

促进肺部健康研究方法

低收入国家项目开展指南

国际防痨和肺病联合会
中国防痨协会

促进肺部健康研究方法

低收入国家项目开展指南

Donald A Enarson MD

Susan M Kennedy PhD

David L Miller MD

Per Bakke MD PhD

国际防痨和肺病联合会 出版
中国防痨协会 译

**Research Methods for
Promotion of Lung Health**

**A guide to protocol development
for low-income countries**

**From materials developed for
International Course on
Research Methods for Promotion of Lung Health
Sponsored by the IUATLD
Publication supported by a grant from USAID**

**International Union Against Tuberculosis And Lung Disease
68 boulevard Saint-Michel, Paris 75006 FRANCE**

促进肺部健康研究方法

低收入国家项目开展指南

译：贺晓新

校：屠德华 王擷秀

审：张立兴

中国防痨协会
2002年5月

前 言

1999年10月23—30日由国际防痨和肺病联合会与中国防痨协会共同举办的“国际青年医生肺部健康促进研究方法研讨班”在北京举行。共有来自越南、尼泊尔、蒙古、柬埔寨、美国以及台湾、北京、天津、山东、黑龙江、广东、浙江、云南、辽宁等国家和地区的20名青年肺科和结核病医师参加。该研讨班就是采用“肺部健康促进研究方法”作为教材的。

该研讨班由本书的两位作者 Donald A Enarson 和 David L Miller 医生以及来自澳大利亚 Guy Barrington Marks 医生和我国端木宏谨和张立兴医生共同担任授课任务。

参加人员对该书很感兴趣，认为理论联系实际，深入浅出，对研究工作具有重要参考价值。原教材经过进一步修订和扩展，由国际防痨和肺病联合会于2001年出版为“肺部健康促进研究方法”一书。

该书由贺晓新博士翻译，屠德华副理事长和王擷秀副理事长校。

本文的出版得到天津结核病控制中心和北京结核病控制研究所的帮助。

由于时间仓促，译文不妥之处在所难免。热忱欢迎读者不吝指正。

张立兴 中国防痨协会 2002.8.23

序言

国际防痨和肺病联合会（IUATLD）是世界上最早的非政府国际卫生组织。其任务是制定促进肺部健康以及预防肺部疾病的策略和机制。IUATLD 关注影响肺部健康最常见和最严重的问题，并将其活动重点放在低收入国家，因为那里是绝大多数高危人群的所在地。

为了有效完成上述任务，IUATLD 将科学研究作为其活动的重点。它已经这样做了，它将科学研究视为开展有效卫生活动的三大支柱之一，其它两大支柱是公共卫生项目的技术支持和宣教活动。卫生研究发展委员会是以促进低收入国家人民健康为目标的组织¹，它认为科学研究“是实现上述目标（促进健康）的强有力工具—但实际上人们在很大程度上忽略了这一点”。接着指出，实现此目标依赖于政府承诺、卫生规划对人群的覆盖面以及为实现上述目标提供所需资源的政治意愿和指导上述卫生活动所需要的理论知识。知识产生于科学研究，并且对“健康促进的有效活动必不可少”。

IUATLD 的经验是通过国际呼吸系疾病研究单位，首先确立在低收入国家开展有效科研项目的基本要求²：

- 1.完成该任务所需要的机构；
- 2.完成该任务所需要的人员；
- 3.研究人员之间的交流以刺激学术气氛。

为了满足这些要求，IUATLD 与追求同一目标的其它组织合作建立了一个肺部健康促进研究方法的国际教程网络。其目的是：

- 1.促进低收入国家研究机构的研究工作；
- 2.帮助这些研究机构建立基本研究方法方面的培训能力；
- 3.扩展先进研究技术方面的专业知识；
- 4.充实研究机构的继续教育教材；
- 5.鼓励低收入国家的研究者进行国际合作研究；
- 6.加强研究机构间的研究合作。

本教材中的资料主要服务于第 4 个目的。这些资料最初是为 1997 年在伊斯坦布尔举办的肺部健康促进研究方法培训班准备的。在该培训班上使用了教材“流行病学和公共卫生演讲要点”³。本教材就是根据上述材料并结合 IUATLD 在其他地方举办的相似培训班上

1 Evans JR, Castillo GT, Abed FH, et al. Health research: essential link to equity in development. New York: Oxford University Press, 1987; 1-136.

2 Becklake MR. 国际呼吸系疾病研究单位报告. 巴黎: IUATLD, 1993.

3 Farmer R, Miller D, Lawrenson R. 流行病学和公共卫生授课要点, 第四版. 牛津: Blackwell Science, 1995.

所获得的经验编写而成的。

本书包含两部分。第一部分是课程要点，第二部分是练习以及 Epi Info 软件包在肺部健康促进研究中的使用指南，Epi Info 软件包在不断升级。本资料有印刷版和网页版（www.iuatld.org）两种形式可供索取。练习部分可以根据具体需要进行扩展或改编。

本教程旨在帮助公共卫生工作者以及研究者根据低收入国家的实际情况制定研究方案。这些资料按照课程格式编排，包含四个部分：基础流行病学讲义、基础流行病学讲义练习、运用 Epi Info 软件进行数据管理、以及根据实际情况制订研究方案。

通过伊斯坦布尔其他国际组织人员的努力、其他培训班获得的经验以及我们与学员之间的交流，使上述各部分内容得到进一步丰富。通过同行们将来在培训班上的运用，这些材料无疑会得到进一步扩展和改进。另外，我们期望世界上低收入国家的地方研究机构能应用该材料作为教学内容。并将有关内容进一步修订以反映特定群体的特殊需求，适合于具体的当地条件。

目 录

序言	1
1.肺部健康促进.....	7
1.1 如何理解肺部健康促进? 7	
1.1.1 对于健康的理解.....	7
1.1.2 什么是肺部健康?	7
1.1.3 如何促进肺部健康?	8
1.2 什么是研究, 为什么它非常重要? 8	
1.3 如何使科研优先? 9	
1.3.1 与其他疾病比较, 待研究疾病的发生频率如何?	9
1.3.2 什么是疾病所导致的残疾或失能等级?	9
1.3.3 是否存在符合成本—效益原则的治愈、控制或预防肺疾病的方法?	10
2.流行病学, 肺部健康促进的基础学科.....	11
2.1 什么是流行病学, 如何应用流行病学? 11	
2.1.1 流行病学简史.....	11
2.1.2 定义.....	12
2.1.3 流行病学研究的特征.....	12
2.2 流行病学的应用 13	
2.2.1 测量群体健康水平.....	13
2.2.2 描述疾病的自然史.....	13
2.2.3 探索疾病的影响因素.....	14
2.2.4 疾病的控制和预防.....	14
2.2.5 卫生服务的规划和评价.....	15
2.3 因素间的联系和因果联系 15	
3.着手一项流行病学研究.....	18
3.1 研究方案 18	
3.1.1 为什么要重视研究方案?	18
3.1.2 研究方案的经典格式和要素.....	18
3.2 决策研究内容: 研究课题的选择 26	

3.2.1 研究课题形成中的基本概念.....	26
3.2.2 选择研究问题的要点.....	27
4.科学研究的构建：研究设计.....	29
4.1 研究设计：与医学知识发展的关系	29
4.2 研究结构概述	30
4.3 研究设计的细节	31
4.3.1 描述性研究.....	31
4.3.2 分析性研究.....	32
4.3.3 试验性（干预）研究.....	39
5.研究对象：研究人群的选择.....	42
5.1 确定研究人群	42
5.2 样本选择	42
5.2.1 定义和抽样程序.....	43
5.2.2 如何进行抽样？.....	44
5.2.3 如何保证得出的结论正确（研究的效度）？.....	45
5.3 样本量的确定	45
5.3.1 为什么研究样本量很重要？.....	46
5.3.2 计算研究样本量所需要的信息.....	46
5.3.3 如何利用这些因素计算样本量？.....	48
6.流行病学中的变量测量.....	51
6.1 总则	51
6.1.1 信息收集的原则.....	51
6.1.2 数据类型.....	51
6.1.3 健康状态或特征的测量工具（数据收集）.....	52
6.1.4 有关测量值的问题.....	53
6.2 疾病的记数和健康测量	54
6.2.1 信息来源.....	54
6.2.2 死因研究.....	55
6.2.3 发病率研究.....	57
6.3 调查表的使用和设计	58

6.3.1 调查表的设计原则.....	59
6.3.2 调查表的内容.....	59
6.4 暴露的测量 62	
6.4.1 环境因素暴露的记录和测量.....	62
6.4.2 暴露信息的收集工具.....	63
6.4.3 暴露评价方法的选择.....	64
6.5 发病或死亡的报告：率的计算、危险度指标 65	
6.5.1 率的计算.....	65
6.5.2 危险度指标.....	67
7.研究实施的实际步骤.....	69
7.1 项目执行：假设你的研究方案获得了资助—现在怎么办？ 69	
7.1.1 研究实施计划.....	69
7.1.2 研究实施过程和研究技术的质量控制.....	70
7.1.3 保证研究效率.....	71
7.1.4 严格控制项目的研究内容.....	71
7.2 收集信息的处理 72	
7.2.1 原始数据收集表和信息记录的核对.....	73
7.2.2 数据的编码.....	73
7.2.3 资料的计算机录入.....	74
7.2.4 计算机文件的核查.....	75
7.2.5 缺失数据的处理.....	75
8.研究结果的解释.....	78
8.1 保证结果的有效性 78	
8.1.1 定义.....	78
8.1.2 研究人群抽样过程中的误差和偏倚（选择偏倚）.....	79
8.1.3 变量测量中的误差和偏倚（信息偏倚）.....	80
8.1.4 检测试验或检测工具内在误差的评价.....	82
8.1.5 混杂（选择偏倚的一种特殊类型）.....	83
8.2 了解结果的含义—数据分析 85	
8.2.1 熟悉数据.....	86
8.2.2 统计学分析方法.....	87
8.2.3 用于不进行比较条件下描述结果的统计量.....	87

8.2.4 离散型变量两组间比较的统计学检验.....	89
8.2.5 连续性测量值的统计学检验：两组或多组间的比较.....	93
8.2.6 连续性测量值的统计学检验：两测量值间的关联.....	95
8.2.7 在控制其它变量条件下，进行组间比较的统计学检验方法.....	97
8.2.8 统计学分析中的其它问题.....	101
8.3 结果的报告 103	
8.3.1 研究报告的准备.....	103
8.3.2 科学论文的撰写.....	103
9.研究中的其它问题.....	106
9.1 该项研究的拥有者是谁？ 106	
9.1.1 构建研究责任.....	106
9.1.2 尽量满足社区在研究过程中提出的需求.....	106
9.1.3 保证研究结果的追踪.....	107
9.2 争取有关方面对研究计划的支持 107	
9.2.1 地方资源.....	107
9.2.2 国际资源.....	108
9.3 研究中的伦理问题 109	
9.3.1 基本原则.....	109
9.3.2 在低收入国家中与赞助性研究相关的原则.....	111
9.3.3 伦理学评估的机构.....	112

第一章

1.肺部健康促进

改善社区卫生保健所必须开展的活动，需要由科学研究来提供卫生知识。无论卫生保健研究的性质如何，其最终目标都是为了增进个体以及群体的健康状况。

1.1 如何理解肺部健康促进？

1.1.1 对于健康的理解

在开始讨论肺部健康促进研究之前，我们必须就健康本身有一个明确的概念，并明白健康促进是什么意思。为了明确、统一使用这些术语，除特别说明外，这些术语的定义均出自一个来源⁴。

什么是健康？如何测量健康？世界卫生组织将健康定义为“*健康是一种躯体、心理以及社会的完好存在状态，而不仅仅是没有疾病或虚弱*”。该定义由于缺乏精确性难以测量而受到批评。

本教材中的研究对象包括引起发病和死亡的疾病，健康服务的结构及其利用，还包括那些可能对健康产生影响的政策。

- 疾病是指躯体或心理的失能状态。
- 健康服务是指由卫生专业人员提供或在卫生专业人员指导下进行的，为了促进、维持或恢复健康状态的服务。
- 政策是指那些对健康产生正面或负面影响的社会心理及政治性问题的结合体。

1.1.2 什么是肺部健康？

肺是个体与外界环境之间一个巨大的脆弱的界面。其功能是从环境中汲取氧，氧对于维持生命非常关键；肺的另外一个功能是排出废气，这些气体如果滞留在体内就会威胁生命。因此在肺与环境之间的界面上必须存在最低限度的载体以运转这些气体。任何损害上述气体交换效能的情况都将损害肺功能。这种损害可发生于许多疾病过程。

绝大多数肺疾病是由于暴露于环境中的有害成分而造成的。肺疾病的两大类型是指由于

4 Last JM ed. A Dictionary of Epidemiology 3rd ed. New York: Oxford University Press,1995, pp 180.

暴露于微生物病原体而发生的疾病（肺炎和肺结核），以及由于暴露于颗粒物、烟尘和有害气体所造成的疾病（慢性阻塞性肺疾病、哮喘、肺癌以及肺纤维化疾病）。

肺疾病的预防和处理以及肺部健康促进依赖于我们对个体与其周围环境中多种有害因子间相互作用的清晰理解。

1.1.3 如何促进肺部健康？

健康促进是指通过开展针对影响健康因素的卫生活动，提高民众控制以及促进其自身健康的过程。

健康促进的工具是卫生服务，包括临床型及预防型卫生服务。因此就我们的目标而言，必须考虑卫生服务的机构及其效能。

许多国家关于卫生服务机构及其运行机制的研究已经将重点导向于卫生服务的技术效能、责任性以及投入—产出的关系上。以最低的投入获得及时的和最佳的卫生服务是现代卫生服务的追求目标。

但是，什么是最佳卫生服务？如何对卫生服务进行监测以保证其在执行过程中维持良好状态？强调成本—节约是否会导致卫生服务质量的下降？这些问题对于当今所有卫生服务和公共卫生工作人员而言都是至关重要的。

促进、维持或者恢复肺部健康的卫生服务往往是全科卫生服务的一部分，此类服务极少只涉及肺部疾病（高水平的专科服务除外）。发展或评价此类服务的研究工作（卫生服务研究）应包括以下几个领域：

- 卫生服务机构，涉及资源、设备以及人员，
- 卫生服务过程，涉及在哪儿、由谁以及如何提供卫生服务，
- 卫生服务产出，涉及服务的数量和特征，
- 卫生服务结局，涉及结果。

1.2 什么是研究，为什么它非常重要？

研究的目的是为了创立对于健康促进的卫生活动至关重要的知识。如果没有这种知识，由于缺乏逻辑或经验性基础，就不可能开展任何有效的健康促进活动。实际上，对于持续性卫生活动而言，如果不包括一个研究规划在内，常常会变得脱离实际，其结果不是具误导性就是存在投入浪费的问题。因此国家结核病规划（其他公共卫生项目的先导）的初始概念中就把研究规划作为其重要成分。

科学研究是一个永无休止的探索行为。虽然公共卫生实践是建立在调查、标准化以及系

统实践的基础之上，但科学研究需要怀疑一切的头脑，要时刻准备对周围的事物进行评价并提出疑问。当这种探索和评价纳入一个系统的框架时，就会产生出新知识，而这种新知识是开展卫生活动以及对卫生活动进行持续性修正所必需的。这就是科学研究，这就是科学研究为什么重要的原因。

