

描述性研究

Descriptive Study



# 引言

**1854年**秋季，伦敦宽街暴发霍乱，**10**天内死去**500**多人，在霍乱暴发后的**6**天内发病严重的街道有**3/4**以上的居民离去。据**1853**年统计，在过去几次霍乱流行中，该地区虽曾流行，但远比其他各区轻微。



死亡无数



人们纷纷逃散



英国医师John.Snow运用了流行病学调查，分析了霍乱的流行情况。

当时霍乱病原体尚未发现，**Snow**集中精力调查发生疫情的地点和死亡病例，发现几乎所有的死亡病例都发生于离宽街水井不远的地方，且他们都饮用宽街供水站的水，**Snow**根据疾病分布进行分析，教区霍乱暴发与宽街供水站的水井有密切关系，而以后的研究进一步证实了这一假设。经封闭水井，暴发即告终止。

Snow关于霍乱的调查，正是运用了**描述性研究**的方法，分析霍乱的人群现象、地区差异等情况，了解霍乱的分布，根据分布特点，提出霍乱暴发与宽街供水站有关这一假设。

**描述性研究**通过描述疾病或健康状况的三间分布情况，找出某些因素与疾病或健康状况间的关系，提供病因线索。

它既是流行病学研究工作的起点，也是其他流行病学研究方法的基础。

# 目 录

第一节 现况研究概述

第二节 现况研究的设计与实施

第三节 研究实例

第四节 生态学研究

# 第一节 现况研究概述

- ✚ 基本概念
- ✚ 研究目的与应用范围
- ✚ 研究特点与研究类型

## 一、基本概念

- ◆ 现况研究又称横断面研究 (**Cross-sectional study**) 也称患病率研究 (**Prevalence study**)
- ◆ 研究特定时点或期间和特定范围内人群中的有关变量(因素)与疾病或健康状况的关系

## 二、研究目的与应用范围

- ◆ 掌握目标群体中疾病或健康状况的分布
- ◆ 提供疾病病因研究的线索
- ◆ 确定高危人群
- ◆ 评价疾病监测、预防接种等防治措施效果

## 三、研究特点与研究类型

### ◆ 研究特点

- 开始时一般不设对照组
- 现况研究的特定时间
- 在确定因果关系时受到限制
- 对不会发生改变的暴露因素，可以提示因果关系

## ◆ 研究类型

### ● 普查（Census）

调查特定时点或时期、特定范围内的全部人群(总体)。

### ● 抽样调查（Sampling survey）

随机抽样，调查特定时点、特定范围人群的一个代表性样本，以样本统计量估计总体参数所在范围。

## (一)普查 (census)

**1、概念：**调查特定时点或时期、特定范围内的全部人群(总体)。

### **2、条件：**

- 有足够的人力、物资及设备用于发现、治疗疾病
- 普查的疾病患病率较高
- 普查的疾病有较简易而准确的检测手段或方法

### 3、目的：依不同的研究而异

- 早期发现和及时治疗疾病：高血压普查；
- 了解疾病的疫情及分布：血吸虫普查；
- 了解健康水平：儿童发育状况普查；
- 建立某些生理指标的正常值：如血脂；
- 了解某病的患病率以及流行病学特征，为防制工作提供依据。

## 4、普查的优点与局限性

---

### 优点

确定调查对象比较简单

可获得疾病的三间分布资料，对疾病的流行因素有一定的启示

### 局限性

普查对象多、期限短，易漏查

调查质量不易控制

不适用于患病率低，诊断复杂的疾病

---

## (二)抽样调查(sampling survey)

### 1、概念：

随机抽样，调查特定时点、特定范围人群的一个**代表性样本**，以样本统计量估计总体参数所在范围。

“以小窥大”，“以局部估计总体”

