

# 第三章

## 描述性研究

# 本章内容

第一节 概 述

第二节 现况研究

第三节 生态学研究

# 第一节 概述

## 一、概念

利用已有的资料或对专门调查的资料，按不同地区、不同时间及不同人群特征分组，描绘、叙述疾病或健康状态的分布情况

流行病学研究基础

## 二、种类

现况研究

病例报告

病例系列分析

个案调查

历史资料分析

随访研究

生态学研究

## 三、用途

1. 描述医学事件的分布
2. 获得病因线索，提出病因假设

## 第二节 现况研究

- 一 概 念
- 二 目 的
- 三 特 点
- 四 种 类
- 五 设计与实施

# 一、概念

★研究特定时点或期间、特定范围内人群中某病或健康状况的分布，并研究有关变量与疾病或健康的关系的一种调查方法。

★又称：横断面研究 (Cross-sectional study)

患病率研究 (Prevalence study)

## 二、现况研究的目的

- (一) 描述疾病或健康状况的三间分布。
- (二) 描述某因素与疾病或健康之间的关联。
- (三) 评价防制措施的效果。
- (四) 确定高危人群



# 三 特 点

- ★ 一般不设对照组
- ★ 特定时点或期间的疾病状况
- ★ 确定因果联系受限
- ★ 对不改变的暴露因素，可作因果推论
- ★ 用现在的暴露（特征）来替代或估计过去情况是有条件的
- ★ 一般不能计算发病率，定期重复进行可获得发病率资料

## 四、现况研究的种类

### （一）普查

#### 1. 概念

于一定时间内对一定范围的人群中每一成员所作的调查或检查。

## 2. 普查的目的

实现描述性研究的所有目的。

**特点：普查普治。**

### 3. 普查的适用条件

- (1) 适用于患病率较高的疾病。
- (2) 疾病的检测手段和方法简单准确，最好有可靠有效的治疗方法。
- (3) 调查范围小，人数少时，可直接用普查。
- (4) 具备人力、物力和设备等条件，有完成普查的可行性。
- (5) 应能保证一定的普查率。

$$\text{普查率} = \frac{\text{实查人数}}{\text{应查人数}} * 100\%$$

**一般不低于 80 %。**

### 3. 普查的优缺点

#### (1) 优点

- 1) 设计特别是确定调查对象上比较简单。
- 2) 不存在抽样误差。
- 3) 可查出某人群中患某病的所有病人，使其得到及时治疗。
- 4) 普查的同时也开展了一次医学科普教育。

## **( 2 ) 缺点**

- 1) 普查对象多，调查期限短，漏查是难免的。**
- 2) 参加普查的工作人员多，他们掌握调查技术和检验方法的程度不等，调查质量不易控制。**
- 3) 患病率低，诊断技术复杂的疾病不宜进行普查。**
- 4) 人力、物力耗费大。**