1.3 如何使科研优先？

在医学以及卫生服务当中建立科研优先时，需强调下列问题：

1.3.1 与其他疾病比较，待研究疾病的发生频率如何？

就全球的健康负担而言，肺疾病仅次于所有其它传染病的总和排在第二位⁵。肺疾病占所有疾病病例数的 1/8 强，最常见的肺疾病每年致死人数在一千万以上。这些常见肺疾病包括：

- 急性呼吸道感染，是 5 岁以下儿童最常见的死因；
- 结核病，是 15-49 岁青壮年最常见的单一死因；
- 吸烟导致的疾病，其造成的死亡与结核病不相上下。在许多工业化国家，吸烟导致的死亡占全部死亡的四分之一之多；
- 空气污染导致的死亡，仅在亚洲的大城市每年就达到百万之众；
- 失能性气道疾病（慢性阻塞性肺病和哮喘），这是由于暴露于矿尘、有机尘、烟尘以及致敏性化学物质所致，在新兴工业行业尤甚。

急性呼吸道感染和结核病所导致的死亡几乎都发生在低收入国家。空气污染以及职业暴露所导致的死亡在低收入国家也很常见。虽然吸烟所导致的死亡目前还大部分发生在工业化国家，但是吸烟所造成的健康负担在低收入国家正在迅速增长。

1.3.2 什么是疾病所导致的残疾或失能等级？

某些肺疾病不是引起死亡的主要原因，但由于其多发性或长期性而造成诸多残疾，并因此加重了卫生服务的负担。

- 急性呼吸道感染是初级卫生保健水平卫生服务需求的三大原因之一；
- 目前世界上有超过 1 亿的哮喘病人；
- 哮喘病例的不正确管理造成对卫生服务的滥用，并导致巨大的资源浪费；
- 结核病人以及急性呼吸道感染病人的不正确管理是导致耐药性的主要原因。

5 World Bank. World Development Report 1993: Investing in health. World Bank, Oxford University Press, Oxford, 1993.

1.3.3 是否存在符合成本—效益原则的治愈、控制或预防肺疾病的方法？

绝大多数肺疾病的发病或恶化是由于环境因素的暴露引起的。因此从理论上讲，肺疾病可以预防。例如，

通过如下措施保证空气质量：

- 控制烟草，预防主动和被动吸烟；
- 控制由于化学燃料和汽车尾气所导致的周围空气污染；
- 减少工作场所的有毒物质暴露。

对于某些传染性肺疾病，可采用以疫苗为基本手段的预防措施；对于许多肺疾病，可以运用标准化的病例管理规则。

- 结核病免疫接种是世界上推行最广泛的预防接种；
- 针对细菌性肺炎最常见病原体的有效疫苗，已经使用或将要使用；
- 标准化的结核病例管理是低收入国家中最符合成本—效益原则的卫生干预措施；
- 五岁以下儿童肺炎病例的标准化管理能够降低死亡率并减少抗生素的不当使用；
- 对于哮喘病例的标准化管理可减少不必要的卫生服务消费，降低卫生服务成本。

上述所有这些知识都是从良好设计的科学研究中获得的，这些知识为开展促进社区健康的卫生活动提供了必需的科学依据。

第二章

2. 流行病学，肺部健康促进的基础学科

流行病学为医学研究提供了适宜的方法学，因而人们将其作为健康促进的基础学科。为了充分发挥流行病学的作用，将其优势发挥到卫生研究中去，我们必须了解什么是流行病学以及如何应用流行病学。

2.1 什么是流行病学，如何应用流行病学？

流行病的本质是从群体角度研究健康和疾病，而不是个体角度。它不仅研究疾病的有效诊断手段和治疗措施，而且还研究疾病的病因和预防措施。

流行病学需要对研究对象以及相关事件进行计数，并按照其一般特征进行分类。其价值则严格依赖于测量的准确度。本手册将对有关测量技术进行叙述。

2.1.1 流行病学简史

在许多历史文献中都有将流行病学方法应用于研究疾病以及健康促进活动的事例。希波克拉底在其论文“空气、水和地区”中，鼓励其追随者要积极考虑生活环境以及生活方式与疾病的关系。更古老的希伯来法律已经根据流行病学基本原理规定公共卫生行为。

在 17 世纪的英格兰，John Graunt 是将定量描述方法应用于医学的先驱者。他利用发表的死亡率资料以及教区关于出生和死亡的登记资料计算出该地人群的年龄、性别以及病因别死亡率，该结果显示了不同疾病的相对重要性以及当时的流行病对人群健康的影响。在 1836 年英格兰建立死亡证明制度后，William Farr 作为第一位官方医学统计学家，通过分析可得到的数据，对人群的健康状态及其影响因素进行了全面描述。该结果可以直接运用于疾病的控制与预防。

19 世纪中叶，John Snow 通过对伦敦霍乱死亡数据的分析，得出最可能的病原来源是污染饮用水的结论，并提出了有效的预防措施。该发现的提出远在霍乱病原体分离之前。

流行病学原理同样可以应用于治疗及预防措施的效果评价。现在所有的新药或疫苗在上市之前必须通过随机对照实验（RCT）的严格考核。在 1948 年发表的关于链霉素治疗结核病效果的实验研究，作为首项 RCT 而备受信赖⁶，但在早期也有一些不严格的实验

6 MRC Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment of pulmonary tuberculosis. Brit. Med. J. 1948; 2; 769-83.

研究。例如在 1747 年，一位苏格兰医生 James Lind 完成了一项实验研究，证实柑橘类水果在预防和治疗水手中坏血病的价值。

在近 50 年中流行病学概念已经超越了传染病及其病因和控制研究的范畴。Sir John Ryle 在 1948 年预测“流行病学的研究范畴将发展”到包括诸如心血管疾病之类的慢性病、癌症和肺疾病。今天流行病学已经发展到了这种地步。另外，流行病学研究方法已经应用到对健康服务需求以及对卫生服务的效果和效率的评价上来了。

2.1.2 定义

关于流行病学及其相关术语有许多种定义，它们分别侧重于研究对象的不同方面。下面的定义出自 Last's 流行病学字典，本书采用了此定义：

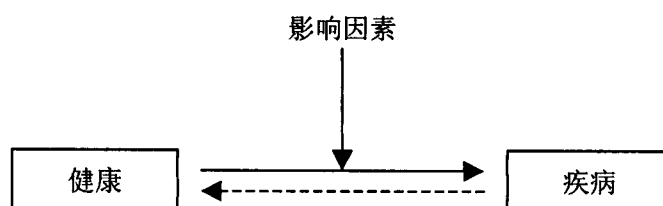
流行病学是研究特定人群中健康相关状态或健康相关事件的分布及其影响因素，并将研究结果应用于控制健康问题的学科。

本定义中有关术语的含义如下：

- 研究包括监测、观察、假设检验、分析性研究以及流行病学实验。
- 分布是指按照时间、地点以及研究人群的社会地位所进行的分析结果。
- 影响因素是指所有能够影响健康状态或能够导致健康相关状态发生变化的物理、生物、社会、文化以及行为因素。（许多文献运用“暴露”一词代替“影响因素”）
- 健康相关状态包括疾病、行为、死因、政策以及卫生服务的结构或利用。
- 特定人群指具有某种明确特征的人群。
- 应用于疾病控制说明流行病学研究的目标—促进、保护和恢复健康。

2.1.3 流行病学研究的特征

流行病学研究运用于肺部健康和运用于其他疾病一样，需要假定一个从健康到疾病或从疾病到健康的转换模型，该转换过程为影响因素所促进或延缓。这就反映了由于暴露于有害物质或微生物病原体而导致的患病过程，例如它可以反映由于治疗而导致的疾病消失（治愈）或缓解（好转），也可以反映由于环境干预措施而导致的疾病预防效果。



这个图形描述了流行病学研究的方法，流行病学研究的主要目的是掌握疾病影响因素的

更多信息，以便有效地预防或治疗疾病并提供最佳的卫生服务。

流行病学的具体目的是：

- 描述人群中不同健康相关状态及其决定因素的分布并判断其相对重要性（描述性研究）；
- 探索疾病的主要影响因素，确定疾病自然史（分析性研究）；
- 评价疾病预防、治疗以及缓解措施的效果、作用和效率（干预或实验研究）；
- 对卫生服务的过程和结局进行评价（卫生服务或实施性研究）。

2.2 流行病学的应用

流行病学的应用包括关于卫生服务规划和评价的实践性应用以及对疾病病因、管理以及预防手段的深入理解两个方面。下面就是关于此类应用的实例：

2.2.1 测量群体健康水平

对于疾病发生频率以及影响因素的直觉估计或印象很可能具有误导性，因为许多因素会影响这种判断结果，如医生的个人兴趣、病人的就医途径以及健康问题的公众意识等。因此需要使用更严格的方法：

- 测量特定群体的疾病负担
如了解卫生服务需求，确定资源分配的相对优先权；
- 监测疾病的发病率和患病率变化趋势
如比较不同群体哮喘和结核病的变化趋势；
- 确定疾病及其结局的特征变化
如对耐多药结核病的追踪观察；
- 确定高危人群以制定适宜的干预措施
如确定环境中引发哮喘的因子；
- 确定健康状态、功能程度或损害程度
如测量职业暴露对肺功能的影响。

2.2.2 描述疾病的自然史

疾病的自然史在不同个体可能存在显著差别。因此对任何个体或少量病例的观察结果都不具有代表性。同样通过回顾病史也不可能准确地了解到疾病临床前期的特征。对于以非选择样本为研究对象进行的流行病学研究，要求完成以下任务：

- 确定期望值的正常值范围
如特定群体的肺功能正常值；

■阐明临床症状谱并确定其诱因

如对于急性呼吸道感染和哮喘建立严重程度分级标准，以指导治疗、预测结局；

■揭示疾病临床前期的发生频率及其病程

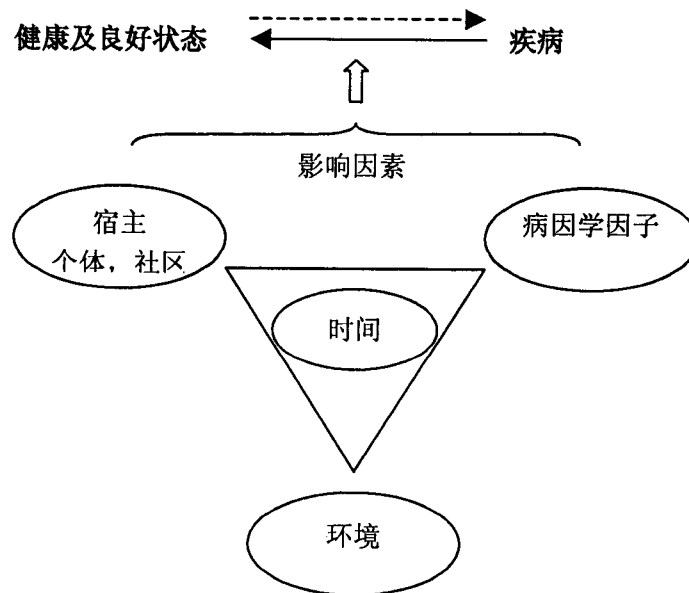
如对于非结核病人的结核分枝杆菌感染研究；

■预测干预条件和无干预条件下的疾病进程（预后）

如对结核病预防性化疗的临床实验，哮喘治疗的临床实验。

2.2.3 探索疾病的影响因素

疾病的影响因素常常是模糊的、多因素的。多数流行病学研究的重点是确定疾病的影响因素，并建立这些因素与健康相关状态之间的联系。影响因素指能够影响健康状态或能够导致健康相关状态发生改变的因素，包括物理、生物、社会、文化以及行为因素等，见下图。



2.2.4 疾病的控制和预防

只有对个体与环境之间的联系进行准确、特异的干预，才能达到对疾病的有效控制。

可以通过以下四种途径实现对个体与环境之间联系的准确、特异的干预：

- 去除或破坏主要的病原和影响因素。这里需要了解病原自然储存宿主（传染源）、传播模式（暴露）以及作用部位等知识

如结核病治疗在减少结核菌传播中的作用；用其他类型油漆替代含异氰酸酯油漆

以预防哮喘；

- 通过环境控制（环境卫生）保护个体避免暴露于有害因子
如控制疾病的医院内传播，改善通风以降低粉尘和有害气体的职业性暴露；
- 加强宿主的抵抗力（保护机制、免疫力）
如良好的营养摄入、被动或主动免疫接种；
- 改变个人生活方式以避免健康危害行为、促进健康行为
如提倡不吸烟，当必须暴露于有害因子时使用呼吸保护设备。

在确定疾病的控制或预防方法时应选择：

- 面向描述性研究中发现的高危人群；
- 针对分析性研究中发现的不同重要程度的因素；
- 采用实验性研究中发现的有效措施。

2.2.5 卫生服务的规划和评价

在过去，卫生服务的计划和卫生资源的分配是决策者根据以往的需求和消费情况，通过“判断”以及特别申请平衡后作出决定的。对卫生服务进行合理规划及有效管理依赖于：

- 通过对服务社区的发病率和死亡率进行准确描述，估计其卫生需求；
- 确定主要的可避免的健康危害；
- 确定其技术以及财政资源的供应能力；
- 对卫生服务消费的准确测量；
- 对干预措施效果方面的了解；
- 对卫生服务的作用和效益评价。

只有通过流行病学研究，才能获得开展有效健康促进活动所必需的信息（疾病的范围及其分布、疾病的病因以及干预措施的效果）。

2.3 因素间的联系和因果联系

观察性流行病学研究有时可以确定某可疑因子与某特定疾病或健康相关状态之间的关联。这种联系甚至具有统计学“显著性”（即按照概率标准，因素间的联系强度已不能用偶然性来解释）。例如某项研究的结果显示，在两种不同地理环境的人群中其肺炎死亡率差异具有统计学显著性。

但这种观察到的联系也可能并不是由于地区内某些特殊因子所造成的因果关联。这种死亡率差异也可能是由于研究方法上的错误或偏倚造成的。

理想的因果关联证据应该是流行病学实验性研究的结果，但由于伦理学方面的问题，在

人群中开展流行病学实验常常行不通。因此，当我们检验流行病学研究的结论时，常常需要根据多个研究的证据来评价这种联系是否为因果联系。

有多种类型的证据可用于评价因果关联。Austin Bradford Hill 提出了评价因果联系的标准⁷，这些标准随后得到了进一步的发展。虽然我们不能根据任何一条标准就可以下结论，但数项标准结合起来就可以增加因果联系的说服力。

下面是流行病学字典中所列举的因果联系判断标准：

联系的一致性

在不同环境应用不同方法重复某观察性研究时，就疾病与可疑因子之间的联系可以获得一致性结论。换言之，多项研究发现了相似的联系。

联系强度

指关联的大小，例如以危险度比形式测量的关联（或组间差异）大小。如果某因素存在时某疾病（或健康相关状态）就非常多见的话，那么我们就说这种联系很“强”。反之如果某因素存在时，某疾病只是略微多见的话，我们就说这种联系“很弱”。这个概念应该与“统计学显著性”区别开来。不管是强的联系还是弱的联系都可能具统计学显著性。

联系的特异性

如果某可疑因素（或某些可疑因素）总是产生同样结局的话，我们就说这种联系具有特异性。

剂量反应关系

指随着某因子或影响因素暴露水平的增加（数量和/或时间上），某疾病的发生危险度也增加。

明确的时序关系

对于某影响因素的暴露一定要发生在结局事件出现之前。尽管暴露日期有时很难确定，但这的确是一条重要标准。

生物学合理性

这种联系要与当前关于该可疑因素与病理学过程的知识相一致。应当慎用这条标准，因为迄今为止该相关过程可能尚不清楚。

与现有知识的连贯性

这种联系应该与该病自然史方面的现有知识相一致。

实验研究的证据

适宜的实验性干预措施（如从环境中去除可疑因素）可以预防或缓解该疾病。

7 Hill AB. The environment and disease: association or causation. Proc. R. Soc. Med. 1965; 58: 295-300.

参见关于病因与联系的相关章节练习内容。

3.着手一项流行病学研究

3.1 研究方案

3.1.1 为什么要重视研究方案?