### 2、原理：

- 随机化的原则
- 每个单位都有同等机会被抽中
- 样本足够大

### 3、抽样调查的优点与局限性

#### 优点

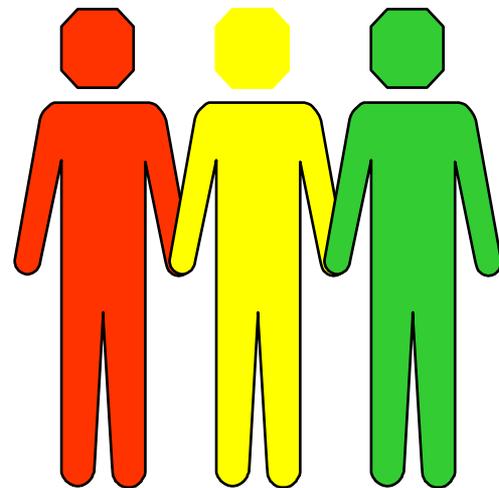
节省人力、物力、时间  
调查范围小，工作易做细致

#### 局限性

设计、实施比较复杂  
重复和遗漏不易发现  
不适用于变异过大的资料  
不适用于发病率很低的疾病

## 第二节 现况研究的设计与实施

- 明确调查目的和类型
- 确定研究对象
- 确定样本量和抽样方法
- 资料的收集
- 资料的整理与分析
- 常见偏倚及其控制
- 研究的优点与局限性



## 一、明确调查目的和类型

- ◆ 根据研究提出的问题，明确调查目的
- ◆ 根据具体研究目的确定是普查还是抽样调查

## 二、确定研究对象

根据研究目的规定对象的人群分布特征、地域范围以及时间点

- 某个区域内的全体居民或其中一部分
- 某一时点上的流动人员
- 某些特殊群体

- ◆ 进行二级预防：选择高危人群
- ◆ 描述疾病的三间分布：选择有代表性的人群

## 三、确定样本量和抽样方法

### ◆ 确定样本量

- 预期现患率 (  $P$  )
- 允许误差 (  $d$  )
- 显著性水平 (  $\alpha$  )