## **(二) 抽样调查**

### **1. 概念**

**指只调查从某人群中抽出的一部分有代表性的人（统计学上称为样本），根据调查结果估计出人群某病的患病率或某特征的情况。**

2. 要抽取一个有代表性的样本，  
就必须遵循

随机化原则

样本大小适当的原则。



## **( 1 ) 随机化原则**

**指整个研究人群中的每一个单位（可以是个人，也可以是个人的集合体，如学校、连队、班级或居民委员会等）被选入样本的概率相等。**

## (2) 样本大小适当的原则

是指样本应达到一定数量， 样本过小时可能所抽出的样本的代表性不够； 样本过大不但浪费人力、物力，而且工作量过大， 容易因调查不够细致而造成偏倚。

### 3. 抽样调查的优缺点

#### (1) 优点：

省时、省力、省材料和省经费。

调查样本相对较小，因而较易集中人力、物力和器材设备，调查结果也易作到细致、准确。

#### (2) 缺点：

设计、组织实施以及资料分析等方面比较复杂。

重复和遗漏不易发现。

不适用于变异过大的材料。

## 四、现况调查的设计与实施

### （一）选题和确定本次调查研究的目的

**选题：**

创新性、重要问题、实用性

**目的：**

明确、具体；

一次调查的病种不宜太多

## **（二）确定研究对象和调查方法**

**研究对象：**

**目标人群、调查人群**

**调查方法：**

**普查？抽样调查？**

## **（三）抽样方法与样本含量**

### **1 抽样方法**

#### **（1）非随机抽样**

选择样本时，加入人主观因素，使总体中每个个体被抽取的机会是不均等的

如典型抽样：试验者根据试验调查的目的、要求和被调查对象的总体情况，有意识地选择那些具有代表性的对象进行试验

## ( 2 ) 随机抽样

遵循随机化原则，保证总体中每一个对象都有同等机会被选入作为研究对象

### 常用的随机化抽样方法

- ( 1 ) 单纯随机抽样
- ( 2 ) 系统抽样
- ( 3 ) 分层抽样
- ( 4 ) 整群抽样
- ( 5 ) 两级或多级抽样

# 单纯随机抽样

## Simple random sampling

- ◆ 也称简单随机抽样，最简单、最基本的抽样方法
- ◆ 从总体  $N$  个对象中，利用抽签或其他随机方法抽取  $n$  个
- ◆ 总体中每个对象被抽到的概率相等



例：从 10 人中随机抽取 1 人

编号	1	2	3	4	5	6	7	8	9	10
----	---	---	---	---	---	---	---	---	---	----

查随机数字

# 1 0 0 0 个随机数字

		00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
26927	01	15389	85205	18850	39226	42249	90669	96325	23248	60933	
	02	85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
	03	61149	69440	11286	88218	58925	03638	52862	62733	33451	
77455	04	05219	81619	10651	67079	92511	59888	84502	72095	83463	
	05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
	06	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
75577	07	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
	08	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
	09	82995	64157	66164	41180	10089	41757	78258	96488	88629	
90279	10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
	11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
	12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
11551	13	62111	52820	07243	79931	89292	84767	85693	73947	22278	
	14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
	15	98614	75993	84460	62846	59844	14922	48730	73443	48167	34770

## 1 0 0 0 个随机数字

		00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
26927	01	15389	85205	18850	39226	42249	90669	96325	23248	60933	
	02	85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
	03	61149	69440	11286	88218	58925	03638	52862	62733	33451	
77455	04	05219	81619	10651	67079	92511	59888	84502	72095	83463	
75577	05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
90279	06	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
	07	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
	08	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
	09	82995	64157	66164	41180	10089	41757	78258	96488	88629	
	10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
11551	11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
	12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
	13	62111	52820	07243	79931	89292	84767	85693	73947	22278	
	14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
	15	98614	75993	84460	62846	59844	14922	48730	73443	48167	34770

例：从 10 人中随机抽取 1 人

编号	1	2	3	4	5	6	7	8	9
10									

随机数字	34	35	78	80	40	53	36	47	17
26									

排序

随机数字最小者抽中

# 系统抽样

## Systemic sampling

- 又称机械抽样

- 按照一定顺序，机械每隔若干单位抽取一个单位

- 方法

- 将总体各个个体单位按某种标志排列、连续编号
- 根据总体数  $N$  和确定的样本数  $n$ ，计算抽样距离 ( $N/n$ )
- 第一段距离内，随机抽取一个号码，作为第一个调查样本单位
- 将第一个样本单位的号码加上抽样距离，得到第二个样本单位，以此类推，直至满足样本量