在实验室研究中，科研人员一般按照下列步骤运用实验性研究设计：

- 提出假设；
- 写出研究方案或研究程序；
- 将实验程序细致地无差别地施加于实验组或实验系列；
- 设置一个对照组或对照系列，除没有接受实验程序外，其他条件等同于实验组；
- 比较实验组和对照组的结果（结局）。

与简单的、无对照观察研究的结果相比，按照上述方法进行的实验室研究其结论更科学。在下面章节将介绍经典研究方案的要素。这些要素将在讲义的其他章节进一步讨论。

3.1.2 研究方案的经典格式和要素

在以人群为基础的研究中，虽然相当部分是观察性研究而非实验性研究，但鉴于类似实验室研究中的理由，拟订研究方案并按照该方案完成研究仍然十分重要。这样的操作程序将提高研究结论的科学性。

1. 摘要
2. 研究介绍
 - a) 研究问题
 - b) 立题依据, 研究现状
 - c) 研究目的和研究内容
 - d) 研究设计以及研究方法
 - 研究设计
 - 研究人群
 - 样本大小以及统计学检验效力
 - 研究对象: 选择和定义
 - 数据收集方案: 测量值, 定义
 - 数据管理和统计学分析
 - e) 项目管理
 - 所需要的研究人员
 - 研究时间
 - 研究结果的随访程序 (需要的话)
 - f) 研究实力及局限性
 - g) 参考文献
3. 伦理学问题的考虑
4. 研究意义 (或期望的影响)
5. 研究预算
6. 研究人员: 每个人的分工及履历表

本教程就是要帮助你制定、书写 (并且最终执行) 人群基础的科学研究方案。

研究方案在制定阶段可以自由修改。一旦制定, 则须严格执行。

摘要

摘要要简洁, 但必须能够使读者了解研究的主要目的、如何实施以及期望收益如何。摘要是研究的纲领性计划书, 它帮助评阅者在检查研究计划的细节之前, 能够了解研究的全貌。虽然研究摘要要放在研究方案的开始部分, 但是起草摘要却要在研究方案完成之后进行。

研究介绍

研究介绍部分要提供所有必须的信息以便于评估者检查下列问题:

- 你想回答什么问题, 这个问题的期望答案 (或研究假设) 是什么? 这个问题 (和研究假设) 重要吗? 如何利用该领域的现有知识解决这个问题?

- 本研究方案能够回答这个问题吗？
- 利用研究方案中的设计和方法能够实现既定目标吗？
- 利用方案中指定的资源能够完成该研究吗？
- 当研究结束并完成结果分析后，就该研究问题是否有足够的统计学效力得出有效结论？
- 研究发现对人群健康有益吗？是直接有益还是间接有益？

在制定研究方案的过程中，要时时刻刻记住上述问题。你得把自己放到评阅人的位置上！

前面提到的研究方案要素，应该按照下列问题的答案形式对整个研究进行解释说明：

- WHY？ 提出研究问题以及相关背景信息；
- HOW？ 叙述研究设计以及选择该设计的理由；
- WHO？ 确定目标人群、研究人群以及样本量；
- WHAT？ 确定测量变量，测量仪器以及待分析的结局；
- SO WHAT？ 研究结果的期望意义及其对相关知识的贡献。

研究问题

研究方案应当始于对研究问题清晰、准确的描述。将这种描述写成问题格式可能要比陈述格式的效果好。

例如：

为什么伊斯坦布尔儿童的哮喘患病率很高？

立题依据、研究现状

本部分的目的是叙述如何根据现有知识提出该研究问题。需要按照一定的逻辑顺序将你的思想发展过程叙述出来。要简洁，列出关键文献，而不是对相关文献的全面综述。

- 讨论题目的重要性；
- 对相关文献以及现有知识进行综述（包括现有知识的缺陷之处，进一步说明开展本研究的必要性）；
- 描述你在该领域已经获得的结果；
- 说明你是如何提出该研究问题的，与前面所述内容的逻辑关系如何；
- 从大体思路上描述你将如何解决该问题；
- 解释该研究如何增加知识、促进健康以及/或节约成本。

研究目的，研究假设以及具体步骤

研究目的

对一项研究需要回答的问题而言，即使是一个非常准确的研究问题，通常也还是显

得太广泛。例如，如果研究问题是“为什么伊斯坦布尔儿童的哮喘发病率非常高？”你不能指望一项研究就能解决该问题的所有答案。因此在具体的研究中，你必须将这个问题分解成多个目的。

例如：

本研究的目的是探讨伊斯坦布尔哮喘病例超常之多是否与遗传倾向性（以遗传性过敏症的程度估计）、社会经济状况和/或室内空气污染的联合作用有关。

研究假设

为了实现这些目的，研究计划需要包含多个组疾病率或暴露率的比较，并利用统计学检验对比较结果下结论。如果是这样的话，这些目的还应该写成检验假设的形式。这些假设应该写成待反驳的形式，即所谓的无效假设。

例如：

下列假设有待于检验：

1. 在伊斯坦布尔，无论社会经济状况如何，儿童的哮喘患病率没有差别。
2. 随着家庭内空气污染程度的加重，儿童的哮喘患病率并没有升高。
3. 哮喘患病率与社会经济状况之间的联系以及哮喘患病率与室内空气污染程度之间的联系在不同程度遗传倾向性的儿童组别间并无差别。

具体步骤

到这里你就可以把研究组需要完成的操作步骤总结出来了，这些步骤是实现研究目的所必须的。这就是所谓的*具体步骤*。

例如：

为了实现这些目标，确定了如下具体步骤：

1. 按照一个具体的病例定义，确定儿童哮喘病例的适当来源，并选择 200 个病例。
2. 确定并选择适当的对照个体（非哮喘病患者）。
3. 采用标准化调查表调查记录病例和对照的个人、人口统计学以及社会经济状况等方面的信息。
4. 对病例和对照进行皮肤变态反应性试验的检测。
5. 对于每个研究对象，随机选择三天测定其家庭内的微尘粒暴露数据。
6. 比较病例与对照之间遗传过敏症、低社会经济状况、以及室内空气污染较严重等因素的危险度比。

很明显，你不可能马上就能完成具体步骤这一部分。你必须首先要选择好研究设计和研究方法，然后才能列出具体步骤。将具体步骤安排在这里，只是为了让评阅人

能很快地容易地理解你的研究计划。

研究设计和研究方法

研究设计

这一部分主要用于阐述你所选择的研究设计。记住要根据研究目的选择研究设计。因此你需要解释所选择的研究设计与其它研究设计相比有什么优势。

研究人群

在这一部分中要规定本研究的适用范围。即当获得研究结果后，这些结果适用于那些对象？哪些人群存在所研究的疾病？是所有的人群吗？还是只有处于某种地理条件、职业或年龄组的人群才会受到本疾病的影响？本部分还得交代如何肯定研究结果可以推广到所确定的研究人群中去。

例如：

- 我们要研究伊斯坦布尔的哪些儿童？所有儿童？某所学校的儿童？去医院就诊的哮喘儿童？还是报告有哮喘样症状的儿童？
- 根据本研究人群所获得的结果能够为我们提供关于所有儿童的信息吗？能提供关于土耳其所有儿童的信息吗？
- 儿童指“谁”？是16岁以下者，还是2岁以下者？对于儿童这个概念有无设定上下限年龄？

样本大小和统计学检验效力

估计需要多大研究样本才能回答所提出的问题。指定计算样本量的假设条件，并给出根据不同假设计算出来的相应样本量列表。

例如：

- 需要召集多少儿童？
- 他们代表了整个人群的多大比例？
- 该儿童数对于回答研究问题足够了吗？
- 儿童中研究因素（空气污染，遗传过敏性）暴露率的变异有多大？
- 在暴露于研究因素与不暴露于研究因素的儿童中我们能够检测到多大程度的结局差异？

研究对象：选择和定义

这一部分应包括：

- 详细解释本研究需要召集多少个体，从哪儿找，为什么；
- 关于合格对象标准、纳入标准、剔除标准以及中止观察标准的定义；

- 对潜在合格个体数的现实估计；
- 对研究对象抽取方法的描述；
- 讨论召集到规定数目研究对象的可行性，并估计所抽取样本中同意参加研究的个体比例。

例如：

- 研究将包括患有（不患有）哮喘的儿童吗？
- 要剔除哪些儿童？
- 如何找到这些儿童？
- 如果他们拒绝参加研究你怎么办？
- 有多少儿童是合格的研究对象，你希望实际参加研究的儿童比例达到多高？

资料收集方案：定义和测量值

说明如何收集数据、确定健康结局（疾病或其它健康相关状态）及其影响因素是很重要的。这就是说在设计书中必须明确本研究如何测量或定义上述变量。同时也应该规定质量控制程序。如果该程序是以前讲述过的标准化程序，则标明其参考文献出处。

上述内容包括：

- 精确规定所有术语的定义；
- 针对研究方法和测量仪器，考虑进行预实验；
- 讨论所作定义的准确性和可靠性；
- 讨论测量工具以及所作定义的限制性（即误差或错误分类的影响）。

例如：

- 在本研究中如何确定哮喘病例：调查表、临床检查还是通过查阅病人病历、住院登记？
- 如果某个非哮喘儿童被误判为哮喘病例，会产生什么样的结果？
- 本研究中室内空气污染的概念是什么？如何进行测量或估计？
- 准备对室内空气污染中的哪种成分进行评价？
- 如何确定研究对象的社会经济状况？
- 本研究中遗传性过敏症的定义是什么？

数据：管理与统计学分析

本部分内容将讲述：

- 将调查数据编码以及录入成计算机文件的过程；

- 信息完整性和准确性的保证措施；
- 结果报告以及组间比较示例；
- 对所有假设进行检验的统计学分析方法。

如果将这一部分内容按照研究问题、研究目的以及研究假设的顺序组织编排，效果会更好。

还须给出关于统计学检验方法的参考文献：包含统计学检验方法的统计学书籍或文献或者所运用的统计学软件。

例如：

- 如何将分析内容从病人记录（或者调查表）中提取出来？
- 如何保证对调查表资料的准确编码？
- 用什么样的表格、图形以及数据来表达结果？
- 用哪种统计学检验方法进行假设检验？

项目的全局管理

这部分将规定：

研究人员：实施研究所需要的人员及其任务。要按照需要完成的任务以及所需时间说明研究人员数的合理性。要具体规定每个研究成员的责任。

研究期限：规定以下各研究阶段所需要的预期时间：

- 预试验；
- 选择研究对象；
- 调查表以及其它资料的准备；
- 数据收集；
- 随访；
- 数据核对和统计学分析；
- 结果报告—向研究对象、资助方、所涉及的社区以及学术机构报告研究结果。

并在适当之处给出对研究对象的随访程序（即向每个研究对象都通报研究结果吗？如果你发现某个未接受治疗的个体出现了临床相关症状该怎么办？）

研究优势和研究局限性

在制定研究计划时，研究设计以及信息收集方法是折衷选择的。完美的研究设计几乎不存在，多数研究方案均有不当之处。所以在研究设计中要专门留一段来阐述本研究设计、研究方法可能面临的批评，并说明为什么你认为自己的选择所带来的局限性并不严重。同样，也很有必要把那些你认为特别有意义，值得财政支持的方面强调出来。

参考文献

本部分将列出上述各部分（包括立题依据、研究现状、标准化方法以及统计学分析）的参考文献。一般按照文献在研究方案中的出现顺序进行列示。

伦理学问题的考虑

在本部分必须详尽叙述本研究设计已经充分考虑并将严格遵守赫尔辛基宣言中的基本条例。需要说明：

- 如何保证技术方面的质量；
- 研究程序的预期损害以及预期受益；
- 进行该研究的依据及其合理性；
- 研究对象的利益如何优先于所谓科学或社会的利益；
- 如何维护研究对象的利益；
- 研究对象受到损害时研究方需要承担的责任；
- 如何向研究对象通报研究事宜；
- 如何获得他们的知情同意书。

例如：

- 这些儿童在接受皮肤过敏试验时会发生危险吗？
- 谁为儿童提供参与同意书？
- 谁将了解检查结果？
- 如果发现了疾病将采取什么措施？

如果该研究是由社区外机构甚至是国外机构资助的话，就必须详尽说明研究各方的利益以及研究对象、当地研究者和社区能够从研究中获得的受益。

- 研究结果属于谁？
- 谁将是有关论文的作者？
- 资助方如何处理研究结果？
- 研究中诊断出的病例能得到适宜的治疗吗？

研究意义（预期影响）

本部分将以预期结果的形式再一次强调研究的必要性和合理性。这里需要说明：

- 预期结果的意义；
- 本研究组成员、其它研究者、决策者以及社区将如何利用本研究结果。

预算

必须详细列出研究执行过程中预期的每一项支出，即使是被常规卫生服务或研究外其它卫生资源所覆盖的项目成本也应列出。要尽可能按照单位价格列出支出数目（如根据人员资格所支付的小时工资）。

预算部分应该包括一份详细解释各种支出费用的书面预算证明。

研究成员：每个人的任务及其履历表

本部分要说明每个研究者在研究中所承担的角色，并明确研究各部分的责任人。

履历表要明确交代研究者的资格和经验，包括所受的培训、学位、研究经历以及已发表的科学论文。

这一部分还要交代研究者的其他义务。目的是要使评阅者相信研究者有足够的经验和能力完成课题，要使评阅者明白研究者有足够的时间来完成该研究。

3.2 决策研究内容：研究课题的选择

选择研究课题是研究工作最重要的第一步。选题是否恰当不但关系到能否书写出清晰的研究方案，而且将决定你能否获得经费支持完成研究。因此选题是研究工作中不可缺少的重要部分。

3.2.1 研究课题形成中的基本概念

研究课题常常以问题的形式提出，为什么呢？

问题

开展科学研究的目的是扩展相关知识。因此为了形成创造知识的新途径，关于待探索问题以及探索方法的描述必须准确。表达准确就是要按照研究方案之所述内容进行陈述。

可以将研究课题改编成特定问题的形式，这个问题必需能够明确地表达特定的研究目的。该问题越准确，研究工作越有可能创造新知识。将研究课题改编成问题形式，也能较容易地列出探索问题答案的步骤来。如果连课题表达都不够清晰准确的话，自然也不可能对如何获得新知识有清晰的认识。