**计量资料:** 
$$n = \frac{4s^2}{d^2}$$

**计数资料** 
$$n = \frac{PQ}{S^2} = 400 \times \frac{Q}{P}$$

n: 样本含量

s: 总体标准差的估计值

d: 容许误差, 即样本均数(率)与总体均数(率)之差, 一般定为0.1P

P: 总体率的估计值(某病现患率),  $Q=1-P$

## ◆ 抽样方法

- 非随机抽样

- 随机抽样

## 非随机抽样

选择样本时，加入人的**主观因素**，使总体中每个个体被抽取的机会是不均等的。

### 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。

## 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

**方法** 单纯随机抽样

系统抽样

分层抽样

整群抽样

多级抽样

## 单纯随机抽样(简单随机抽样)

### Simple random sampling

- 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

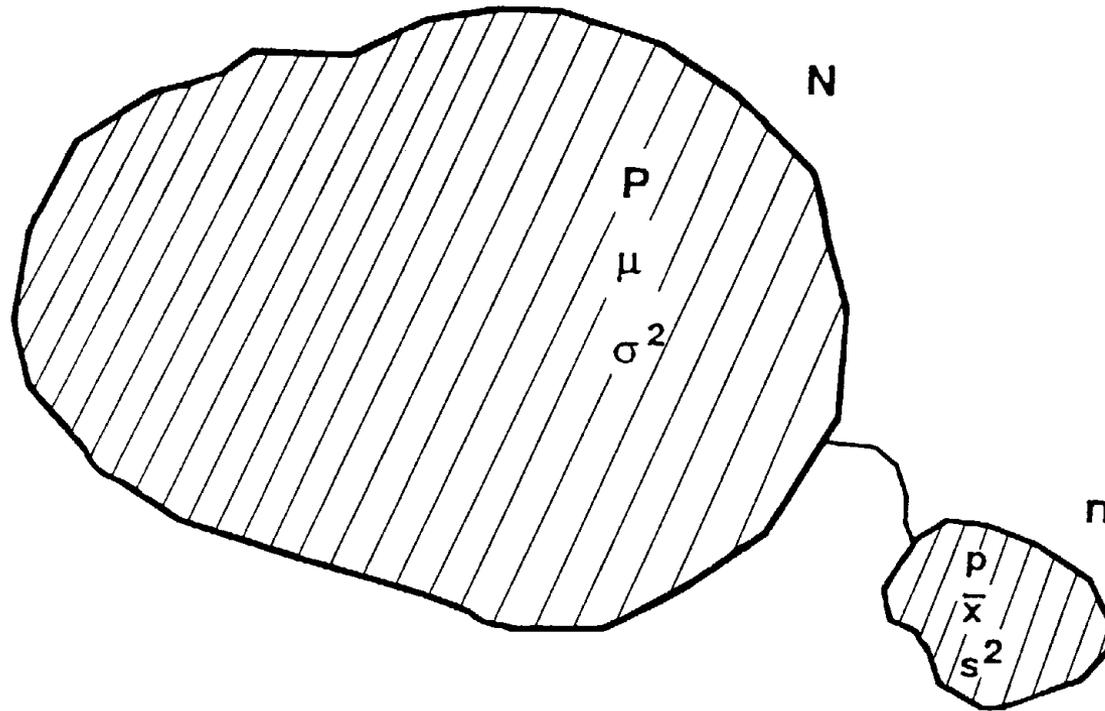


图3-1 总体与样本示意图

## 系统抽样(机械抽样)

### Systemic sampling

按照一定顺序，机械地每隔若干单位抽取一个单位的抽样方法。

- 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 ( $N/n$ )
- 用单纯随机方法在第一组中确定一个起始号
- 从此起始点开始，每隔 $K$  ( $K=N/n$ ) 个单位抽取一个作为研究对象

