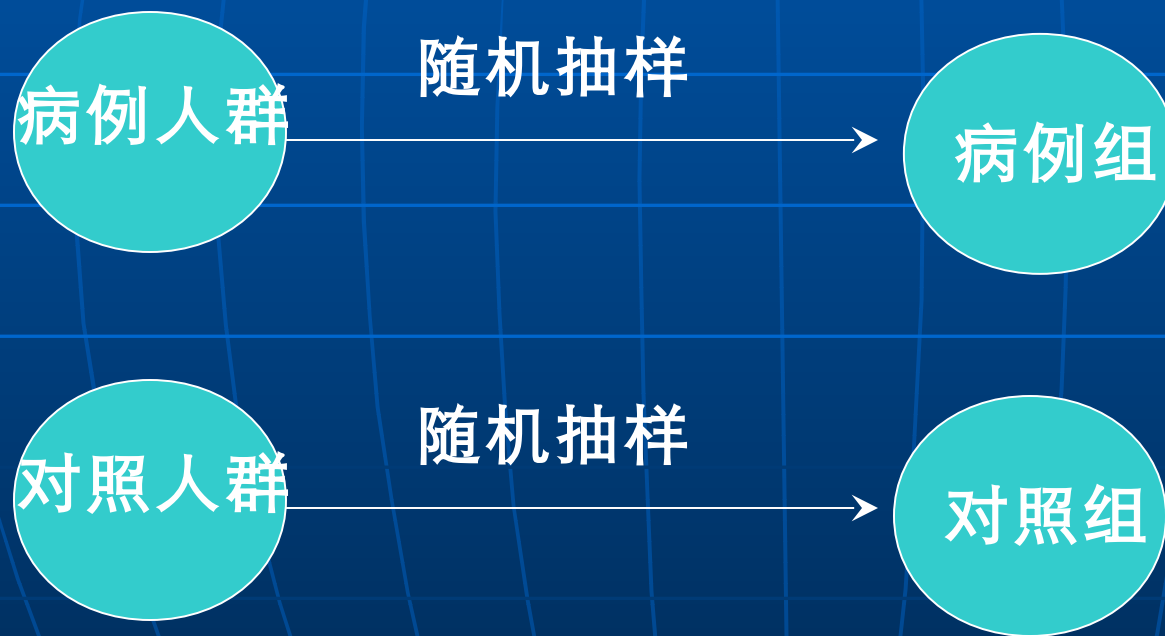
### 三、特点

1. 属于观察法
2. 设立对照组
3. 观察方向由“果”及“因”
4. 不能确实证明暴露与疾病的因果关系。
  -



# 第二节 病例对照研究的种类

## (一) 病例与对照不匹配的病例对照研究



要求对照人数应多于或等于病例人数。

## (二) 病例与对照匹配的病例对照研究

**匹配：**即要求对照组在某些因素或特性上与病例组保持相同。

**匹配的目的：**排除匹配因素对研究结果的干扰。

一般与疾病及所怀疑的病因因素都有联系的因素，可能是混杂因素，可选为匹配因素。

所选的匹配变量的作用不能评价，因此欲分析其作用的变量不能匹配，否则会导致匹配过头。

**匹配过头：**就是把非混杂因素而可能是危险因素的因素进行了匹配，导致该因素不能分析。

**匹配的优点**（1）增加流行病学效率。  
（2）排除非研究因素的影响。

**匹配的缺点：**  
（1）选择对象复杂。  
（2）需匹配因素多时，难以决定合适的匹配因素。  
（3）有时可导致匹配过头。

## 匹配的种类

**1. 频数匹配：**就是要求对照组匹配因素所

占比例与病例组一致。

**2. 个体匹配：**以每一病例为单位，选择在

某些特征或变量方面与病

例

一致的一个或几个对照，

组

# 第三节 病例对照研究的实施

## (一) 明确目的

## (二) 选择适宜的对照形式

1. 广泛地探索疾病的危险因子——不匹配，频数匹配。
2. 罕见病——一个体匹配。
3. 对照的性质——兄弟，姐妹，邻居等为对照——一个体匹配。

# **(三) 研究对象的选择：**

## **基本原则**

### **➤ 代表性**

**病例能代表总体的病例**

**对照能代表产生病例的总体人群或源人群**

### **➤ 可比性**

**两组主要特征方面无明显差异**

**1. 病例的选择：**病例是指患所研究疾病者。

(1) 诊断标准：应尽量采用国际和国内统一的  
标准。

(2) 病例来源：

1) 来源于医院：

2) 来源于社区

(3) 对病例其他特征的规定：

如性别、年龄、民族等。其目的是控制外部因素，增加可比性。

(4) 病例有三种：新发病例、现患病例和死亡  
病例。

**2. 对照的选择：**对照是指未患所研究疾病者

，

与

疾

可以是健康人，也可以是患

所研究疾病病因无关的其他

病的病人。

**(1) 对照应代表病例的源人群。**

**(2) 对照的来源：**

**1) 同一或多个医疗机构中诊断的其他病例；**

**2) 病例的邻居或所在同一居委会、住宅区中的健康人或非该病病例。**



### **(3) 选择对照应注意:**

- 1) 对照必须不患所调查的疾病，不是隐性病例，也不是处于潜伏期的病例。**
- 2) 与病例的病因因素无关或无联系；如肺癌，慢性支气管炎，肺结核均可能与吸烟有关，不能互为对照。**
- 3) 必须与病例选自同一人群，即在研究时他如患病，他将是被研究的病例。**
- 4) 选择的方法应事先规定，规定后要严格遵守，不可随意更改。**
