

队列研究

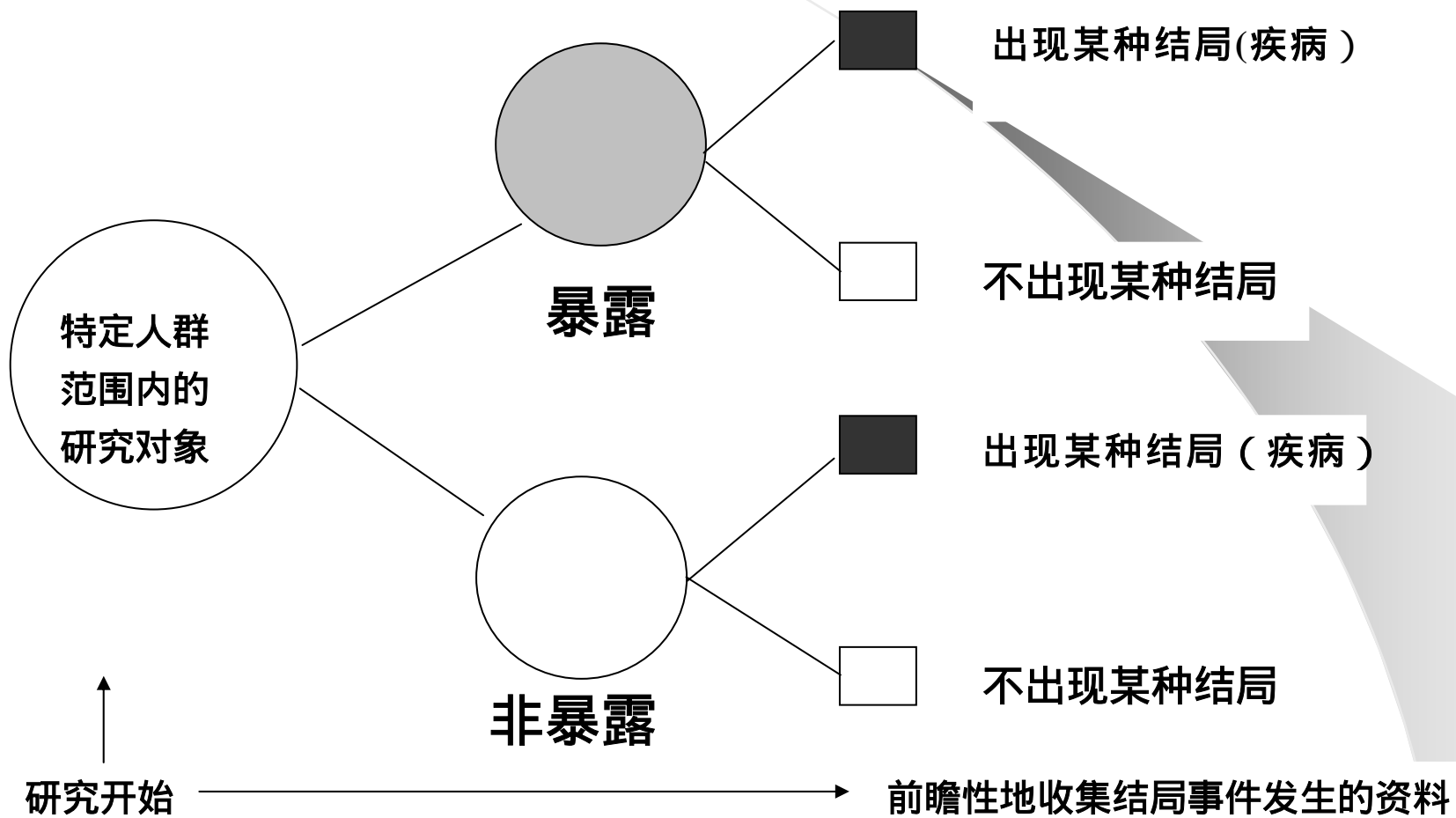
赵根明

复旦大学公共卫生学院

概念

将特定人群分为暴露于某因素与非暴露于某因素或不同暴露水平的亚组，追踪观察一定时间，比较两组或各组的发病率或死亡率，以检验该因素能影响发病的假设。

又称前瞻性研究、发病率研究、纵向研究、随访研究、定群研究。



队列研究的结构模式图

- 暴露的含义
- 队列研究从研究因素着手，追踪发病或死亡，从“因”到“果”。
- 目的是检验某因素能影响某疾病发生概率的假设。

分 类

根据研究开始时病例是否发生，主要分为：

- 前瞻性队列研究
- 历史性队列研究

前瞻性队列研究

- 研究开始时，暴露因素已存在，但疾病尚未发生
- 在随访过程中，可获取暴露因素变动情况
- 属于规模巨大的研究

降低前瞻性研究费用

- 利用现有的疾病登记系统收集病例
- 只收集暴露人群的资料
- 在随访的研究人群中，再进行巢式病例对照研究
- 选择疾病高发地区进行研究
- 考虑运用历史性队列研究