- **系统抽样的优点**

- ( 1 ) 可以不知道总体单位数

- ( 2 ) 大样本易于操作。

- **系统抽样的缺点**

- 总体单位有周期性，会影响代表性

# 分层抽样

## Stratified sampling

- ◆ 将总体单位按某种特征分为若干次级总体（层）
- ◆ 从每一层内单纯随机抽样
- ◆ 按比例分配，最优分配
- ◆ 抽样误差最小

# 整群抽样

## Cluster sampling

- ◆将总体分成若干群组，抽取部分群组

- ◆单纯整群抽样

  - 调查被抽到的群组中的全部个体

- ◆二阶段抽样

  - 调查部分个体



## **整群抽样特点：**

**易于组织实施**

**抽样误差较大，增加  $1/2$  样本量**

**群间差异越小，抽取群越多，抽样误差越小**

# 多级抽样

## Multistage sampling

◆用于大型流行病学调查

◆方法

- 从总体中抽取范围较大的单元（一级抽样）
- 从一级单元中抽取范围较小的单元（二级抽样）
- 依次类推

## 2. 样本含量估计

### ( 1 ) 决定样本大小的因素

1) 容许误差

2) 预期患病率或标准差

## (2) 样本含量的计算

### 1) 计量资料

$$n = \frac{4s^2}{d^2}$$

公式中  $n$  为样本含量， $d$  为容许误差，即样本均数与总体均数之差，由调查设计者根据实际情况规定。 $s$  为样本标准差。

**例：**欲调查某病病人血红蛋白含量，根据以往的经验， $s = 3.0\text{g}/100\text{ml}$ ，要求误差不超过  $0.5\text{g}/100\text{ml}$ ，则该调查样本大小为：

$$n = \frac{4s^2}{d^2} = \frac{4 \times 3.0^2}{0.5^2} = 144(\text{人})$$

## 2 ) 计数资料

当容许误差  $d=0.1P$  时, 则

$$N=400 \times Q/P$$

当容许误差  $d=0.15P$  时, 则

$$N=178 \times Q/P$$

当容许误差  $d=0.2P$  时, 则

$$N=100 \times Q/P$$

上式中  $P$  是某病患率,  $Q=1 - P$ ,  $N$  即样本数量。 此公式适用于呈二项分布性质的资料, 且患病率不太大或太小的情况。

### 3) 患病率很低疾病的样本含量估计

按泊松分布期望值的可信限进行估计。

例如：

某地的肝癌患病率估计为 20/10 万，欲对该地的肝癌患病情况进行调查，问应抽多少人？

如果随机抽 1 万人，按估计的患病率计算，可能发现 2 例病人。

参考 Poisson 分布可信限表，期望值为 2 的  
90% 可信限下限为 0.355，上限为 6.30。

Poisson 分布期望值的可信限

期望值	0.95		0.90	
	下限	上限	下限	上限
0	0.0000	3.69	0.0000	3.00
1	0.0253	5.57	0.0513	4.74
2	0.242	7.22	0.355	6.30
3	0.619	8.77	0.818	7.75
4	1.09	10.24	1.37	9.15
5	1.62	11.67	1.97	10.51

这样我们就有可能见不到病例，使调查工作失去意义。

当期望值为4时，90%可信限下限为1.37，即有90%的机会可见到病例。

这样，抽取20000人可满足期望值为4的要求。

如采用整群抽样，样本量须加大，可粗定为加大简单随机抽样的样本量的1/2，即应抽取30000人。

$$n = \frac{1 - 0.0002}{0.0002} \times 100 = 499900$$



## （四）确定疾病的测量方法

应尽量采用简单、易行的技术和灵敏度高的检验方法。

## （五）确定研究变量及其定义和测量方法

明确的定义、可靠的测量方法  
如：腰围

## **(六) 拟订调查表**

**调查表类型：**

**一览表式**

**一人一表式**

**问题类型：**

**开放式、封闭式、复合式**

# 调查表（一人一表）

主要包括三个部分。

第一部分，即一般性项目

包括姓名、年龄、性别、出生年月、出生地、文化程度、职业、民族、工作单位、现住址等。

便于核查，补填或更正而设置的。

人口学资料

## 第二部分，即**调查研究项目**

**调查研究的实质部分。**

### **几项原则供参考：**

- ① 与本次调查有关的项目一项也不能缺。
- ② 每个调查项目都要用通俗的文字准确无误地表达出来 不应使被调查者产生误解或出现不同的理解。
- ③ 应尽量选用那些能以客观指标来回答的问题询问调查对象。
- ④ 问题由易到难