例如：

假定我们对曼谷的耐多药结核病感兴趣。我们就可以提出这样的研究课题： 曼谷的耐多药结核病，或 曼谷的耐多药结核病是由于不合理的病例管理造成的吗？

将课题改编成问题形式的优点是比平淡的描述表达清晰。将研究课题改编成问题形式要求对研究课题的逻辑判断方面有更多的精确性和线索。

研究假设

如果对研究问题的回答要涉及多组间的比较或因素间的相关性分析，那么研究假设必须符合研究问题以及研究目的的表达。提出假设的目的是将研究问题设置成通过一系列统计学检验能找出其答案的格式。

研究假设应按照涉及组间比较或因素关联分析的陈述形式提出，而不应以问题格式提出。传统的陈述格式是将其表达成准备拒绝的假设格式。这就是所谓无效假设。这么作的理由是绝大多数统计学检验所提供的结果是拒绝无效假设时所犯错误的概率。在规定的概率水平下拒绝无效假设将使我们更加相信研究因素（健康相关状态、因子以及影响因素）间存在联系。

例如：

- 研究问题—曼谷的耐多药结核病是由于不合理病例管理造成的吗？
- 研究目的—对一组接受不合理病例管理的病人与另一组接受合理病例管理的病人，进行耐多药发生率比较。
- 研究假设—在曼谷接受合理病例管理病人的耐多药发生率与那些接受不合理病例管理病人相比没有差别。

将研究课题改编成一个问题，再将问题分解成更精确的研究目的，再形成研究假设，我们对研究内容的认识会越来越清晰。研究问题设定了研究的基本结构，研究目的则要求我们精确操作，提出诸如耐多药、结核病、病例管理、病例管理质量以及曼谷之类的关键词来。

研究假设的提出迫使我们认真考虑需要进行哪些比较来回答研究问题，研究假设的提出也同时建立了在解释结果时如何应用统计学检验的格式。在这种情况下，通过比较接受不同病例管理的病人组别中耐多药的发生率，我们就可以建立不良病例管理（影响因素）与耐多药结核病（健康相关状态）之间的联系（或认为不存在这种联系）。在确定这种联系的特征时，我们可以评价此联系的强度（即不同组别间耐多药结核病发生率之间的差异有多大），并估计本研究发现的差异的可信度。这是考虑因果关联的两个关键要素。

3.2.2 选择研究问题的要点

在某些情况下，科学被认为是纯粹的“智力”活动，应该强调其追求知识的特点。但是我们不是生活在一个有无限资源的世界里，我们必须建立研究优先权。即便是现有资源

也常常是通过模糊的机制和网络进行分配。因此在选择研究问题时，我们需要了解当今现实，并考虑如何在这种条件下开展工作。

选择研究问题的要点：

- 与一般社区健康的关系（发生频率、严重程度）；
- 对研究发现进行有效干预或研究发现的_{应用}可能性；
- 完成研究以回答研究问题的可行性；
- 研究结果向其它条件的可推广性；
- 研究组成员的兴趣与经验；
- 与政治和社会（社区、科学）需求之间的关系。

参见关于制定研究方案中的研究课题选择相关章节练习内容。

第四章

4.科学研究的构建：研究设计

4.1 研究设计：与医学知识发展的关系

科学研究的目的是获得知识、了解自然世界及其运行机制。这就需要将一系列系统观察按照一定的逻辑顺序组织起来进行比较分析。通过这种比较可以得出解释观察真实性的结论（演绎推理），并形成解释普遍真实性的思想（归纳推理）。然后便可以通过把这些思想运用于其它条件对其进行验证。

医学中此类概念的形成过程一般都按照一系列系统性步骤进行。这种过程常常是从对单个病人（病例）的观察开始。然后收集一定数目的类似病人（病例系列）以确定该疾病的特征。

通过运用标准化的方法记录某个大人群（社区）中关于某疾病特征和自然史方面的信息，就可以全面了解该疾病对该社区的疾病负担（描述性研究）。通过此类工作也可以对疾病的可疑影响因素获得一定程度的了解（产生假设）。通过对同一人群内不同地点或亚组人群的比较来完成对可疑因素的检验，可以提供更具说服力的证据（分析性研究）。

最后，从观察性研究和组间比较研究转移到评价针对人群某特征所采取干预措施的效果上来，并将该效果与另一组没有接受干预措施的人群进行比较（试验性研究）。

试验性研究

此类研究与其它领域研究者所进行的经典实验室研究很类似。二者的区别在于我们的研究对象是人群。运用此类研究设计的最大限制是伦理道德方面的制约。关于这方面的示例是临床或干预性研究。

观察性研究

既然利用人体进行试验性研究常常既不可行又不符合伦理道德，因此绝大多数流行病学研究就只好对人群进行观察性研究（非试验性）。

观察性研究的目的是“系统地观察并测量自然试验的结果”。开展这种研究的难题是怎样找到一种容许研究人员进行此类观察的研究设计。

描述性研究

此类研究常常要借助为其它目的所收集的现成数据（例如疾病的死亡登记数据、传

染病登记数据以及医院的住院资料)对现有状态进行描述,而无事先形成的假设或设计比较。

分析性研究

此类研究的目的是通过对两组或多组间研究变量的比较分析或在同一人群内对不同因子间的相关分析来检验假设。

横断面研究

- 类似于描述性研究;
- 在同一时点上收集特定人群的疾病和影响因素资料;
- 可以通过对不同亚组的比较来检验假设。

队列研究

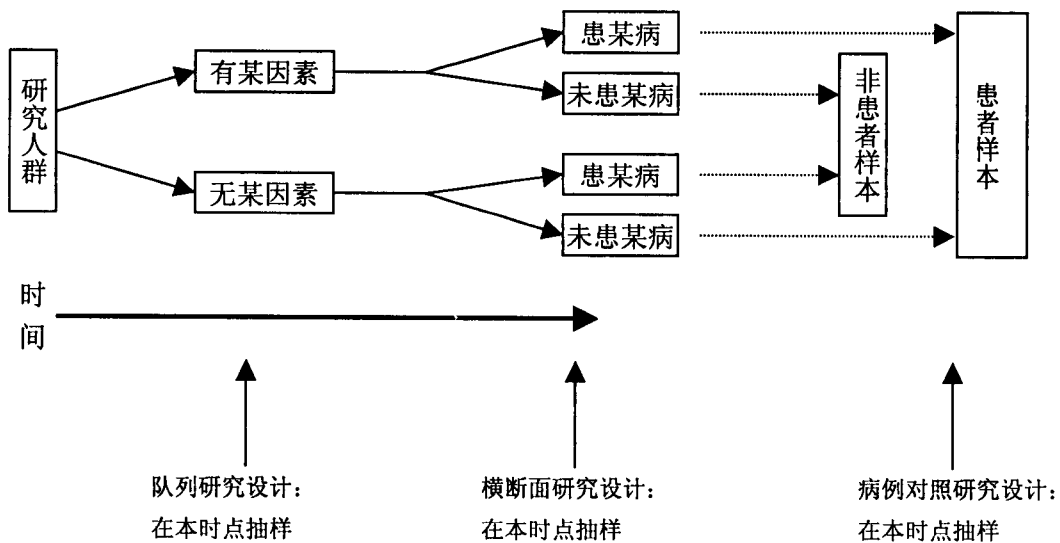
- 确定一个人群,并将其分成两个组,即可疑因素(危险因素、暴露等)暴露组与对照组;
- 对于该两组人群进行随访,确定其发病率并进行比较以检验假设。

病例-对照研究

- 确定患有特定疾病的个体(病例);
- 确定病例来自的人群,并从中选择一个非病例样本(对照);
- 测量病例和对照中各种影响因素的暴露频率(对此进行比较以检验特定假设)。

参见关于研究设计选择的相关章节练习内容。

4.2 研究结构概述



流行病学研究设计的基本框架结构是指通过调查，将有无某健康相关状态（常常指疾病）与有无某研究因素（如治疗、暴露、个人危险因素）联系起来的结构安排。

根据这两个成分（疾病与影响因素）的关系，可以划分为四种组合：

- 1.有影响因素（或危险因子、暴露、治疗），同时患有疾病，
- 2.有影响因素（或危险因子、暴露、治疗），而未患疾病，
- 3.无影响因素（或危险因子、暴露、治疗），而患有疾病，
- 4.无影响因素（或危险因子、暴露、治疗），未患疾病。

4.3 研究设计的细节

4.3.1 描述性研究

*描述性研究*对研究问题提供了一个初始的概述。这有助于我们了解疾病负担，并且有助于卫生保健以及预防性卫生服务的规划。

目的

- 测量指定社区中不同健康问题的相对重要性，并监测其变化情况；
- 描述指定社区内不同疾病或健康问题的发生频率。

研究方法

描述性研究通常是常规来源的发病率及死亡率统计数据或健康相关变量（如肺功能、卫生服务利用情况）数据汇集起来，然后将这些信息按照有意义的分类进行报告，如：

- 时间 （什么时候发生的）
- 地点 （在哪里发现的）
- 人物 （谁受到了影响）

描述性数据常常按照这些标题进行分组，以探索相应情况下疾病患病率的变异情况。

时间 疾病的发病率变化可能有如下几种时间模式：

- **长期趋势**—在多年的时间跨度上表现出的变化；如结核病，哮喘；
- **周期变化**—疾病的发病率以一定时间间隔规律升降；如百日咳、麻疹；
- **季节性变化**—疾病的发病率在一年当中的特定时间升高；如哮喘、流感；
- **流行**—疾病的发病率无规律地短时升高；如军团病。

地点 就某种疾病而言，病例的空间聚集性可能有几种表现形式：

- **地理地貌性**—指疾病的发病率在不同气候、不同人种、不同社会和文化背景以及不同饮食习惯的国家或地区之间存在的变异；如肺癌、哮喘等；
- **城乡差别**—人群密度、贫穷程度、环境污染、饮食以及职业差异导致的发病率差异；如结核病、慢性阻塞性肺病；
- **局部性**—居住条件、环境污染、家庭取暖和烹调的燃料类型以及通风条件等因素的差异，导致发病率即使在同一社区，也可能在不同地点或不同家庭间存在差异；如肺炎、气道阻塞性疾病；
- **机构性**—如果按照组织单位如班级、营房、病房以及车间进行分析的话，即使在同一机构内，如学校、军营、监狱、工厂和医院等，其疾病的发病率也可能存在差异；如急性呼吸道感染、职业性肺病。

人物 不同的个体特征可能通过影响个体对病因因子的暴露机会或者个体的易感性而影响其患病的危险性：

- **年龄和性别**—包括生理以及解剖学上的差异、免疫力、暴露于危险因子的特点和时 间、疾病史以及伤残史、退行性变；如结核病、肺炎、哮喘；
- **婚姻状况**—生活方式、与儿童的接触程度；如急性呼吸道感染；
- **种族**—遗传易感性、环境、文化背景；如哮喘、结核病；
- **家庭**—共同的基因遗传性与生活环境、共同的关系与传统；如哮喘，吸烟；
- **职业和社会经济状况**—对有害物质以及感染性因子的暴露机会、居住地点、出行方式以及经济来源；如哮喘、肺炎、慢性尘肺。

优点

- 可以对现有资料进行低成本、快捷以及符合成本-效益原则的利用；
- 是对研究课题的起始概述；
- 确定进一步研究的参数。

缺点

- 很难发现所有病例，特别是那些非致死性或者较少进行医学处理的疾病；
- 关于某些疾病及其相关变量，可能没有现成的资料可供利用或者其格式不符合研究要求；
- 数据的收集方法和诊断标准非标准化。

4.2.2 分析性研究

根据观察到的疾病或健康状况与影响因素之间的关系，对描述性数据进行分析常可为形成假设提供依据。

形成的假设需要通过严格设计的现场研究来检验：

- 病例是否具有共同特征，或是否暴露于同一因子；
- 暴露于某特定因子的群体其患某疾病的机会是否要高于非暴露者。

分析性研究有三种类型：

- 横断面研究；
- 队列研究；
- 病例—对照研究。

横断面研究

定义

在某一特定时点获取所有测量值的研究方法。利用这些测量值确定病例（或其他健康相关状态）和可疑影响因素（过去或现在的暴露情况、或特征），并对亚组人群进行比较。

目的

- 通过计算一系列可能影响因素（如个人特征、暴露、行为以及治疗措施）与某疾病之间的关联程度，检验其因果关联假设。关联程度的计算有两种方法：
 - 一对研究群体中具有不同类型或不同水平影响因素的亚组人群进行患病率比较，
或
 - 一对具有某可疑影响因素的人群与另一个其他条件相似但没有该因素的人群（或根据该地区患病率计算的期望值）进行患病率比较。
- 提供关于以下方面系统的、可比较的信息，为制订卫生服务规划提供依据：
 - 一 人群中不同亚组的疾病或伤残负担；
 - 一 人群中的高危人群，这是医疗服务和预防性卫生服务的重点人群；

研究方法

必须明确规定研究目的，因为它将决定研究人群和研究方法的选择。

研究方案应该：

- 确定目标人群以及研究样本的抽样方法（见“研究人群的选择”章节）；
- 叙述如何根据研究对象的纳入或剔除标准确定比较组（如病例组/对照组；暴露组/对照组）；
- 提供关于如何测量疾病和暴露以及如何将研究对象分配至疾病组和暴露组（操作定义）的明确说明；
- 规定测量仪器，并对研究人员进行仪器使用和研究程序的培训；
- 制定数据管理及分析计划，包括错误核对以及其它质量控制程序。

优点

- 可以较快捷地获得结果，成本也低；
- 可以探讨多种联系；
- 可以运用标准化手段测量暴露和结局。

缺点

- 研究因素与疾病之间的时序关系常常不清；
- 当研究的疾病为少见病时，需要较大的研究样本；
- 对过去事件的回忆可能存在偏倚；
- 研究人群只能包括特定队列的存活者，这种样本对于目标人群而言是有偏倚的。

队列研究

定义

所谓队列是指一组具有共同经历的个体。队列研究指在某一时刻确定一个未患有研究疾病的人群（队列），然后根据某因素的有和无（如是否暴露于某因子）将其分组。再对他们进行一定时间的随访，以确定暴露组和对照组的发病率。

目的

- 观察某特定因素（暴露于某因子或具有某特征）的暴露者其患病危险性是否高于对照组（非暴露者或无研究特征者），即确定与发病率相关的危险因素；
- 测量归因于危险因素的超额危险度（危险度或率比）。

研究方法

研究队列可能是下列样本之一：

- 一般人群；
- 研究疾病发病率高的人群（如某特定年龄段、性别或职业的人群）；
- 具有某些特征的个体，这些特征利于鉴别，并易于随访（如医生、护士或其它流动性较低的企业职员）。

如果研究计划确定在“今天”选择研究队列，而随访时间是将来，这种队列研究就称之为前瞻性队列研究。

在研究起点要对所有研究对象运用标准化方案（如调查表、标准化检测试验）进行调查，记录其个体特征、疾病史以及相关生物学变量的信息。在随访期间要连续地或在某些时点上测量并记录危险因素（诸如可疑有害因素的暴露）的信息。

前瞻性队列研究中的“对照”常常是“内源性”的，即研究队列中那些未暴露于研究危险因素的个体即组成了前瞻性队列研究的对照组。如果研究队列中的个体都暴露于危险因素（如矿工和煤尘），就需要建立外部的比较组。它可能是该地区的全部