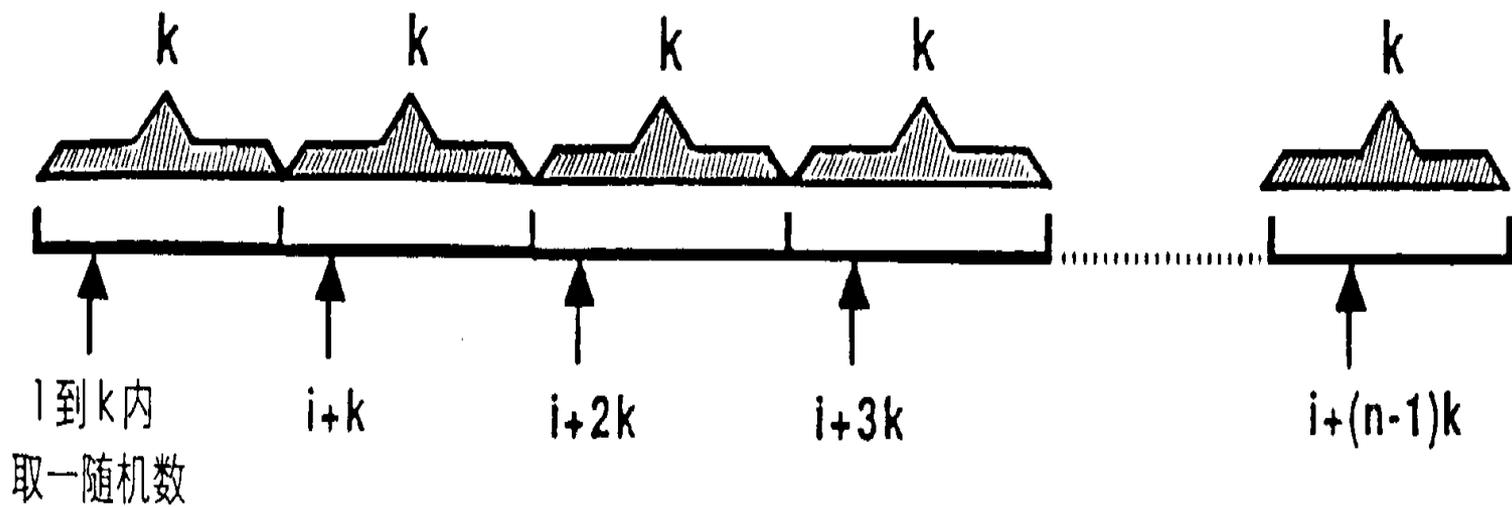


图3-2 系统抽样示意图

## 整群抽样

### Cluster sampling

将总体分成若干群组，抽取其中部分群组作为观察单位组成样本。

#### 分类：

- 单纯整群抽样（Simple cluster sampling）

被抽到的群组中的全部个体均作为调查对象

- 二阶段抽样（Two stages sampling）

通过再次抽样后调查部分个体

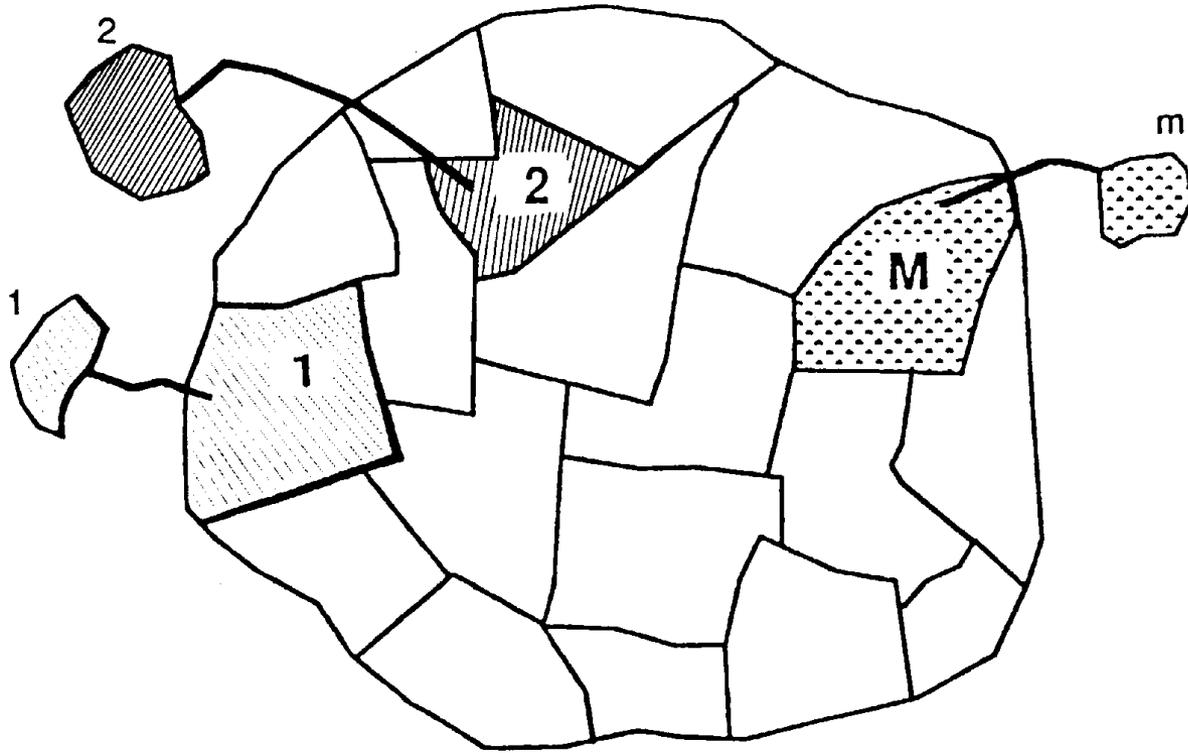


图3-3 整群抽样示意图

## 分层抽样

### Stratified sampling

将总体单位按某种特征分为若干次级（层），然后从每一层内单纯随机抽样组成一个样本。

#### 分类:

- 按比例分配（proportional allocation）分层随机抽样  
各层内抽样比例相同
- 最优分配（optimum allocation）分层随机抽样  
各层抽样比例不同，内部变异小的层抽样比例小，  
内部变异大的层抽样比例大