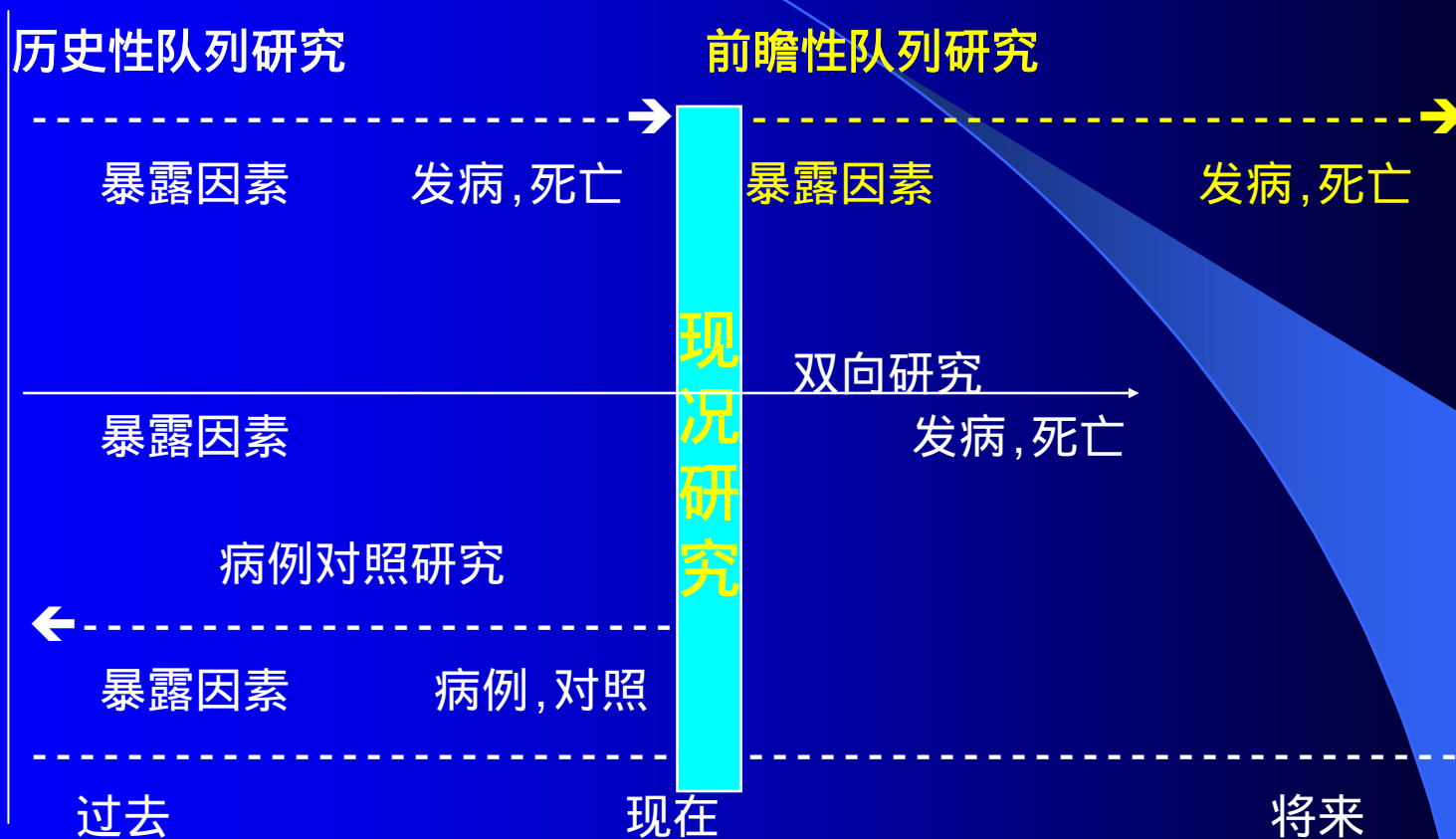
历史性队列研究

- 研究开始时，暴露与疾病均已发生
- 先追溯历史资料，研究暴露和非暴露组，然后通过各种途径查出研究对象的发病和死亡情况
- 历史资料的完整性和真实性将影响到研究的可行性和结果可靠性
- 缺乏混杂因素的资料，可能歪曲研究结果

巢式病例对照研究

首先进行队列研究，收集每个队列成员的暴露信息以及有关的混杂资料，确诊随访期间内发生的每一个病例，然后以队列中的病例为病例组，对照来自同一个队列，进行病例对照研究。

不同研究类型



队列研究典型实例

- 吸烟与肺癌
- 放射治疗与白血病

吸烟与肺癌

- Doll 和 Hill于1951开展的
 - 对象：在英国登记的所有开业医生
 - 调查：信访
 - 方式：调查表(是否吸烟，吸烟量；过去是否曾吸烟，现戒烟；从未吸烟)
 - 结果：总死亡率0.65‰(不吸0.07‰, 吸烟0.71‰)

放射治疗与白血病

- 白血病是真性红细胞增多症(PV)中的一种自然结局还是放射治疗的结果?
- 1965年开始研究
- 查阅1937-1955年的诊断记录(988例PV)
- 追溯患者到1961年12月31日(急性白血病)

放射治疗与白血病

治疗方法	病人数	病例数	发病率(%)
非放射	434	1	0.2
X 线	127	8	6.3
P^{32}	330	35	10.6
X 线+ P^{32}	97	14	14.4
合计	988	58	5.9

队列研究的步骤