人群，也可能是其它特征相似的非暴露人群。

经过一定时间的随访后，计算每一组的疾病率或疾病平均水平，并进行比较。

如果研究计划要求在指定人群中根据其成员在过去某时点的信息确定队列，然后从那个时点（过去）开始随访到“今天”，那么该研究设计就称之为回顾性队列研究。

研究人群是根据某指定人群中在过去某时点上其成员的信息确定的（利用现有资料—即雇员记录、学校注册登记本、城市名单）。

将研究对象分为暴露组（暴露于研究因素）和对照组所需要的信息也是从历史记录中获取的。这种信息既可能是来自个体自身记录，也可能来自其家庭记录。

对于非致死性结局的研究，需在今天确定研究对象，并测定其疾病状态。对于死亡率研究，则需在今天确定每个研究对象的生命状态（利用公共记录资料），对于已死亡者，则需确定其死因。

计算每个组别的发病率（或死亡率）并进行比较，这种比较常常通过计算两组的发病率比（相对危险度）来实现。这部分内容将在下面章节详细讨论。

优点

- 可以观察到事件间明确的时序关系，易于将病因与关联因子区分开来；
- 可以确定特定群体在规定时间内的新发病例，并计算发病率；
- 可以同时研究某危险因素与几种结局的关系；
- 可以对危险因素与结局进行精确测量。

缺点

- 如果发病率较低，则需要较大的人群样本；
- 如果疾病潜伏期较长，则疾病发生需要较长的时间跨度；
- 研究费用相对较高；
- 研究中的失访数据可能导致结果偏性；
- 在长期随访过程中，标准化方法甚至标准可能会发生变化。

例如

研究目的:

调查木料烟尘暴露在儿童哮喘发病中的作用。

前瞻性研究:

- 按照今天的日期确定本市 50 家幼儿园中所有非哮喘病例的儿童，
- 将这些儿童分为两组：
 - 生活在以木材为燃料家庭（暴露）的儿童，
 - 生活在不以木材为燃料家庭（非暴露）的儿童，
- 对这些儿童随访十年，并每年进行相关检查，
- 十年后，计算每组儿童的哮喘发病率并进行比较（危险度比或危险度差）。

回顾性研究:

- 利用学校和公共卫生档案记录，确定本市 10 年前就读于 50 所幼儿园中任何一所的所有儿童，
- 根据记录中的家庭住址信息，将这些儿童分成两组：
 - 生活在以木材为燃料家庭的儿童（暴露）
 - 生活在不以木材为燃料家庭的儿童（非暴露），
- 在“今天”找出这些儿童（如到目前的学校去找），并确定在过去 10 年中哪些儿童患了哮喘，哪些儿童未患哮喘，
- 计算每个组别的哮喘发病率，并进行比较（危险度比或危险度差）。

病例—对照研究

定义

将病例组某可能影响因素的暴露频率（对某可能因子的暴露史和/或个人特征），与病例来自的总体该因素（暴露或个人特征）的“期望频率”进行比较的研究方法。所谓“期望频率”常常是通过对一组无病人群（对照组）进行调查获得的。

这种研究设计常常作为假设检验的初步研究方法使用，因为它具有相对于其它研究方法的明显优势（见下）。

目的

- 检验病例组某因素的暴露率是否高于对照组；
- 利用“比数比”估计超额疾病频率的相对危险度。

研究方法

病例—对照研究的关键在于如何能够从来自同一人群的病例和对照样本中获得无偏数据，以保证对两组暴露情况比较的公正性。

关于暴露史资料以及个体相关特征信息可以通过直接询问获得，也可以通过查阅历史记录获得。

为避免偏性，对照组的资料获取方式以及观察程序应该与病例组完全一致。

病例选择

- 虽然习惯的做法是从易于操作的来源中抽取病例样本，但理想的方法是选择规定人群中的所有病例；
- 无论采用何种病例抽取方法，必须保证人群中的所有病例均有相同机会被抽取为研究对象。

常见的病例来源包括

- 就诊于医院或家庭医生的病人；
- 登记在册的病人（如肿瘤登记）；
- 在社区某疾病系统筛检所发现的病人；
- 死亡证明。

对照选择

- 因为设置对照的目的是提供可能危险因素的“期望频率”估计值，因此对照应该是**病例来自总体的代表（理想随机）样本**；
- 一旦被选，就不能遗弃或替换对照中的个体，除非是由于对照的错误纳入；
- 可以通过对年龄、性别、社会阶层、婚姻状况等重要混杂变量的匹配来保证病例组和对照组其它相关特征的相似性。（注意病例—对照研究不能分析匹配变量及其相关变量作为危险因素的作用。当危险因素和潜在混杂因素未知时就有发生“匹配过头”的危险。因此如果不能肯定，最好不要对混杂进行匹配和调整）；
- 每个病例匹配二或三个对照可以提高统计学检验效力。

对照的常见来源包括：

- 与病例生活在同一地点或在一起工作的个体，与患儿在同一学校的儿童等；
- 人群登记资料，如出生登记、学校以及选举人名册、家庭医生名册；
- 其它疾病的医院病例（选取病例的医院）；
- 病人亲属—容易获取，但由于环境暴露相似且遗传特征接近，而可能对比较结果产生混杂作用；
- 随机数字电话拨号。

优点：

- 与队列研究相比获得结果快捷便宜；

- 所要求的样本量小；
- 可以较容易地找到病例组；
- 对于少见病而言这是唯一可行的研究方法。

缺点

- 事件间的时序关系有时不清；
- 因为分母总体未知，所以不能测量疾病的发病率或患病率；
- 很难保证对照能代表病例人群的总体（存在混杂）；
- 特别是在对照组，存在记录不完整或对过去事件以及暴露史回忆不可靠的现象。

例如

研究目的：

- 调查木料烟尘暴露在儿童哮喘发生中的作用；
- 确定本地区过去一年中在主要保健门诊接受治疗的所有哮喘儿童（6-10岁）（病例）；
- 确定与病例年龄相同、非哮喘患者、与病例在同一地区上学的对照组儿童，并从中抽取一个随机样本；
- 对上述儿童的父母进行调查，询问在孩子出生以及0-6岁期间家庭是否使用木材作为燃料。

分析

在病例一对照研究中没有总体人群大小的记录，不可能计算疾病的发病率或患病率，因此我们计算病例组和对照组的暴露频率，并应用比数比指标对此频率进行比较。关于这一点将在“危险度指标”章节中进行叙述。

当有一个以上的病因或危险因素或者因素之间存在交互作用时，就需要运用更复杂的统计学检验方法来评价这些因素的病因学意义（如 logistic 回归）。

“巢式”病例一对照研究

假定某研究人群本来是为回顾性队列研究设计的，但后来由于研究资源的限制，不可能获得整个队列重要影响因素（暴露或个人特征）的详细信息。这时就可以进行一次巢式病例一对照研究。

这是一个病例一对照研究，其中病例是队列中已经发生的全部病例，而对照则是同一队列中非病例的随机样本。然后只需收集上述病例和对照有关影响因素的详细信息就可以了。

该方法在研究起始就确定了暴露以及其它危险因素或影响因素的分类，因此它具有队列

研究方法的优点。

优点

- 无暴露回忆偏倚；
- 疾病前资料可信；
- 比队列研究的花费要少。

巢式病例—对照研究在一定程度上弥补了上述普通病例—对照研究的不足。

4.2.3 试验性（干预）研究

病例—对照研究和队列研究的结果能够对疾病的可疑病因以及疾病的预防和治疗方法进行提示。但对于干预措施的效能及其安全性则需要正规检测。此类研究常常采用**干预研究或临床试验**的形式。

定义

指选择某人群进行有设计的试验以评价某干预措施效果的研究。干预措施的效果是通过对试验组（接受干预或治疗措施）与接受传统疗法或安慰剂（无主动干预）的对照组的相关结局进行比较来评价的。

目的

- 评价医学干预措施的效果及其副作用；
- 验证因果假设。

绝大多数干预研究采用随机对照试验（RCT）的形式。试验研究通过比较试验组和对照组的结局来评价特定干预措施（治疗或预防）的效果。研究对象经随机分配形成试验组和对照组。

RCT 主要用于

- 与对照方案（传统处理方法或安慰剂）相比，评价新治疗或预防干预措施的功效率和安全性；
- 比较备选治疗方案的效果；
- 评价卫生服务不同提供形式的效果和效率；
- 为暴露于某可疑因素即可导致某疾病，或去除该因素即可预防或降低该疾病发生频率的结论提供直接证据。但此类试验研究也会带来伦理学方面的问题。

临床试验研究的基本要求是：

- 结局受益应具有临床意义并可以测量；
- 干预措施必须与试验对象的卫生保健需求相一致；
- 对于该干预措施的功效问题确实存在疑问；

- 干预措施对于病人和提供者双方必须都可以接受；
- 要有合理的根据认为受益高于风险。

临床试验设计

- 与队列研究的原则相似；
- 研究人群应该能够代表干预措施将要推广的人群（目标人群），该人群流动性要小且易于接近，不得使用志愿者；
- 必须按照随机化原则将研究对象分配至试验组和对照组（有时需要进行分层随机化分组—见下面内容）；
- 干预措施施加到试验组，对照组不接受干预措施，而是接受安慰剂或传统治疗方法；
- 为避免疾病报告（或其它相关事件/行为）或结局评价偏性，最好是研究对象与评价者都不知道研究对象的分组情况（双盲）。有时由于治疗措施的特征（如手术）明显，实行双盲是不现实的，这时可以采用单盲设计，即或者是病人或者是结局评价者不清楚治疗分组情况；
- 通过对试验组和对照组按照同一标准，采用一个或几个明确定义的主要和次要结局指标来评价干预措施的效果；
- 结局不仅应包括正面效果也应包括负面事件。

无论哪个组别发生结局的错误分类，都将减小两组间的效果差异，并导致受益降低的假象。

对于试验组和对照组，随访程序应相同，要点如下：

- 结局的定义和资料收集方法要简便，而且要足够敏感，能检测到所有相关事件；
- 必须按照标准方式收集与记录结局资料；
- 对于试验组和对照组的随访必须同样严格；
- 随访从随机化分组开始并持续到能够确定所有个体的结局为止；
- 必须报告所有的失访，并尽力减少之。

分析

分析过程必须包括所有随机化分组的病人—这就是所谓的意向分析，意思是：

- 所有病人均按照原始分配的组别进行分析，而不论后期处理过程中发生的偏离于研究方案的变化；
- 计数随访过程中的所有事件；
- 应分析研究方案中规定的所有正面和负面结局。

$$\text{功效} = \frac{(E - I)}{E} \times 100$$

这里 E=期望发病率（根据对照组推算）

I=干预组发病率

序贯分析是一种连续监测试验组和对照组结局对比的分析方法。当获得统计学显著性的阳性或阴性结果时即停止试验。它主要用于迫切需要得出结果、有可能发生严重副反应或预期收益较高的情况。由于避免了不必要的研究部分，因此所耗用的资源最少。

伦理学考虑

如同所有流行病学研究一样，在批准临床试验之前必须进行伦理学方面的评价。应考虑如下内容：

- 新疗法可能风险的未知特征或传统疗法治疗失败的代价；
- 某些个体暴露于新疗法受到损害的后果，剥夺了其它人可能的从中受益；
- 在没有前期试验的基础上将新治疗手段采纳为常规应用的可接受性；
- 向研究对象解释本试验研究的程度以及获得知情同意书的程序；
- 在保持“盲法”评价时，要关心研究对象的利益和安全。

交叉试验

每个组别按照随机顺序接受试验疗法和安慰剂疗法。本方法有时也运用于由于研究对象的选择或医学原因而导致不能使用双盲法的 RCT 中。在这种情况下，为避免偏倚必须按照所有个体的原始分组进行结局分析（即按照“意向分析”原则进行处理）。

参见关于病例—对照研究、队列研究、临床试验以及不同研究设计方案的比较相关章节练习内容。

5.研究对象：研究人群的选择

5.1 确定研究人群

必须根据研究目的从目标人群中抽取研究对象。样本人群指那些从目标人群中抽取的个体总和，样本人群的选择必须遵循代表性和可行性的原则。

研究问题 普遍真理		研究计划 研究真理
步骤1 目标人群	步骤2 可操作人群	步骤3 样本人群
特定的临床和人口学特征	特定的时间和地理特征	规定的抽样方法
<i>选择标准</i>		
适合于研究问题	代表目标人群 易于研究	代表可操作人群 易于操作

5.2 样本选择

利用整个人群进行研究几乎是不可能的。作为可行性方案，常常在一个由可操作个体组成的样本中完成研究，即所谓的*可操作人群*，可操作人群要代表其来自的总体，即所谓的“*目标人群*”。如果样本的抽样方法得当，那么就可以根据样本统计量估计目标人群中相关变量的发生频率。

对总体人群（*目标人群*）进行研究不仅成本高而且还有复杂的组织工作问题，因此需要从目标人群中抽取一个样本（*样本人群*），再对样本人群中的个体进行相关测量。但经常面临的问题是从总体人群中抽取一个样本也是不可能的，这样样本人群往往是从目标人群中的可操作部分（*可操作人群*）中抽取的。如果可操作人群与目标人群不存在差异（关键变量的分布相同），那么从可操作人群中抽取的样本就能反映目标人群的特征。在这种情况下就可以根据样本的测量值对目标人群作出推论。如果可操作人群或样本人群与目标人群存在差异，则不能根据样本的测量值对总体作出推论。

5.2.1 定义和抽样程序

在抽样过程中，全部抽样单位（普遍性）称为**总体**，样本就是从中抽取的。总体不一定是人群；其组成单位也可以是机构、记录或事件。所谓抽样程序就是要产生一个能代表总体的研究人群。在抽样时必须规定所使用的抽样单位，因为抽样单位决定抽样方法和资料分析方法。

目标人群是我们准备得出结论（作出推论）的个体（项目、测量值）集合。推论所反映的总体人群可能是从中抽取样本的单个人群（可操作人群，如锯木厂），也可以是所有类似的目标人群（如所有的锯木厂）。

样本（研究）人群是从可操作人群中抽取的用于研究目的的个体（单位）集合。样本人群的特征对于目标人群而言应该具有代表性，因为其结论将推广至目标人群。

如果可能的话，应该给出目标人群和样本人群特征间的比较结果，以确定可能对结果产生偏倚的特征差异。同样，应该对研究人群内的研究参与者和非参与者的主要特征进行详细全面的描述和比较。

研究人群的选择要点

根据下列原则选择研究人群：

- 合理性：利用该人群能否实现研究目标；
- 可行性：从诸如可操作性、人群中的成员参加研究的意愿以及研究费用等方面，考虑利用该人群进行本研究的可能性。

例如：

在提出木尘暴露对于肺功能影响的研究计划时，有人可能选择调查锯木厂工人。附近有多家锯木厂，最后选择了其中一家，因为你的内弟是该厂经理。它代表了一组具研究可操作性的人群。

应该根据地点、时间以及其它相关特征规定研究人群。

样本选择的一般要求

样本应具有如下特征：

- 能够代表目标人群；
- 足够大使数据随机变异的影响最小；
- 能够代表具有共同利益的所有群组。

例如:

所选锯木厂的类型亦即其加工的原料类型，必须能够反映研究问题。如果我们探索木尘暴露导致的哮喘，则必须选择这么一家锯木厂，其所加工的木料正是我们怀疑能导致哮喘的那种类型。从该地区几百家锯木厂中可以选择一家，其工人足够多，能反映回答研究问题所需要的木尘暴露水平的变异情况。

抽样过程中的关键决策包括:

1. 如何确定研究个体（即抽样框是什么）？如人口普查名单、登记名册、病人名单以及雇员名册；
2. 如何保证研究样本能代表目标人群，而不是将导致错误结论的有偏样本？
3. 什么是样本量？所谓样本量就是保证研究能够回答研究问题（即有足够的统计学效力拒绝或接受研究的无效假设）所需要的研究对象的数目。决定样本量需要考虑如下方面：
 - 测量值的准确性（内部变异）；
 - 所要求的精确度；
 - 具有实际意义的组间差异程度；
 - 可利用的资源。

从哪里抽取研究样本（即抽样框是什么）？

选择研究人群的常用方法包括:

- **人群基础的样本。**即从人群登记册、人口普查资料抽取或通过直接接触的方法如电话抽样法抽取；
- **机构基础的样本。**这类样本是从工作场所（职业性肺疾病研究）、职业协会（吸烟与肺癌关系的研究）、学校（进行结核菌素试验观察的班级抽样框抽样）或根据卫生服务使用者名单（哮喘、结核病或吸烟研究）选取；
- **其它途径。**多为方便样本或用其他随意方法得到的样本，有可能包含未知的选择偏倚。

5.2.2 如何进行抽样？

常用的方法有:

随机抽样，每个研究单位均有相同概率被抽取为样本（如采用随机数字表）。

系统抽样，根据名单按照一定间隔抽取个体组成样本（如根据工厂的雇员花名册每五人抽取一人组成样本）。

整群抽样，以“群”（如家庭）为单位进行随机抽样获得的样本。

分层抽样，将总体划分成亚组或层（如年龄组、性别或种族），在每一层内抽取不同样本量的随机样本组成总样本。这种抽样方法可以提高较小组别的代表性（如 1：2 抽样），但降低了较大组别的代表性（如 1：5 抽样）。

多阶段抽样，将上述两个或两个以上的抽样方法联合运用的方法。

5.2.3 如何保证得出的结论正确（研究的效度）？

按照上述方法进行抽样就可以避免在研究终点得到不正确的结论。样本选择可以在三个层次影响研究的准确度：

■ **精确度：**

研究人群的大小影响统计学推论的精确度。这一点将在后面的章节中详细讨论。

■ **内部效度：**

指与真实值相比，估计值的准确度或无歪曲程度（偏倚或系统误差）。它受抽样变异以外其它因素的影响，如选择人群的特征。抽样过程本身也可能导致系统误差。

■ **外部效度：**

这是指样本研究结果可向其它人群外推的程度。作这样的推断应该谨慎。内部效度好的研究结果未必就可以向其它总体外推。

样本选择中产生的偏倚（选择偏倚）所导致的效度问题，对结果的歪曲可能要远大于统计学推断不当所导致的结果歪曲。在后面部分将对这个问题进行详细讨论。

见关于人群选择的相关章节练习内容。

5.3 样本量的确定

在研究计划实施之前，研究者要确信研究样本足够大，根据研究结果得出的结论具有真实性，这很重要。另一方面，保证不会由于不必要的研究样本量导致卫生资源浪费（或包括了太多不恰当的个体）也同样重要。

例如:

某研究以不吸烟妇女为研究对象探讨被动吸烟与肺癌的关系,对一组不吸烟的护士随访了30年,并分别计算了在家庭或工作场所被动吸烟(暴露)者与非被动吸烟者的肺癌发病率。

在研究开始前,保证护士样本足够大,如果在30年后观察到肺癌发病率存在差异,该研究者就可以肯定这种差异确实存在,而不是随机变异的结果。另一方面,由于该研究费时费力,因此也不必对过多的护士进行调查研究。

5.3.1 为什么研究样本量很重要?

绝大多数研究要对组间疾病率进行比较或评价影响因素与疾病间的联系。为了实现这个目的,研究者需要研究或观察目标人群中的一个样本,并根据样本人群的结果得出关于“人群内真实情况”或“其它相似人群真实情况”的结论。

然而,由于在对研究对象进行选择或测量的过程中存在随机和非随机(系统)变异,样本结果可能并不反映目标人群中疾病与危险因素间的真实联系。

很明显,研究样本越大,结果的随机变异或随机误差导致错误结论的机会就越小。同样,如果组间疾病率的真实差异非常大的话,即使利用一个相对较小的研究样本也能检测出这种差异来。然而,如果疾病率的真实差异较小(但确实存在),那就需要较大的研究样本才能肯定结果的真实性。最后,如果两组间的疾病率的确存在差异,但由于对疾病的测量存在较大误差,也将需要较大的研究样本来保证研究中观察到的差异非误差所致。

参见关于随机变异与样本大小相关章节练习内容。

5.3.2 计算研究样本量所需要的信息

在确定研究样本大小时必须考虑以下内容。

无效假设

检验假设不同样本量的计算也不同,因此必须对无效假设进行明确阐述。通常根据主要的研究假设进行样本量计算。也可以根据利用最小亚组进行检验的无效假设计算样本量。

例如:

无效假设 (H_0):

被动吸烟与肺癌之间不存在关联 (或暴露于被动吸烟的非吸烟者, 其肺癌发病率与非暴露者相同)。

根据结果不会得出错误结论的确定性

“确定性”这个概念包括两层含义, 见下面的表格及示例。

根据对上述两组人群肺癌发病率的比较研究结果, 可能得到的结论列于下表:

根据样本结果得出的结论	人群真实情况	
	被动吸烟与肺癌有关	被动吸烟与肺癌无关
拒绝无效假设 (即两组的肺癌发病率不同)	符合	I 型错误 概率 = α
接受无效假设 (即两组的肺癌发病率相似)	II 型错误 概率 = β	符合

对上表两种类型错误结论的说明:

1. 如果研究结论为肺癌发病率在两组间不同, 而实际上被动吸烟与肺癌之间并无生物学联系的话, 你将错误地“拒绝了无效假设”(I 型错误)。

α 犯 I 型错误的概率 (即在实际上不存在关联时, 得到组间有差异或存在关联的结论);

可信水平

没犯 I 型错误的确定性 ($1 - \alpha$)。

2. 如果研究发现两组间的肺癌发病率相似, 但实际上被动吸烟可以增加肺癌的发生风险, 那么你将错误地“接受无效假设”(II 型错误)。

β 犯 II 型错误的概率 (即实际上存在关联时, 得到组间无差异或不存在关联的结论);

效力 (把握度)

没犯 II 型错误的确定性 ($1 - \beta$)。

预期研究结果 (效果规模)

为了估计在确定的可信水平下需要多少研究个体才能有效地检验假设, 则需要有关研究结果期望大小及其变异程度的信息! 这就是所谓效果规模。

效果规模取决于两个因素:

- 两组间的期望效果差异,
- 各组测量值的期望变异程度。

换言之, 必须根据“噪音强度”估计“信号”大小。

5.3.3 如何利用这些因素计算样本量?

很明显,样本量、研究效力、可信水平以及效果规模不可分割地联系在一起。拟了解(或估计)其中的任何一个参数,就必须先掌握(或估计)其它三个因素的信息。例如,为了估计研究的样本量,研究者就必须事先确定可信水平、可接受的研究效力以及期望的效果规模。

样本量可能是你希望计算的信息。但也可能受可供研究的病人数量限制,样本量是固定的;或者由于财力、时间或其它资源限制,样本量是相对固定的。在此情形下,需要确定的因素就变成了研究效力、可信水平或者效果规模。

公式中各因素的估计值

如何选择可信水平 ($1 - \alpha$)

- 由研究组来决定 (或者这是你要计算的因素);
- 研究者常常采用默认值 $\alpha = 0.05$ (给出 95% 的可信区间);
- 根据犯 I 型错误的风险来选择。

例如:

如果某新疗法有严重的副反应,那么研究者就希望在得出新疗法的效果确实不同于(优于)原疗法的结论时,要非常有把握。这种情形下,研究者就会选择 $\alpha = 0.01$ 的可信水平,并给出 99% 的可信区间。

如何选择研究效力 (或 $1 - \beta$)

- 由研究组来决定 (或者这是你希望计算的因素);
- 研究者常常采用默认值 $1 - \beta = 80\%$ (即有 80% 的把握没犯 II 型错误)。之所以选择这个值的理由一般是因为研究者认为当不同疗法间实际存在差异,而结论是不同疗法效果相同时所犯的误差要比犯相反的错误更安全;
- 根据犯 II 型错误的风险和代价来作出选择。

例如:

如果新疗法成本低廉且快捷,而研究者只有一次机会探讨该药的效果,那么他们就希望,如果新药确实有效,就别犯认为新药无效的错误。这时他们就希望研究具有较高的效力水平 (如 95%)。

效果规模有关信息的来源有:

- 其它类似研究;

■ 预试验;

■ “最佳猜测” 或对 “重要差异” 的估计。

对于研究者而言上面的样本量计算常常令人难以接受。下面的情形很常见，研究人员向统计学家咨询关于样本量计算的问题，结果很惊讶地发现统计学家坚持要知道两组间疾病率的期望差异或研究变量的期望变异。

研究者回答“如果我知道这种差异的话就不要研究啦!”。很遗憾，样本量估计的唯一方法是通过“效果规模”进行经验基础上的猜测，或通过选择一个具有卫生服务工作的“效果规模”值来实现的。

样本量（或可信水平、研究效力、效果规模）的计算公式

样本量/效力计算最常用的公式是用于两组间疾病率比较（如耐多药结核病的患病率）或两组间测量变量（如 FEV₁）平均值比较的公式。如果研究方案涉及其它类型统计学分析方法的话，那么在开始研究之前最好就样本量/研究效力的计算咨询统计学家。

1. 两组间疾病率的比较（如队列研究、横断面研究或试验性研究）:

EPI-INFO 有一个统计计算器，如果向它提供了其它因素的值，它就可以计算样本量（或研究效力、效果规模）。

公式是:

$$\text{样本量} = \frac{[(P_0 \times (1 - P_0)) + [(P_1 \times (1 - P_1)) \times (Z_{(1-\alpha)} + Z_{(1-\beta)})^2]]}{(P_1 - P_0)^2}$$

这里:

P₀=非暴露组（或组 1）的期望疾病率

P₁=暴露组（或组 2）的期望疾病率

Z_{1- α} =所选择 α 水平的 Z 分（见表）

Z_{1- β} =所选择 β 水平的 Z 分（见表）

Z 分:

绝大多数统计学软件只要求输入 α 值和 β 值，然后软件根据研究类型将其转换成数值（即所谓 Z 分）。如果没有计算机而进行样本量计算的话，就得从 Z 分表中读取适当的 Z 分值，绝大多数统计学教科书都附有 Z 分表。

2. 两组间暴露频率（或危险因素）的比较（如病例一对照研究）:

公式同上，但这里的 P₀ 和 P₁ 指对照组和病例组中暴露或危险因素的发生频率或比例。

3. 两组间测量值均数的比较（如横断面或试验性研究）:

$$\text{样本量} = \frac{2(SD)^2 \times (Z_{(1-\alpha/2)} + Z_{\beta})^2}{(\text{均数 1} - \text{均数 2})}$$

这里：SD=总体中测量值的期望标准误

4. 相关或回归分析中的样本量计算：

也可以利用相关或回归分析的假设计算样本量。但应该向统计学家咨询。也可以运用“掐指一算”（凭经验估计的方法）的办法给出样本量估计值来，这对于研究设计的目的而言非常实用。所谓“掐指一算”是指对于回归模型每个待检验的因素而言大约需要 30 个研究对象。例如，如果你计划研究肺功能（如 FEV₁），并需要考虑年龄、性别、哮喘史、社会经济状况、吸烟史以及环境暴露（6 个因素）的影响，那么关于样本量适宜的假设是需要约 6×30 即 180 个研究对象。

在进行样本量计算时关于单侧还是双侧检验的简评

在上述计算中可以从标准 Z 分值表中查找相应的 Z 分值。对于 α 的 Z 分，假定 I 型错误的 1/2 分布在正态分布的一端，另外 1/2 分布在另一端。这就是所谓双侧检验。双侧检验通常是正确的，除非有特殊理由认为从阴性结果中也能获得有意义的新知识时则采用单侧检验，否则就应该采用双侧检验。

关于单侧和双侧检验的进一步介绍

在绝大多数情况下，研究者在执行一项研究计划时都有一个关于研究结果的假设。例如，在一项关于哮喘危险因素的研究中，研究者可能假设市区的空气污染将导致哮喘发病率的升高。然而，如果研究结果提示生活在高度污染市区的个体其哮喘发病率反而显著为低，那也是一项有意义的发现。换言之，无论初始假设如何，在绝大多数的研究中，不管沿着哪个方向探索，其研究结果均能提供有意义的信息。

偶尔，研究者只关心某个方向的结果。

例如，计划评价一项新的结核病治疗措施，看它是否可以缩短疗程减少访视次数。研究有三种可能的结果：(1)标准治疗方案效果好；(2)两种治疗方案之间无差异；(3)新方案效果好。在这种情况下，结果(2)和结果(3)的意义相同，因为无论结果是(2)还是(3)为真，都将推动我们采用新治疗方案。也就是说，我们根本不需要对结果(2)和结果(3)进行区分。

在该示例中，计算研究效力或样本量时，适合于运用单侧检验，因为对于标准治疗方案而言你并不关心新方案的效果是否要好一而只关心其效果是否要差。这种情形下你得选择 α/2 的 Z 分值。此时切记对于相反效果（即新治疗方案的效果要好）的分析效力为 0。

见关于研究方案制定中的样本量和研究效力计算相关章节练习内容。

6. 流行病学中的变量测量

6.1 总则

6.1.1 信息收集的原则

数据的收集方法和变量条目的录入标准，必须在研究起始就明确规定，并在整个研究过程中严格执行。反对在研究中对上述方面进行修改。如果作了修改，必须清楚地记录修改的理由。这些标准应涉及所有对研究有重要意义的因素，如时间、地点、年龄、性别以及种族等。

变量测量的一般结果是数字或类别。可以根据这些数字或类别结果的排列方式得出逻辑结论。变量间的相互关系告诉我们周围世界的秘密。

6.1.2 数据类型

研究方案应规定需要收集的数据或信息，以便选择恰当的分析方法。数据有两种主要类型：

连续型资料

这是一些落在连续性衡量刻度内的数据点（如身高或年龄以及某些功能测量值如FEV1）。反映连续性资料的数值无离散化意义，它们往往在同一衡量刻度内相互交融。

例如：

尽管我们说某人 52 岁了，但这并不绝对准确。它仅仅是按照最近年份的近似年龄。加上小数点（例如 52.69 岁）后就可以较精确地描述实际年龄了。增加的小数点位数越多，估计值就越准确。因为不可能获得“绝对”精确的数据，所以此类测量值常常记录为“位于某区间”（也就是说某人 52 岁，指其年龄在 52.00000... 到 52.99999... 之间）。

离散型（分类）资料

离散型资料代表唯一取值的组别。经常利用对某类别的回答“是”或“否”来表示某个体或研究对象是否属于该类别的信息。因此对于性别而言，个体或男或女，再没有其它类别。这些类别相互排斥而无交叉覆盖。如不是女性，则必定是男性。但也存在两个以上独立离散值的可能性，例如家庭中的儿童数。