- 5) 对照与病例组除要比较的因素外，其他各方面的特性应尽量一致，即所**

## (四) 样本含量的估计

### 1. 影响样本大小的因素

- (1). 研究因素在对照组中的暴露率  $P_0$  ;
- (2). 估计该因素引起的暴露的比值比 OR 。
- (3). 希望达到的检验显著性水平，即假设检验第 I 类错误的概率  $\alpha$  ;
- (4). 希望达到的检验把握度  $(1-\beta)$  ,  $\beta$  为统计学假设检验第 II 类错误的概率 。

## 2. 方法

❖ 查表法

❖ 公式法:

$$N = \frac{\left[ Z_{\alpha} \times \sqrt{2\bar{P}(1-\bar{P})} + Z_{\beta} \sqrt{P_1(1-P_1) + P_0(I-P_0)} \right]^2}{(P_1 - P_0)^2}$$

❖ 近似公式:

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2 \times \bar{P}(1-\bar{P})}{(P_1 - P_0)^2}$$

### 3. 注意事项

(1) 同时探索几个因素，而每个因素都有各自的

OR 及  $P_0$ ，这时估计样本大小常以最小的 OR

和最适合的  $P_0$  为准进行估计，以使所有的因

素都能获得较高的检验效率。

## (2) 样本含量估计中几个参数的相互关系

1) 对一定的  $\alpha$  ,  $\beta$  和  $P_0$  来说, 当 **OR** 增

大时, 所需样本含量减少

2) 对一定的  $\alpha$  、  $\beta$  和 **OR** 来说,  **$P_0$**  接近

**50 %**时, 所需样本含量最小。

3) 对一定的  $\beta$  、 **OR** 、  $P_0$  来说,  $\alpha$  增大

时, 样本含量减少

4) 对一定的  $\alpha$  、 **OR** 和  $P_0$  来说,  $\beta$  增

## **(五) 研究因素的选择:**

- 1. 变量的选择**
- 2. 变量的规定**

## **(六) 资料的收集**

**询问、查阅档案，采样化验，  
实地查看等**

## 第四节 资料整理与分析

### 一. 资料的整理

1. 原始资料的核查

2. 原始资料的分组、归纳或编码输入计算机。

## 二．数据的分析

### 1. 描述性统计

- ( 1 ) 描述研究对象的一般特征
- ( 2 ) 均衡性检验



## **2. 统计性推断**

### **(1) 不匹配（或成组）不分层资料的分析**

#### **1) 按每个因素将资料整理成四格表**

## 病例对照研究资料整理表

暴露史	病例	对照	合计
有	a	b	$a + b = n_1$
无	c	d	$c + d = n_0$
合计	$a + c = m_1$	$b + d = m_0$	$a + b + c + d = t$