## 第三部分，即调查者部分

- 调查质量的评价
- 记录特别的情况
- 列出“调查者”和“调查日期”，有助于查询和明确责任。

## **（七）挑选和培训调查员。**

### **调查员选择**

**工作认真、吃苦耐劳、遵守纪律**

### **培训**

**工作的意义**

**科学严谨的态度**

**相关技术**

### **考核**

## **（八）进行现场调查**

**❖通过测定或检查的方法收集**

**❖直接用调查表询问研究对象**

**方式：面访、通信调查、电话调查、自我管理式调查。**

## (九) 调查资料整理、分析

### ① 检查与核对原始资料

检查原始资料的准确性、完整性，填补缺漏，删去重复，纠正错误。

### ② 数据库建立



### ③ 统计描述

#### ✓ 率的计算

患病率、感染率等

人群分布

地区分布

重复的现况调查可分析时间分布

#### ✓ 计量资料

标准差

均数、中位数、几何均数、百分位数、标

人群分布

地区分布

### ④ 对比分析

分布差异

影响分布的因素

# 五、现况调查中常见的偏倚及其防止

## （一）偏倚的概念

1. 误差：测得值与真实值之差。

### （1）抽样误差

由于抽样产生得样本指标与总体参数之差。

### （2）系统误差

由于某些较恒定的因素造成的测得值倾向性得偏离真实值。

在流行病学上系统误差被称为  
**偏倚**

## 2. 偏倚产生的环节

设计、实施、资料处理和分析

## 3. 偏倚的控制

找到偏倚产生的原因，消除偏倚。

如：

通过正确的设计控制偏倚；

严格校正仪器、培训实验人员控制偏倚。

## **（二）现况调查中可能发生的偏倚**

### **1. 选择偏倚**

**（1）研究对象选择方法不当引起的偏倚**

**（2）无应答偏倚**

**应答率要高于 90%**

**（3）幸存者偏倚**

### **2. 信息偏倚**

**（1）调查对象所引起的偏倚**

**报告偏倚**

**（2）调查员偏倚**

**（3）测量偏倚**

## **（二）防止产生偏倚的措施**

**针对偏倚产生的原因采取相应措施。**

- 1. 在设计中明确规定为随机样本的，必须严格遵守随机化的原则。尽量提高应答率。设法补查一部分无应答者并作分析。**
- 2. 训练调查员并对其进行监督和质量控制。**
- 3. 选用精良的仪器设备并事先作好校准。在整个调查中所用试剂力求一致，以消除可能引起的差异。**
- 4. 采用适当技术控制调查者偏倚**

## 第三节 生态学研究

# 一、定义

以群体为观察、分析单位，通过描述不同人群某因素的暴露情况与疾病的频率，分析该因素与疾病的关系。



## 二、生态学研究的方法

### 1. 生态比较研究

比较不同人群中某疾病的发病率、死亡率或健康状态的差别，了解某些因素的出现率与疾病率的关系，从而获得病因的线索。

### 2. 生态趋势研究

连续观察一个或多个人群中平均暴露水平的改变和某疾病的发病率、死亡率的变化关系。

## 不同国家反应停销售量与短肢畸形关

系	国家	反应停销售量 (公斤)	短肢畸形例数
	奥地利	207	8
	比利时	258	26
	英 国	5769	349
	荷 兰	140	25
	挪 威	60	11
	葡萄牙	37	2
	瑞 士	113	6
	西 德	30099	5000