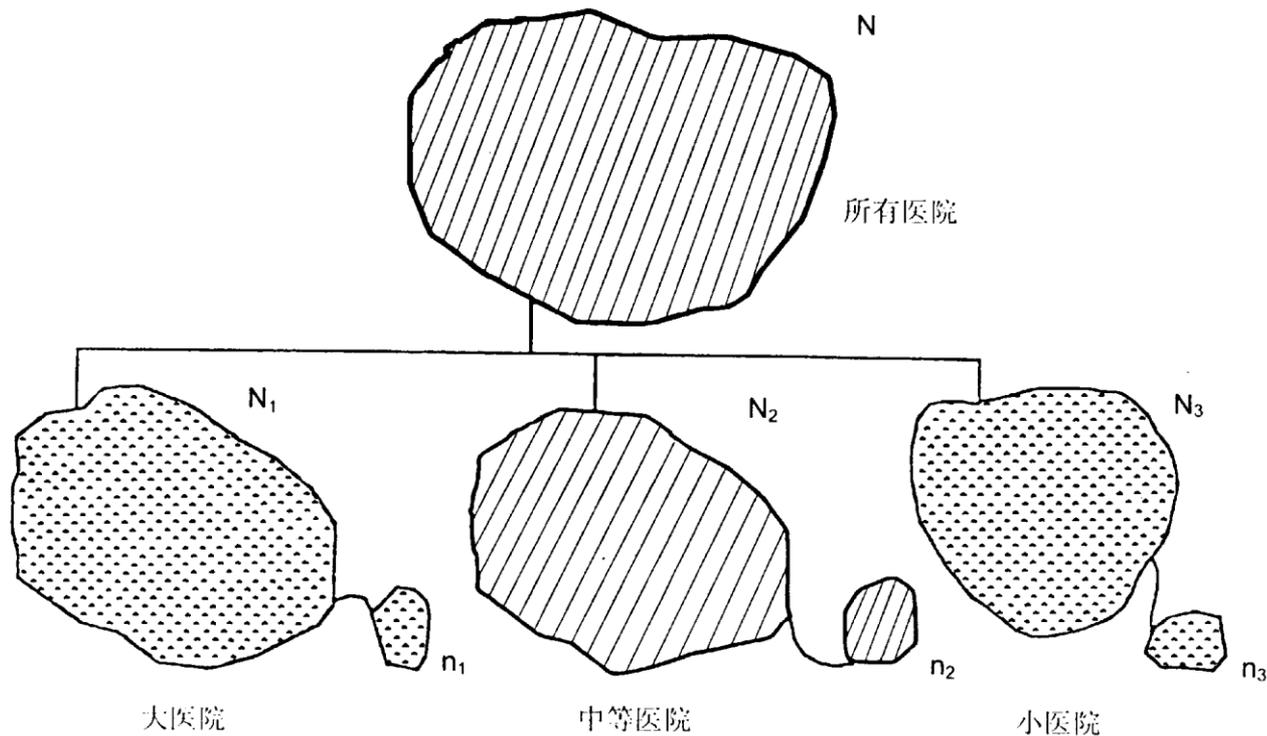


图3-4 分层抽样示意图

## 多级抽样

### Multistage sampling

将抽样过程分阶段进行，每个阶段使用的抽样方法往往不同，即将以上抽样方法结合使用，在大型流行病学调查中常用。

#### 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）
- 依此类推…

名称	方 法	举 例
单纯随机抽样	先将总体中每个抽样单元编号,然后从总体 $N$ 个对象中,利用抽签或其他随机方法(如随机数字)抽取 $n$ 个,构成一个样本	如某校有 300 名学生随机抽取 80 名,先将 300 名学生编号,再用抽签或随机数字表进行抽样
系统抽样	又称机械抽样,是按照一定顺序,机械地每隔若干单位抽取一个单位的抽样方法	如某镇有 1 000 户,拟抽 100 户,先将每户编号,从 1 ~ 10 号随机抽取 1 个号,如为 6,以后则每隔 10 户抽取 1 户,即抽取 6, 16, 26, ..., 996 号组成户数为 100 的样本
分层抽样	先将总体的单位按某种特征分为若干层,然后再从每一层内进行单纯随机抽样,组成一个样本	如欲对某县农村儿童的生长发育进行调查可先按乡镇的经济状况分层(如好、中、差三层),在各个乡镇再随机抽样
整群抽样	是指将总体分成若干群组,从总体中随机抽取若干群组作为观察单位组成样本	某校有 56 个班,欲了解学生的近视眼的患病率,可在学校随机抽取 8 个班进行调查
多级抽样	是将上述方法综合运用方法。即先从总体中抽取范围较大的单元,称为一级抽样单位(如省、市),再从每个抽得的一级单元中抽取范围较小的二级单元(县、乡、镇),依次类推,最后抽取其中范围更小的单元(如村、居委会)作为调查单位	要调查某城市初中生的吸烟情况,将全市中学按质量分成好、中、差三层,每层抽出若干学校,再在抽出的学校中,按年级分成三层,每个年级按整群抽样抽取若干班进行全部调查