死亡可以认为是离散数据。一个人或者是生或者是死，确定某个体的生死状态非常简单。但当涉及到死因时就不那么容易精确测量了。

绝大多数疾病很难采用离散型变量进行分类。例如为了分类的目的，我们假定某个体患有或者未患诸如哮喘、结核病或者肺炎之类的疾病（即离散型数据）。然而患病与未患病的界限却非常模糊，一个人称之为肺炎的疾病，另一个人却可能不这么叫。因此通常根据标准定义中的人为标准进行个体的分类。

因此，研究方案必须对研究的状态（疾病）或特征（因子/因素/影响因素）规定尽可能精确的定义。如果作不到这一点，将导致测量值的不精确，而且会让调查人员无所适从，并损坏其根据收集或报告信息作出结论的能力。

6.1.3 健康状态或特征的测量工具（数据收集）

需要利用工具或仪器进行数据收集。在进行肺部健康研究时，可能用到的工具类型有：

- 访谈、调查表、日记、表格以及常规记录；
- 测量工具，如测量结核菌素或空气变应原皮肤反应的尺子，肺活量计，对环境空气污染进行采样的工具等；
- X-光片；
- 微生物学技术；
- 组织标本的检查技术。

某些测量仪器的结果要比其它仪器精确。这常常可使各类别间无重叠以及在判断个体类别时结果相对明确。

例如：

如果要收集关于性别方面的信息，则可以通过观察个体、让个体回答或者通过某种记录资料来得到其性别信息。在绝大多数情形下，区分男性和女性还是很容易的。

相反，如果类别间并非真正离散化，并且存在明显重叠（如肺炎患病个体与未患病个体），或者测量工具比较主观时（如阅读胸部 X-光片），测量值就相当不精确了。当你对某测

量值可能落入的区间进行判断（如测量结核菌素皮肤试验的硬结反应）时，这种不确定性是测量值的必备特征之一。

在肺部健康研究中常规应用许多标准化和公认的方法。如果可能，要严格遵守这些定义和方法，以保证研究结果与前人的研究具有可比性。如果与上述标准发生了偏离，请详细描述。

6.1.4 有关测量值的问题

研究方案必须对收集的数据、数据的收集方法以及数据如何分类等作出精确规定。这是为了保证所记录的数据无模糊之处，也是为了保证其它研究者或评阅人能够证实你的结果或将其与其它公开发表的研究结果进行比较。

保证可比性：定义及操作程序的标准化

标准化是指将不同研究间方法学差异的影响最小化的过程。标准化可以运用于诊断定义、诊断仪器以及数据收集程序。

为研究目的所选择的定义通常与比较目的所使用的定义一致。在主题领域，科研工作是在广泛的范围内协作完成的，这时上述定义通常是在已发表科研工作的基础上通过国际合作调查作出的。在有此类国际推荐定义存在的情况下，研究者或者遵守它或者详细说明偏离这些定义的程度及其原因。

对于测量技术和仪器也是一样。如果存在关于测量程序或仪器类型的国际标准化建议，则应该首选，应对它进行准确描述并严格遵守。对这些建议的任何偏离都必须详细说明其理由。

保证精确度：测量误差的最小化

测量误差反映了部分研究人员的粗心、对测量工具或测量仪器使用不熟悉或不正确或者没有遵守旨在减少此类误差的标准化程序。如后所述，此类误差可能是随机误差也可能是系统误差。

尽可能将收集到的实际数据值全部报告出来，因为这才是数据质量的反映。令人吃惊的是多数研究都没有给出“真实值”。

例如：

测量值误差的经典示例是“尾数偏好”。这在结核菌素皮肤试验中得以体现。在结核菌素皮肤试验中不重视精细的测量技术，研究人员（或卫生服务人员）将结果判读成五级中的某一级。如果努力改进测量精确性的话，则可避免此类误差。尾数偏好是数据收集质量的反映。

测量值的精确度常常是研究人员的个人特征反映。为了保证测量值的最佳精确度，应该将每个研究人员的测量结果与其它人以及“金标准”进行比较（即观察员间误差）。如果某研究人员的测量结果与其它人员相差甚远，那么这个调查员应继续接受培训直至符合要求或者将其排除出局。另外，每个调查员还应该对同一值进行重复测量，以保证在不同条件下对同一值的测量结果无明显差异（调查员内部误差）。

信息记录的完整性和准确性

经验不足的研究人员也许会感到惊讶，测量值的记录和编撰过程难道也会发生诸多误差而且足以损坏一项研究？无论研究人员多么细心地测量，也不论测量仪器有多么精确，如果测量信息记录不细心不完全，编撰不准确，那么所收集的信息就不再具有科研价值。没有任何一种统计学操作可以克服由于信息处理不当所造成的问题。

即使经过精确细心测量，数据的记录和编撰也很认真，但在信息记录和编撰阶段引入的误差可能仍然不小。许多优质研究，此类误差可能占全部信息的百分之几。因此必须努力将此类误差降低至最低水平，并对此类误差的水平进行估计。关于这一点将在数据管理的章节中进一步讨论。

6.2 疾病的记数和健康测量

健康可以看作是个连续谱，因此可以用功能（即个体能够完成规定任务或活动的程度）指标来测量。疾病对健康的影响在于其抑制机体功能并危及生命，因此健康状态的分布是从功能完美到极端失能甚或死亡。用于健康分类的测量值范围包括死亡、发病和患病、功能损害、亚健康状态以及危险人群。

6.2.1 信息来源

健康和疾病的信息来源有多种。用于流行病学目的的信息主要是从医疗服务过程、疾病监测以及对人群的健康监测过程中获得的。另外一种成本较高的信息收集方法是开展专题研究。

为了科研目的，信息必须贴切研究主题，而且要可信、完全、易获得。

常规收集的信息

在病人医疗服务过程中常规记录的信息可以运用于科研目的。虽然此类信息通常是可获得的并且符合研究主题的要求，但它不是不完全就是不可信。即使对于流行病学家而言，查阅其在日常医疗服务中积累的病例记录，从中提取有用信息用于科研目的也是一件需要审慎处理的事情。

例如：

通过查阅常规记录来获得病人吸烟习惯的细节就很能说明问题。尽管我们知道吸烟是肺疾病最重要的相关因素，但常规记录中这种信息的频繁欠缺实在令人吃惊。

由于以上问题，最有意义的常规收集信息是那些按照标准表格收集的信息。这些表格通常是国际推荐表格（例如，IUATLD 在其出版发行的关于结核病、儿童急性呼吸道感染、哮喘以及烟草的控制及其预防等一系列“指导手册”中推荐的信息收集表格）。当规范地使用此类表格时，所记录的信息就可能比较完整，不同病人之间的信息具有可比性。

此类表格也可用于官方统计，诸如生命事件登记（死亡、出生以及结婚等）以及法定报告疾病的登记（如结核病）。

专题研究收集的信息

符合特定主题的常规信息不是不存在就是质量没保证。这种情况下就必须专门为研究目的收集信息。显而易见其研究成本要高，但这经常是获得可信而又完全信息的唯一途径。

正如常规收集的信息那样，运用标准化的、经过现场检验的表格收集到的信息对于研究目的而言是最可信的。

6.2.2 死因研究

在健康谱中最容易界定的健康状态是死亡；但确定其死因却是另外一码事。

死亡记录

死亡信息常常是通过地方生命统计登记处的常规死亡通知获得的。这种信息是运用标准化表格收集的，并按照国际推荐格式进行了死因分类。

一般说来该表格是由病人的主治医师填写的，但有时也由并不了解该病人的医生或验尸官填写。

死亡测量

所有反映死亡的率都是“发病率”，因为死亡是离散型且时间确切的事件。死亡率用以常规测量具有高病例死亡比的肺部疾病（如肺癌）。在干预措施有效时（如结核病或肺炎），用死亡率来反映卫生服务质量和可行性比用社区疾病负担要优越。

死亡证明书

<p>死亡原因</p> <p>I</p> <p>直接导致死亡的疾病或情况* (a) _____ 由于（或作为其结局）</p> <p>(b) _____</p> <p>前期原因</p> <p>指加重上述病因的发病情况，在 (c) _____ 最后要描述该情况 由于（或作为其结局）</p> <p>(d) _____</p>	<p>发病到死亡的大约间隔</p> <p>_____</p> <p>_____</p> <p>_____</p>
<p>II</p> <p>其它导致死亡，但与引起死亡的 _____ 疾病或情况无关的重要情况 _____</p> <p><small>*这里不指死亡方式，如心衰、呼吸衰竭。它是指疾病、损伤、或引起死亡的合并症。</small></p>	<p>_____</p> <p>_____</p>

在进行死亡测量时，会用到不同的率：

- **粗死亡率**指某人群在规定时间内死亡的概率；
- **原因别（或年龄别）死亡率**指某人群在规定时间内由于某种疾病或某个条件（或某年龄组）的死亡概率；
- **病死率**指某病患者在规定时间内死于该病的概率。

在规定人群内计算各种率的计算公式如下：

率	表达式	常用单位
粗死亡率	$\frac{\text{期间的死亡人数}}{\text{同一期间的死亡危险人数}}$	100,000 人年
原因别/年龄别死亡率	$\frac{\text{一定期间内由于某种原因/某年龄组的死亡人数}}{\text{同一期间的死亡危险人数}}$	100,000 人年
病死率	$\frac{\text{某期间内死于该病的人数}}{\text{同一期间的患病人数}}$	百分比

某病患者也可能死于其它原因而非所患疾病。

例如：

最典型的例证就是一个 91 岁的老人，其来年能够继续存活的概率要比一个 21 岁的年轻人小得多，而无论这个年轻人患有什么病。因此在对不同人群进行比较时必须考虑这些增加患者死亡概率的因素（年龄、性别以及社会经济水平）。

6.2.3 发病率研究

发病是生理或心理完好状态的主观或客观偏离。该术语通常指某种疾病或精确定义的功能状态。可以以三种方式测量发病率：

- 病例；
- 病程；
- 疾病影响（病程、功能损害程度以及卫生服务利用情况）。

疾病计数以及健康测量时需要考虑的疾病特点

频率

疾病的发生频率是选择研究设计的一个重要因素。对于少见病而言（肺癌、结核病）运用队列研究设计其研究效率就不会高。而运用病例—对照研究设计则可能更具可行性。

严重程度

在进行疾病或功能影响因素的组间比较时，必须考虑疾病或失能的严重程度。

例如：

- 结核病人痰涂片阳性程度的半定量评价与结核病人的传染性有关；
- 结核病人肺部病变范围与病人的存活概率以及对治疗的反应有关；
- 哮喘病人的肺功能水平及症状发作频率（程度分级）与卫生服务的效用有关；
- 慢性阻塞性肺病病例的生理功能水平与病人的致残以及生存有关。

病程

肺部疾病的病程变异很大。急性呼吸道感染的病程很短，而哮喘则有时会终生相随。疾病的病程对于研究设计的选择具有重要影响。病程较短的疾病不能有效利用横断面研究设计。

潜伏期

由环境暴露所导致的许多慢性疾病（慢性阻塞性肺病、肺癌）以及存在携带状态的传染病（结核病、社区获得性细菌性肺炎、真菌性疾病以及巨细胞病毒疾病）都具有潜伏期。如果潜伏期是研究疾病的特征的话，在选择研究设计时须予以考虑。

测量仪器或测量方法的选择

记录病例、事件或影响等常规收集信息的工具

在各种肺疾病中，绝大多数国家只对结核病进行病例和发病事件的常规记录和报告。在此情况下，有一个按照标准化的定义和程序运行的国际化的信息记录和报告体系。

可以利用现成的培训资料进行培训，以保证记录和报告的质量标准化。另外，也规定了一些用于对报告信息的质量进行一般性评价的指标。这些推荐的操作定义和操作流程构成评价框架，结核病研究就是按照这个框架的要求执行的。

对于另外一些信息也有常规报告，这些信息是以住院形式反映哮喘和急性呼吸道感染的影响。这种信息的标准化程度差且难以用于研究目的。

如果常规记录的病例、事件或疾病影响的信息要用于研究目的的话，则对信息记录体系有一定的技术要求。这些要求包括：

- 记录必须简单化，只包含能够保证准确性的最少信息；
- 应该是常规工作的一部分，有助于鼓励实践者完整报告信息；
- 必须有达成一致的操作定义，要对表格的使用者进行系统培训，以保证答案的一致性；
- 必须有运行良好的信息交流体系，以保证信息的规律传输以及结果或相关后果及时反馈给表格填写者；
- 为了信息记录和报告的持久性，需要一个具有完整性和准确性的常规信息监测体系（包括达成一致的监测指标和常规监督）。

记录专题研究收集信息的工具

关于研究目的的信息通常需要专题收集，而且对于该项工作已经有了一套标准化的定义和操作流程。

在选择测量工具时需要考虑的方法学或测量工具的特点包括：

- 存在标准化的、推荐的研究方法；
- 所用测量工具或方法的准确性（测试特征）；
- 所用测量工具的耐用性；
- 研究对象对该测量工具的可接受性；
- 使用该工具的可能损害；
- 维护个人隐私。

6.3 调查表的使用和设计

在流行病学调查中调查表常常作为信息记录（数据收集）的工具使用。调查表的良好设计对于能否收集到准确可靠的信息非常关键。

目的在于保证按照系统方式收集的信息应尽可能符合事先设计好的、标准化的标准和方法的要求。这样就可以对不同组别研究对象的特征进行有效比较。

调查表可以通过面对面的访谈完成填写，也可以由研究对象自己填写。调查表的填写形

式会影响调查表的格式，但基本原则不变。

调查表中的问题要精确无含糊之处，并且在必要时，向调查员或受调查者提供填表说明，就问题如何解释、如何记录答案以及如何完成调查表等进行详细交代。

希望上述注意事项能对调查表的良好设计起到指导作用。

6.3.1 调查表的设计原则

研究目的：研究目的要清晰准确，这非常重要。

问题的选择：只包括实现研究目的所需要的问题。必须严格排除与研究目的无直接关联的问题。

问题的语言表达：问题的语言表达对所有研究对象都必须清晰易懂无模糊之处。

问题的排列次序：应该仔细安排问题的询问次序。问题的编排要有逻辑性、让人感兴趣，而且要省时间。

答案选项：要详尽说明各问题的答案类型（即离散选项或开放式答案）。

调查表的结构：调查表的结构和布局设计一定要便于调查表的填写和数据分析。

6.3.2 调查表的内容

简述

这是一段简短叙述，内容如下：

- 研究的资助方以及研究目的，
- 研究对象（病例和对照）的选择，
- 调查程序（访谈或自填表），
- 研究对象拒绝参加研究或拒绝回答某问题的权利，
- 保护个人隐私的措施，
- 如果需要，需获得研究对象签名的同意参与研究的书面材料。

研究对象的识别：研究对象的姓名、住址及其识别号码

本部分收集的信息应该记录在调查表的封面上，然后将其拆下来保存到一个单独的文件中，与其余的数据部分分开。只用研究对象的识别号来标识该个体的其他所有数据。个体标志符与研究对象识别码之间的联系必须形成机密卷档，只有研究者才能将二者联系起来，这可以保证个体识别数据的保密性。在其他未经授权者也可进入的文件中，将个体信息与个体标志符或对象识别码联系起来的行为是不道德的。