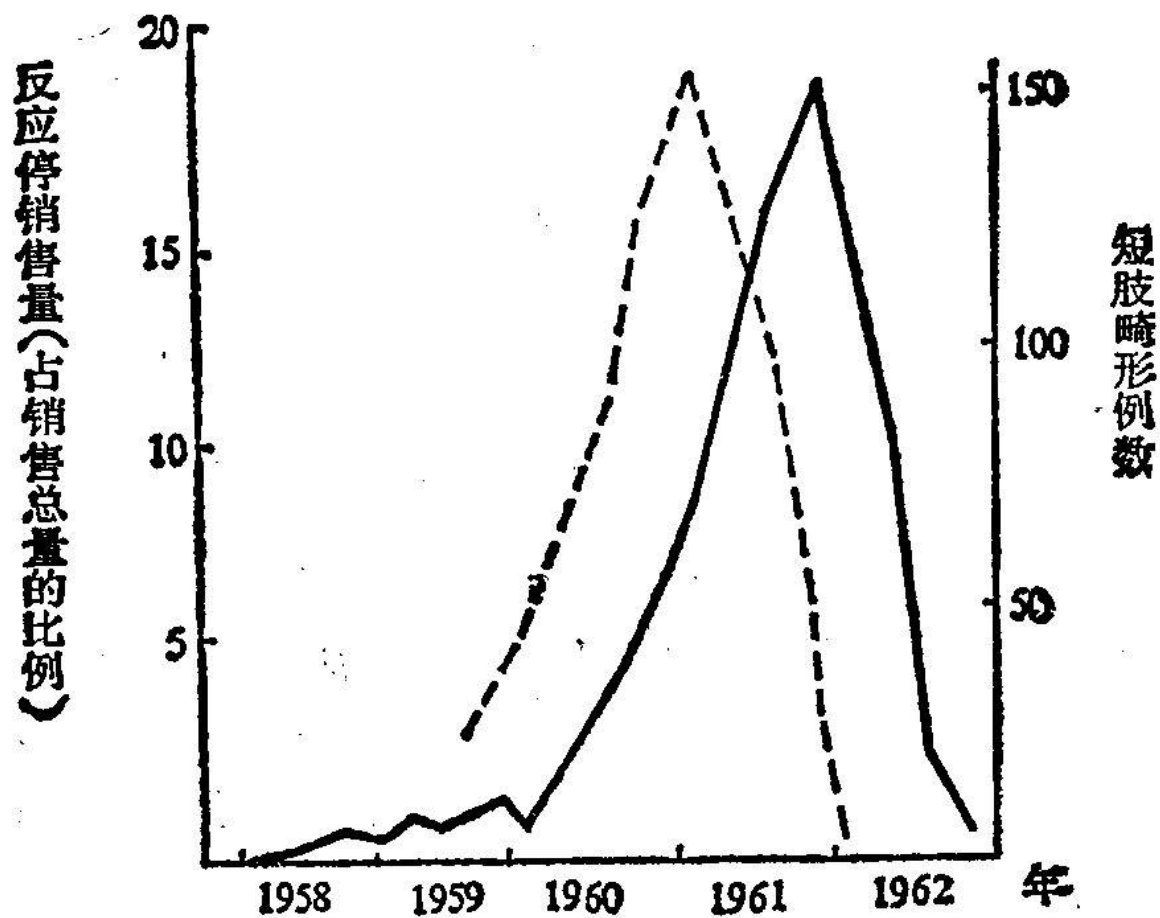


图4-19 西德反应停销售总量（虚线）与短肢畸形病例数（实线）的时间分布  
(流行病学研究实例1984年)

### 三、生态学研究用途

★提供病因线索，产生病因假设

★评估人群干预措施的效果

## 四 优 点

- ★ 经济 ， 出结果快
- ★ 提供病因未明疾病的病因线索
- ★ 对个体剂量无法测量的情况，是唯一可供选择的方法
- ★ 适用于研究因素暴露变异范围小，较难测量暴露与疾病的关系
- ★ 人群干预措施的评价及估计疾病发展趋势

## 五、生态学研究局限性

- ★ 出现生态学谬误
- ★ 难以控制混杂因素
- ★ 存在多重共线性问题
- ★ 难以确定因果关系

# 生态学谬误

- 在生态学研究中，生态学谬误是此类研究最主要的缺点，其是由于生态学研究是由各个不同情况的个体“集合”而成的群体（组）为观察和分析的单位，以及存在的混杂因素等原因而造成的研究结果与真实情况不符。

- **A 班和 B 班**

A 班学一门课的平均时间是 10 小时，平均成绩是 90 分；

B 班学生学这门课的平均时间是 15 小时，平均成绩是 80

结论：学习时间和成绩是负相关的。

学习起点不同，智力水平不同



## 难以控制混杂因素

- 1964-1965 年 28 个国家
- 平均每日猪肉摄入量与乳腺癌猪肉摄入量与乳腺癌死亡率
- 强正相关
- 猪肉摄入量与乳腺癌可能有联系
- 猪肉摄入量可能是乳腺癌危险因素（蔬菜摄入减少、脂肪摄入增加）的一个标志





# 存在多重共线性问题

- GDP 与恶性肿瘤（环境污染）
- GDP 与糖尿病（肥胖）



# 难以确定因果关系

- 时间顺序
- 初步的研究
- 二手资料
- 生态学谬误



***The end***