## 2) 暴露率的假设检验:

$$\chi^2 = \frac{(ad - bc)^2 \cdot n}{(a + b)(c + d)(a + c)(b + d)}$$

## 3) 计算暴露与疾病的关联强度 相对危险度 (RR) :

RR > 1 :

RR < 1 :

RR = 1

---

RR		关联强度
0.9 – 1.0	1.0-1.1	无
0.7 – 0.8	1.2-1.4	弱
0.4 – 0.6	1.5-2.9	中等
0.1 – 0.3	3.0-9.9	强
<0.1	10-	很强

---

## 病例对照研究无法计算 **RR** 。

比值比（OR）：

$$OR = ad/bc$$

又称优势比，交叉乘积比，指病例组中暴露人数与非暴露人数的比值除以对照组中暴露人数与非暴露人数的比值

OR95% 可信区间：

OR 95%C.I.=

$$OR^{(1 \pm 1.96 / \sqrt{\chi^2})}$$

如果所研究疾病的发病率（死亡率）很低（5%）

$$OR \approx RR$$

4) 得出结论：有无联系，联系强度。

## 口服避孕药（OC）与心肌梗死关系的病例对照研究

暴露史	病例	对照	合计
有	39	24	63
无	114	154	268
合计	153	178	331

$X^2=7.70$       $p<0.01$

OR=2.20

OR 95% C.I.:1.26-3.84

## (2) 不匹配分层资料分析

层 — — — 组

分层的目的：排除混杂因素的干扰。

1) 不分层分析

2) 分层资料的整理：



暴露特征	i 层的发病情况		合 计
	病例	对照	
有	ai	bi	n1i
无	ci	di	n0i
合 计	m1i	m0i	ti

3) 计算各层 OR 值：若各层 OR 值无显著性差别，则计算  $OR_{MH}$ 。否则，不宜计算  $OR_{MH}$ ，应计算标准化死亡率或标准化率比。

4) 计算总的 OR 值： $OR_{MH}$ 。如 OR 与  $OR_{MH}$  有差别，则有混杂存在。如 OR 与  $OR_{MH}$  无差别，则无混杂存在。

5) 计算总的  $\chi^2$ ：

6) 得出结论：

	< 40 岁			≥40 岁		
	服 OC	未服 OC	合计	服 OC	未服 OC	合计
病例	21(a1)	26(b1)	47(m11)	18(a2)	88(b2)	106(m12)
对照	17(c1)	59(d1)	76(m01)	7(c2)	95(d2)	102(m02)
合计	38(n11)	85(n01)	123(t1)	25(n12)	183(n02)	208(t2)

$$OR_1 = 2.80$$

$$OR_2 = 2.78$$

$$OR_{MH} = 2.79 \quad OR_{MH} 95\% \text{ C.I.} : 3.18 - 2.22$$

$$X^2_{MH} = 11.79 \quad v = 4-1=3 \quad P < 0.01$$

### (3) 分级资料的分析

#### 1) 资料整理

	暴露分级						合计
	0	1	2	3	4	.....	
病例	a0(=c)	a1	a2	a3	a4	.....	m1
对照	b0(=d)	b1	b2	b3	b4	.....	m2
合计	n0	n1	n2	n3	n4	.....	t

#### 2) 卡方检验

3) 计算各级的 OR 值： 应进行趋势检验

#### 4) 得出结论

	暴露分级				
	0	1 –	5 –	15 –	合计
病例	2(c)	33 (a1)	250 (a2)	364 (a3)	649 (m1)
对照	27(d)	55 (b1)	293 (b2)	274 (b3)	649 (n2)
合计	29 (n0)	88 (n1)	543 (n2)	638 (n3)	1298 (t)

OR    1.0       8.1       11.52       17.93

$\chi^2=43.15$      $\nu = 3$      $P<0.001$

趋势检验：     $\chi^2=40.01$      $\nu = 1$      $P<0.01$

# ( 4 ) 个体匹配资料的分析

## 1) 将资料整理成四格表:

对照	病例		对子数
	有暴露史	无暴露史	
有暴露史	a	b	a+b
无暴露史	c	d	c+d
对子数	a+c	b+d	t=a+b+c+d

暴露史	病例	对照	
	+	+	
	+	-	c
	-	+	b
	-	-	d

2)  $\chi^2$  检验:

$$\chi^2 = \frac{(b-c)^2}{b+c} \text{ 或 } \chi^2 = \frac{(|b-c|-1)^2}{b+c}$$

3) 计算 OR 及 OR95% 可信区间:

$$OR = \frac{c}{b}$$

4) 得出结论:

## 外源性雌激素与子宫内膜癌的关系

对照	病例		对子数
	有暴露史	无暴露史	
有暴露史	27(a)	3(b)	30(a+b)
无暴露史	29(c)	4(d)	33(c+d)
对子数	56(a+c)	7(b+d)	63(t)

$\chi^2 = 19.53 \quad P < 0.005$

OR=9.67(2.31-40.41)



## (5) 病因分值 (EF) :

又称: 归因分值 (AF)、归因危险百分比 (ARP)

表示在一个有代表性且稳定的目标人群中,  
归因于某暴露因素的比例。

## 暴露人群的病因分值

(归因危险百分比) (EFe)

指暴露人群内，某种疾病的发病率中，  
归因于暴露引起的发病占全部发病的比例。  
。

$$EF_e = \frac{(I_e - I_0)}{I_e} \times 100\% \approx \frac{OR - 1}{OR} \times 100\%$$

## 人群病因分值 ( $EF_p$ )

( 人群归因危险度百分比 )

人群中某疾病的发病率中，归因于某种暴露引起的发病占全部发病的百分比。

$$AF_p = \frac{I_p - I_0}{I_p} = \frac{P_e(OR - 1)}{P_e(OR - 1) + 1}$$

# 第七节病例对照研究中的偏倚及其防止

(一) **选择性偏倚**：由于选入的研究对象与未选入的研究对象在某些特征上存在差异而应起的误差。

**1. 入院率偏倚**：**Berkon** 偏倚。由于患某病的人及具有某特征的人住院机会不一样，就可能使与疾病无关的某特征与疾病出现假关联。

**控制**：在多个医院选择病例，同时选择医院和社区的對象。

## 2. 现患病例 – 新发病例偏倚: Neyman bias

病例对照研究或横断面研究中，尤其是病死率高的疾病研究对象通常只是那些存活的病例，而对于那些已经死亡的病例通常很难进行调查。

这些幸存者，可能会改变他们原来的暴露状况，如改变其生活习惯。

这样在病例对照研究或横断面研究中选用的病例的暴露情况可能会与队列研究中不同。

由此造成的偏倚就是现患病例 – 新发病例偏倚。

**控制：** 选用新发病例

与队列研究结合起来进行。

**3. 检出症候偏倚：**病人因某些与致病无关的症状而就医，从而提高了早期病例的检出率，导致过高地估计了暴露程度，而产生的偏倚。

**控制：**选择中、晚期病例作为对象。

**4. 时间效应偏倚：** 将处于潜伏期或疾病早期的病人选为对照所造成的偏倚。

**控制：** 采用敏感的疾病早期检查技术；  
开展观察期充分长的纵向调查。

## **(二) 信息偏倚:**

**1. 回忆偏倚:** 由于被调查者记忆失真或不完整造成的偏倚。

**控制:** 采用客观的标准获得资料;  
用新发病例。

**2. 调查偏倚:**

**控制:** 改善调查方法或环境;  
对调查员进行培训;  
同一个时间, 同一调查员调查病例及其对照;  
校准仪器或试剂。



### (三) 混杂偏倚

当研究暴露于某因素与疾病的关系时，由于一个或多个既与疾病有制约关系，又与暴露密切相关的外来因素的影响，掩盖或夸大了所研究的暴露因素与疾病的联系，这种误差称为混杂偏倚或混杂。

#### 控制：

设计阶段：限制；匹配；随机分组。

分析阶段：分层分析；多因素数学模型。

# 第八节 病例对照研究的优缺点

## (一) 优点:

1. 特别适用于罕见病的研究。
2. 省时间、省人力、省经费，易于组织实施。
3. 一次可调查多个因素。

## **(二) 缺点:**

- 1. 不能判断因果联系。**
- 2. 不能计算发病率、相对危险度。**
- 3. 易产生偏倚：回忆偏倚、选择偏倚等。**

## 第九节 病例对照研究还能干什么？

1. 药物有害作用的研究。
2. 疾病预后研究。
3. 评价疾病预防与治疗措施的效果