- 设计前的准备
- 制定研究计划
- 研究暴露因素的决定、分级及其标准的规定
- 研究对象的选择、分组方法
- 检验两组对象在主要指标方面的均衡性
- 追踪观察
- 核对资料、整理统计与分析

暴露组

- 暴露组的选择
- 暴露因素的规定
- 暴露水平的分级
- 暴露因素的收集

暴露组的选择

- 选择暴露组应考虑的问题

已经处于暴露因素中、能提供可靠的暴露史、职业暴露

- 几种暴露组的人群

职业人群、特殊暴露人群、医疗保险人群、特定地区人群、孕妇

暴露因素的规定

- 必须有一个明确的规定
- 最好用定量或分级指标
- 能测定暴露剂量则更好

暴露水平的分级

可用定量资料分级，但由于不易获得准确的定量资料，故常用暴露水平分级。

如分为：严重暴露、中度暴露、
轻度暴露、不暴露

暴露资料的收集

- 有关记录
- 调查对象
- 医学检查
- 环境监测

非暴露组的选择

除未暴露于所研究的因素外，其它各种因素的影响或人群特征（年龄、性别、民族、职业、文化程度等）应尽可能相同。

非暴露组种类

- 内对照
- 与人群发病率或死亡率比较
- 设立比较组
- 多种对照

队列研究的样本含量

- 有关计算样本量的几个问题

抽样方法

暴露组与非暴露组的比例

失访率

样本大小

样本大小公式：

$$N = \frac{(Z_{\alpha} \sqrt{2pq} + Z_{\beta} \sqrt{p_1q_1 + p_0q_0})^2}{(p_1 - p_0)^2}$$