## 四、资料的收集

- ◆ 通过测定或检查的方法收集
- ◆ 直接用调查表询问研究对象

## 五、资料的整理与分析

- ◆ 检查资料完整性和准确性
- ◆ 按标准归类、核实

- 检查资料的完整准确性，补缺、补漏，删除重复，纠正错误
- 对疾病或某健康状态按规定的标准归类核实
- 按不同空间、时间、人群描述分布特征和差异，进行显著性检验
- 按是否暴露研究因素进行分组作有对照组的比较分析

# 资料的整理、分析及结果解释

整理

查漏补缺、逻辑错检；编码、录入

分析

根据研究目的而异

结果解释

研究对象来源、抽样方法、样本代表性、三间分布规律

## 六、偏倚及其控制

◆ 常见的偏倚

◆ 偏倚的控制

# 1、误差(error)

指研究或实验得出的结果不符合真实情况或者是错误的，有随机误差和系统误差。

## (1) 随机误差(random error)

- 即**抽样误差**，指样本指标(统计量)与总体指标(参数)之间的差异
- 通常是可以估计与测量的
- **无法避免，可减少**：分层抽样、减少调查单位的变异、增大样本量

## (2)系统误差(systematic error,bias)

- ▶ **又称偏倚**，由某些不能准确定量的、但较为恒定的因素所引起，如受试者，观察者，仪器，外环境的非实验因素
- ▶ 其大小和方向恒定，或按照一定的规律变化
- ▶ 可以发生在设计、实施和分析的任一环节
- ▶ **可以避免或减少**：实验设计(如随机、对照、重复)，技术措施(如校正仪器、熟练操作)

# 随机误差和系统误差的区别

## 随机误差

## 系统误差

增加样本量可减少

增加样本量不能减少

可以估计与测量

不可估计与测量

发生在研究的设计阶段

可发生在设计、实施和分析的任一环节

不可避免

可以避免

## 偏倚

### ◆ 定义

从研究设计、实施、到数据处理和分析的各个环节中产生的系统误差，以及结果解释、推论中的片面性，导致**研究结果与真实值之间出现倾向性的差异**，从而错误地描述暴露与疾病之间的联系，称之为偏倚。

## ◆ 偏倚的种类

### 选择偏倚

不正确选择研究对象引起：  
无应答偏倚、志愿者偏倚

### 信息偏倚

资料收集阶段，由于观察和测量方法引起的：报告偏倚、观察者偏倚、测量偏倚

## ◆ 产生原因

- 主观选择研究对象
- 任意变换抽样方法
- 调查对象不合作或因种种原因拒绝参加
- 调查到的对象均为幸存者，无法调查死亡者
- 回答不准确或回忆不清
- 调查偏倚
- 测量误差

## ◆ 偏倚的控制

- 严格遵照抽样方法要求，确保随机化原则
- 提高研究对象的依从性和受检率
- 正确选择测量工具和检测方法
- 培训调查员，统一标准和认识
- 做好资料的复查复核工作
- 选择正确的统计分析方法，辨析混杂因素

## 七、研究的优点与局限性

### ◆ 优点

- 常用抽样调查，结果有较强推广意义
- 有来自同一群体的自然形成的同期对照组，结果具有可比性
- 可同时观察多种因素

## ◆ 局限性

- 难以确定先因后果的时相关系
- 不能获得发病率资料
- 研究对象可能处于临床前期而被误定为正常人

## 八、应用

- ◆ 掌握目标人群中疾病的患病率及其分布
- ◆ 提供疾病致病因素的线索
- ◆ 确定高危人群
- ◆ 对疾病监测、预防接种效果及其他资料的质量评价

## 第三节 研究实例

- 明确调查目的和类型
- 确定研究对象、样本量和抽样方法
- 确定研究内容和资料的收集方法
- 资料整理与分析
- 调查结论

# 2000年第四次全国结核病 流行病学抽样调查

## 一、明确调查目的和类型

### ◆ 目的

获得全国结核病的患病率及相关的基础资料。

### ◆ 类型

普查和抽样调查都可以采用，但结核病属较常见病种，抽样调查是最优的选择。

## 二、确定研究对象、样本量和抽样方法

### ◆ 研究对象

全国的居民

### ◆ 样本量

约**38**万人

### ◆ 抽样方法

分层整群等比例随机抽样

## ◆ 具体方案

- 全国设**257**个调查点
- 抽样比例为**1 : 3200**
- 每个调查的群体人数保持在**1500**人左右
- 折算后全国抽样人数约为**38**万人左右，约每**480**万人口中调查**1500**人