样本量计算实例

某队列研究欲观察放射线暴露与白血病的
关系。已知一般人群白血病的发病率为
1/万，放射线暴露者的发病率为10/
万，如果研究者将显著性水平定为0.05，
研究功效定为0.10，需多少人？

追踪观察

- 从常规登记中收集结果
- 定期对研究对象作健康检查

队列研究的资料分析

- 率
- 显著性检验
- 联系强度等

率的计算

累积发病率(cumulative incidence)

人群数量较多、人口比较稳定、资料比较整齐

队列研究累积发病率资料整理表

暴露史	病例	非病例	合计	发病率
有	a	b	$N_1=a+b$	a/N_1
无	c	d	$N_0=c+d$	c/N_0
合计	$M_1=a+c$	$M_0=b+d$	T	M_1/T

发病密度

队列研究发病密度资料整理表

暴露史	病例	人时数	发病率
有	a	N_1	a/N_1
无	c	N_0	c/N_0
合计	M_1	T	M_1/T

人时的计算

计算原则为：从观察对象中剔除死亡，迁移及失去联系的人数，补充新加入的人数折算人年。

常用方法：精确法、寿命表法、简化的近似法。

大样本人年的计算

以个人为单位计算暴露人年

应用寿命表法来计算暴露人年

男性各年龄级各年末存活人数

年龄组 (岁)	1980年 1月1日	1982年 1月1日	1983年 1月1日	1984年 1月1日	1985年 1月1日	1986年 1月1日	1987年 1月1日	合计 人年
<45	607	547	471	392	324	246	210 2,875.5
45-54	598	596	604	599	603	625	598 5,279
55-65	369	406	433	472	493	496	519 4303.5
65-	62	75	95	111	132	158	180 1,386
合计	1636	1624	1603	1574	1552	1525	15071,3844

计算年龄组死亡率时，以死亡时属该年龄组的对象为分子，以该年龄组合计暴露人年数为分母。

以个人为单位计算暴露人年

- 如样本不太大，可以将研究对象分别计算暴露人年。
- 方法较准确，但较费时。

寿命表法计算暴露人年

- $L_x = l_x + 1/2(N_x - D_x - W_x)$

联系强度的测量

- 相对危险度(relative risk , RR)

反映暴露与发病(或死亡)的关联强度的指标，是暴露组发病(死亡)率与非暴露组的发病(或死亡)率的比值。

$$RR = \frac{I_e}{I_0} = \frac{a/N_1}{c/N_0}$$

$$SE (LnCIR) = \sqrt{\frac{1}{a} - \frac{1}{N_1} + \frac{1}{b} - \frac{1}{N_0}}$$

$$95\% CI (LnCIR) = LnCIR \pm 1.96 \times SE (LnCIR)$$

$$SE (LnIDR) = \sqrt{\frac{1}{a} + \frac{1}{b}}$$

$$95\% CI (LnIDR) = LnIDR \pm 1.96 \times SE (LnIDR)$$

归因危险度

完全由某因素所致的危险度。用暴露组的发病率(或死亡率)减去非暴露组的发病率(或死亡率)的余数表示之。

$$\begin{aligned}AR &= I_e - I_0 = a/N_1 - c/N_0 \\ &= I_0 \cdot RR - I_0 = I_0 (RR - 1)\end{aligned}$$

RR与AR

- RR针对个体而言，具有病因学意义
- AR针对群体而言，具有公共卫生学意义

归因危险度百分比

- AR%

$$AR\% = \frac{I_e - I_0}{I_e} \times 100\% = \frac{RR - 1}{RR} \times 100\%$$

人群归因危险度及人群归因危险度百分比

- $PAR = I_t - I_0$

- $PAR\% = \frac{I_t - I_0}{I_t} \times 100\% = \frac{Pe(RR-1)}{Pe(RR-1)+1} \times 100\%$