## ◆ 实际调查

- 平均每个调查点实检人数为**1420**人
- 抽样人口为**418456**人
- 应检人口为**375599**人
- 实检人口为**365097**人
- 受检人口占应检人口的**97.2%**，占抽样人口的**87.3%**

## 三、确定研究内容和资料的收集方法

### ◆ 收集方法

- 实验室检测
- 问卷调查

### ◆ 调查、检测项目

- 肺结核的患病率
- 涂阳患病率和菌阳患病率
- 野生株的菌种鉴定和药物敏感试验
- 结核病及肺结核的死亡专率
- 结核病防治措施及肺结核患者社会经济情况

## 四、资料整理与分析

主要对结核病现患流行情况及其人群、空间和时间分布特征进行描述。

### 结果:

#### ◆ 流行状况

- 活动性肺结核患病率 **367/10万**
- 菌阳患病率 **160/10万**
- 涂阳患病率 **122/10万**

估算全国有活动性肺结核患者**451万**，  
菌阳肺结核**196万**，  
涂阳肺结核**150万**。

**结论** 我国人群结核患病率依然很高

## ◆ 人群分布

- 35岁年龄组的涂阳和菌阳肺结核患病率的男女性别无差异
- 其余年龄组的各类患病率男性均高于女性
- 患病率随年龄的增长呈不断增高趋势，55岁后更为明显，至75岁达到最高

## ◆ 空间分布

- 将我国分成东、中、西部三类地区,各地区的肺结核疫情有明显差异
- 农村是结核病疫情最严重的地区
- 项目地区活动性、涂阳和菌阳肺结核患病率明显低于其他非项目地区

### 表3-1 结核病流行的空间分布

因素	活动性肺结核 (例)	涂阳肺结核 (例)	菌阳肺结核 (例)
<b>地理位置</b>			
东部地区	245	93	120
中部地区	436	148	178
西部地区	451	137	199
<b>城乡分布</b>			
城市	211	68	117
城镇	319	101	129
农村	393	116	169
<b>地区</b>			
项目地区	328	110	150
非项目地区	429	143	180

## ◆ 时间趋势

### 标准化处理后

- **2000年**活动性、涂阳和菌阳肺结核患病率  
300/10万、97/10万和124/10万
- **1990年**活动性、涂阳和菌阳肺结核患病率  
523/10万、134/10万和177/10万
- 涂阳患病率2000年比1990年下降27.6%，年递降率为3.2%

## 五、调查结论

- ◆ 我国结核病的流行虽然呈缓慢下降趋势，但患病率依然较高
- ◆ 结核病患病率男性高于女性，随着年龄的增長患病率不断提高
- ◆ 经济不发达地区和农村结核病疫情较为严重

## 第四节 生态学研究

- ✚ 概 念
- ✚ 研究目的
- ✚ 研究类型
- ✚ 优点与局限
- ✚ 研究实例

## 一、概念

- ◆ 描述性研究的一种
- ◆ 群体的水平上研究某种因素与疾病的关系
- ◆ 以群体为观察和分析单位
- ◆ 描述不同人群中某因素的暴露状况与疾病的频率，分析该暴露因素与疾病的关系

## 二、研究目的

- ◆ 提供病因线索，产生病因假设
- ◆ 评估人群干预措施的效果

## 三、研究类型

- ◆ 生态比较研究
- ◆ 生态趋势研究

# 生态比较研究

## Ecological comparison study

- 应用较多的一种方法
- 最简单的方法
- 观察不同人群或地区某种疾病的分布，根据疾病分布的差异，提出病因假设

## 生态趋势研究

### Ecological trend study

连续观察不同人群中某因素**平均暴露水平**的改变和(或)某种疾病发病率、死亡率变化的关系，了解变动趋势，比较暴露水平变化前后疾病频率的变化情况，判断某因素与某疾病的联系。

## 四、优点与局限性

### ◆ 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

## ◆ 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
- 难以确定因果联系