研究实例

- 吸烟与肺癌死亡的前瞻性队列研究

RR=13.7

AR=0.89‰

AR%=92.7%

PAR=0.49‰

PAR%=87.5%

- 剂量反应关系分析
- 分层分析

标准化死亡率

- 暴露人群人数较少, 无法获得发病率或死亡率资料

标准化死亡率比(SMR)

研究人群中观察死亡数与以标准人口死亡率计算的预期死亡数之比。

$$SMR = \frac{a_i}{E(a_i)} \times 100$$

标准化死亡率比 (SMR) 计算举例

项目	年龄	年份			合计
		1950-54	1955-59	1960-64	
人年	20-24	1,000	500	200	
	25-29	1,000	1,500	1,000	
	30-34	500	500	1,500	
观察死亡数	20-24	2	1	0	15
	25-29	3	4	2	
	30-34	0	1	2	
人群死亡率 (1/1000)	20-24	1.8	1.8	1.6	
	25-29	1.7	1.5	1.5	
	30-34	1.9	1.8	1.7	
预期死亡数	20-24	1.8	0.9	0.3	12.9
	25-29	1.7	2.3	1.5	
	30-34	0.9	0.9	2.6	

标准化死亡比例比

- SPMR

预期死亡值的计算是根据全人群某疾病死亡数与全死因死亡数之比来计算，而不是用死亡率。

$$\text{SPMR} = \frac{a_i}{E(a_i)} \times 100$$

标准化死亡比例比计算表

年龄组	结核病死亡占全死因的比例(1)	工人死亡人数(2)	工人结核病预期死亡数(3)	工人结核病观察死亡数(4)
20-	6.28	292	18.34	10
25-	8.30	357	29.63	20
30-	9.01	341	30.72	22
35-	7.78	1095	85.19	98
45-	5.19	1784	92.59	174
55-	3.68	1554	57.19	112
60-64	2.77	2051	56.81	104
合计			370.47	540

队列研究中的偏倚与混杂

- 常见偏倚

选择性偏倚

失访偏倚

信息偏倚

混杂偏倚

偏倚的控制

- 严格按照规定选择研究对象
- 提高研究对象的依从性
- 提高设计水平，做好质量控制
- 分层分析和多因素分析

队列研究的优点

- 可直接获得两个组的发病率或死亡率，并可直接获得RR
- 可减少收集资料过程中由主观因素带入的偏性
- 在追踪观察中可以及时了解 and 记录对象某些特征或暴露情况的改变
- 可以获得多种疾病的结局资料

队列研究的局限性

- 实施一次队列研究比较困难，且耗费较多的人力、资金和时间
- 通过定期医学检查获得结局资料时，观察对象容易了解到研究目的而改变他们的行为，给资料分析带来了困难
- 不适用于发病率很低疾病的研究

历史性队列研究

- 所需时间短
- 仅依靠一些记录资料、检查数据或个人提供的资料
- 需收集过去的暴露史
- 研究开始时疾病已发生或死亡

三种流行病学方法的比较

方法学特点	前瞻性队列研究	病例对照研究	现况研究
样本组成	无病个体	病例与对照	暴露者、现患者或存活者
分组标准	暴露 - 未暴露	患病 - 未患病	前两者之一
时间顺序	前瞻性(因---果)	回顾性(果---因)	现况
比较内容	暴露者与未暴露组发病或死亡情况	病例与对照过去的暴露情况	暴露者的患病情况或患病者的暴露情况
率	发病(死亡)率	暴露百分率	现患率、暴露率
暴露与疾病联系指标	相对危险度, 归因危险度, PAR	OR, PAR(在一定条件下)	率比, 率差, PAR
优点			
局限性			