个体特征

调查表测量记录的内容主要取决于研究的特征和目的，但通常包括：

- 研究题目和对象识别码；
- 调查日期、时间、地点以及调查员姓名；
- 性别、年龄、婚姻状况、职业以及与研究目的相关的社会资料（如教育、收入、居住条件、家庭成员数）。

注意：通过比较研究人群与目标人群某些特征的频数分布，个体特征资料经常用以检查在对象选择以及研究参与中发生的偏倚。

问题的选择

所有问题必须严格贴切研究目的。不要包括无关的多余问题，例如那些所谓“如果有机会了解的话则更好”的问题。同样，出于节约成本考虑以及为了维持研究对象的兴趣和合作，调查表应包括最少量的问题，但这些问题能够保证获得回答研究问题所需要的所有信息。

然而，有时如果能将几个主题的内容揉入一次调查则可能更经济。这种情况下每项子研究的目的是要同样清楚，研究对象的参与水平也不能因研究对象过分劳顿而受损害，关于每个主题的问题选编应同样严格。

问题的语言表达

问题的语言表达对于获得服务于研究目的、质量良好的信息非常重要。语言表达应具备如下特点：

- **简单**—问题设计应该简短，一个问题只问一条信息；
- **易懂**—尽量使用常见的、每个人都能理解的词汇（在某些情形下对技术术语进行简短解释或代之以另一个用语也许有好处，为调查员准备关于进一步探查问题以及对答案进行解释的指南也是必要的）；
- **明确**—问题不能有多种含义（如“你在用药吗？”—这里就需要对此进行定义）而且必须精确（如“你经常咳嗽吗？”—需要界定经常咳嗽的频数范围）；
- **无偏**—避免对问题的特定答案进行暗示（如“你认为吸烟会损害你的身体吗？”）。避免使用感情语（如“你认为吸毒不道德吗？”）。不要鼓励对流行事物或社会可接受事物的想象（如“你认为所有结核病人应该隔离吗？”）；
- **兴趣**—问题对研究对象要尽可能贴切并有趣，以维持其注意力及参与意愿。

对于第一语言非英语的研究对象，英语调查表应该由一个语言学家翻译成其母语，再由另一个语言学家翻译回英语。将翻译结果与原始材料进行比较，以保证问题的含义不会在翻译过程中遭受歪曲。

问题的编排次序

问题询问的顺序可能会对给出的答案有重要影响，因此需要仔细考虑问题的编排次序，以避免出现冗长乏味、尴尬以及得到有偏答案。为此下面几点建议很重要：

- 要使调查对象对于询问的问题感到放松。因此调查表通常从容易的问题开始，诸如个体识别资料以及人口学资料，以建立研究对象的信心。关于个人习惯以及生活方式的问题可能具有胁迫性或使人尴尬，应该安排在调查表的后面部分。
- 问题的顺序安排应该流畅并符合逻辑，使研究对象以及调查员感兴趣。复杂问题最好放在调查表的后面部分。
- 对前面问题的回答不应后面的问题产生影响。例如关于吸烟危害的问题就可能影响研究对象对自己吸烟习惯问题的答案选项。
- “核对问题”，即那些用来检查答案一致性的问题（如“年龄”和“出生日期”）应该分开安排。

答案选项

答案格式必须为研究对象提供一个作出准确完整回答的范围。因此应该给出一个适当的答案选项单。答案格式一般有三种类型：

- “是”、“否”（或“不知道”）型。两分类答案的最大缺点是不能区别答案背后的含义。
- 多选型格式列出了包括“不知道”在内的可能答案，研究对象可以从中挑选一个或几个答案。由于很难保证答案选项单能包括所有的可能答案，因此还要包括“其他（请详细说明）”的选项。尽管这可能有助于探查研究对象的“逃脱”答案，但此类答案又具备了“开放性”格式的缺点，因此应该尽可能避免。
- “开放性”格式常用于探索未知领域的调查表，诸如个人想法及态度等，即所谓人类学研究。此类答案具有不受调查员误差影响的优点，允许研究对象用自己的语言充分表达自己的思想。此类研究常常作为进一步研究的前期工作。其缺点是资料难以分析，也难以得到一般性结论。

调查表的结构

质量良好的调查表应该具备易于完成并易于分析的特点。可以通过以下措施实现调查表

易于填写的目的：

- 为调查员（自填表的情况下为应答者）准备清楚明白的填表说明。这包括开始部分的一般性说明以及必要之处对个别问题的说明；
- 对问题进行编号；
- 在可能跳跃问题回答之处进行说明，例如某问题答案为阴性后跳过随后的问题。

可以通过下列措施实现调查表易于分析的目的：

- 离散型的问题和答案要比定性的问题和答案易于分析和解释；
- 使用事先编码的答案框；
- 将答案框以调查表边沿对齐，以利于编码和计算机录入；
- 保证对每个问题都列出了所有的答案选项，包括“不知道”，“不适用”和“未回答”。

最后：

- 使用良好质量的纸张进行印刷；
- 在同一研究中，对于不同目的的调查表宜使用不同颜色的纸张以示区别；
- 简短的调查表易为研究对象所接受，从而可能得到较高的应答率。

6.4 暴露的测量

本章节将对职业性危险因素和环境危险因素的暴露作为影响因素或危险因子的测量进行示例。这里概述的绝大多数原则也适用于其他影响因素的测量。

6.4.1 环境因素暴露的记录和测量

就象疾病信息可以通过常规途径或通过专题研究途径收集一样，关于肺疾病环境和职业暴露以及其它可疑影响因素（危险因素）（诸如社会经济状况、营养状况、合并症）的信息也可以有多种来源。

适用于前述疾病信息收集的条件也同样适用于本类信息的收集：利用标准化表格记录信息，遵守标准化操作程序以及就所运用的数据获取方法而言，信息记录尽可能完全。

需要考虑的暴露因素（剂量）特征

强度	这是指暴露的水平或大小（如每天吸烟的支数、工人呼吸区内的硅尘浓度）。
暴露时间	指暴露发生的时间长度（如从事某工作的年数；在某地区的居住年数；与某传染源的接触时间以及吸烟年数）。根据研究疾病的健康结局特点，需要考虑的时间长度可以长至年（如癌症、慢阻肺），也可短至小时或天（如哮喘病人的喘息症状发作）。
暴露特征	有时非常需要考虑暴露期间的暴露水平是恒定的，还是由一系列的暴露峰组成，而高峰之间则是非暴露期（如工业地区污染物的重复事故性泄漏）。
有效暴露时间	指生物学相关时间内的暴露（强度、时间和/或特征）效应。例如在肺癌的环境或职业性危险因素研究中，获取研究对象目前的职业信息或目前的被动吸烟信息没什么意义。而需要了解的是该个体 10-30 年前的家庭以及职业性暴露的信息。
环境与因子	许多情形下，相关病因并不清楚（如城市环境、在铝矿冶炼厂工作）；换言之，病因是相关因素的复合物。切记暴露也可以指一种“环境”或混合物，而不仅限于某一因子。

6.4.2 暴露信息的收集工具

- 访谈、调查表、日记和记录；
- 大环境的测量（如城市、工作场所空气中的石棉水平）一称之为“区域测量”；
- 个体环境的测量（如由研究对象随身携带的空气采样器所获得的测量值）一所谓“个体测量”；
- 个体暴露剂量的测量（如将呼出气中的 CO 浓度以及血清中碳氧血红蛋白浓度作为个体吸烟量的测量值）；
- 组织中相关标记物的浓度测量（如肺组织样本中石棉纤维的定量测定）；
- 暴露直接效应标记物的测量（如卷烟烟雾中致癌物诱导产生的 DNA 加合物）。

理论上，这里列出的测量工具越深入，对暴露的测量就越直接越有效（与前面的“非直接”测量值相比）。但实际上，这些深入测量工具的局限性往往超过其优点。

不同测量工具的优缺点

测量工具	优点	缺点
调查表、访谈、日记和记录	<ul style="list-style-type: none"> ●简单 ●成本相对较低 ●可以获得大量研究对象的信息 ●可以对测量工具进行设计以获取暴露环境或混合暴露的信息（如工作、居住地区） ●可以获得从前的暴露信息以及暴露时间的数据 	<ul style="list-style-type: none"> ●依赖于个体的信息报告（所得结果可能存在偏倚或者不够准确） ●多是定性资料而非定量信息
区域测量	<ul style="list-style-type: none"> ●相对于个体测量，成本要低 ●有可能利用常规来源资料（如空气污染监测站资料） ●定量资料 	<ul style="list-style-type: none"> ●可能并非特定个体暴露的良好反映 ●暴露强度为某时点的测量值，而该时点可能非相关时点 ●不能获得暴露时间长度信息
个体测量	<ul style="list-style-type: none"> ●可以测量个体的暴露强度 ●可以评价个体的暴露特征 ●定量信息 	<ul style="list-style-type: none"> ●费时费力 ●暴露强度为某时点的测量值，而该时点可能非相关时点 ●测量方法不够敏感（检测限） ●不能获得暴露时间长度信息
个体暴露剂量/组织浓度测量	<ul style="list-style-type: none"> ●反映吸收剂量 ●是个体暴露剂量的直接定量测定 	<ul style="list-style-type: none"> ●成本以及可行性变化大 ●可能对受试个体有损害 ●需要考虑机体对暴露的清除/冲刷作用
暴露效应的生物标记测量	<ul style="list-style-type: none"> ●是生物学有效暴露的最接近测量 	<ul style="list-style-type: none"> ●对个体具有损害性 ●成本高 ●对绝大多数疾病而言尚无现成技术（如哮喘，慢阻肺（COPD））

6.4.3 暴露评价方法的选择

暴露评价方法的选择取决于现有资源、专业知识以及研究相关暴露的特征。

例如：

因为肺癌潜伏期长的特点，测定矿工目前的个体环境暴露水平不适合于肺癌环境危险因素的研究。

卫生专家常感到他们对于暴露测量或暴露记录的专门知识太局限，因此必须依赖与该方面专家的协作或干脆将该部分从研究中去掉。虽然比较理想的方法是就研究中涉及环境暴露或营养学问题的成分在咨询相关专家后再进行设计，但实际上常常找不到此类专家。

简单调查表的作用（如美国胸部学会关于吸烟习惯或简单就业史的标准化问卷）不容忽视。根据调查表答案可以得出定量或半定量结果。

半定量暴露分类:

- 吸烟的强度和时间;
- 低/中/高危险工作的就业史;
- 生活水平相对于本地区处于贫困线以上/贫困线/贫困线以下。

例如:

一些研究也经常采用个体自报告的暴露信息。在加拿大的 4 个城市进行了一项人群基础的哮喘研究, 该研究应用了欧洲社区呼吸健康监测方案。在信访问卷的 12,000 名应答者中哮喘的现患率是 9.6%。该问卷还利用简单问题对灰尘和烟雾职业暴露的作用进行了估计: “在现在的工作中, 你暴露于灰尘、气雾或烟尘吗?” 这一“暴露”几乎增加了 1 倍的哮喘发病相对危险度 (OR=1.92, 95%CI: 1.6, 2.2)。

参见关于确定变量以及疾病测量的相关章节练习内容。

6.5 发病或死亡的报告: 率的计算、危险度指标

6.5.1 率的计算

为了进行比较并得出结论, 测量值必须转换成数值后再计算相关率。

流行病学测量值是对特定人群和特定时间而言的。例如某个率可能是单位人群 (通常是 10 的倍数, 如 100,000) 单位时间 (常常是每月或每年) 内某健康相关状态/事件 (病例或死亡) 的发生数。这些数字可以在不同人群之间进行比较 (例如某人群的死亡率是 12/10 万人年, 而另一个人群则为 55/10 万人年)。率的这种表达形式需要通过分母和分子来获得一个概率或比值。

在分母或分子的测量过程中可能会产生误差和偏倚。

常规统计数字来源的分子误差可能源于以下变异:

- 诊断试验的应用过程;
- 病例确诊过程;
- 所采用的信息记录体系。

分母误差则可能来源于:

- 人群迁移;
- 人群结构的变异;
- 行政区域的变化。

现患

指特定人群在某特定时点上某事件或某疾病的数目（特定现患数）。也可指一定时期内的事件数，那么该指标就称为期间现患数。将事件数除以危险人群数就是患病率。

患病率最常用于病程较长相对常见的慢性病，如哮喘和慢阻肺。

发病

指特定人群在特定期间内某疾病或相关条件的新发事件数。当其表达为单位时间单位危险人群的新发事件数时，就是所谓发病率。死亡率是发病率的特殊形式，说明某单位危险人群在单位时间内新事件（死亡）的发生数。发病率可以是原因别发病率也可以是某组别发病率（如年龄、性别、民族）。

发病率常用于相对少见（如结核病）或病程较短疾病（急性呼吸道感染）的频率测量。发病率常用报告率进行估计。本方法只有在病例确定比例相对高并具备有效报告体系时才适用。

卫生服务的利用率

通常用卫生服务的利用率（诸如住院率、住院时间以及意外就医）指标来估计某疾病的影响（如由于哮喘导致的意外急诊就医）。这些率既能反映疾病的经济负担，也能反映社区的一般性健康负担。为了充分反映后者，服务设施必须易利用且公平，而且关于卫生服务利用率的报告也要全面、准确。

不同测量指标间的联系

疾病（和死亡）的不同测量指标之间存在联系。因此：

患病率=发病率×病例平均病程

死亡率=病死率×平均死亡病程×发病率

例如:

如果某疾病的平均病程超过一年,那么该病的患病率将大于年发病率;如果病程较短(如几天或几周),患病率将小于年发病率。如果某病的病程很长(如哮喘),那么其患病率数据意义较大;如果病程很短(如儿童肺炎),则发病率(或死亡率)指标更有意义。对于传染性疾病(如结核病),患病率指标特别重要,因为它不仅反映了疾病的病程,而且是决定疾病传染给社区内其它易感者的因素之一。

疾病的测量受医学干预措施的影响。正确的治疗措施将显著缩短病程(降低患病率)。任何有效的治疗措施都将迅速降低病死率并因此降低死亡率。

例如:

结核病治疗可以迅速缩短病人的传染期。但如果治疗不当,它又通过挽救未愈传染性病人的生命而延长了病程导致患病率的增加(并因此增加了疾病的传播)。

6.5.2 危险度指标

危险度是指从一种状态转换成另一种状态的概率。它可以反映某影响因素存在条件下发病的概率,也可以反映接受某种治疗后疾病痊愈的概率。危险度是测量人群中疾病及其动力学的通用指标。许多流行病学研究的目的是为了获得此类概率的估计值。危险度的结果通常表示为单位人群的率或比。

危险度的定义

流行病学研究的结果常常以概率或危险度的形式表达。危险度的常用表达形式如下:

绝对危险度

研究中观察或计算出的某人群中某事件的发生概率。常用暴露人群的粗发病率表示。

相对危险度

暴露人群与非暴露人群的疾病或死亡率之比(危险度比或率比)。也可以表达成暴露人群与非暴露人群的累积发病率之比(累积发病率比)。

危险度差

暴露人群与非暴露人群的疾病或死亡率之差。

比数比(OR)

病例组与对照组的暴露比值之比。当发病率较低时比数比与危险度比会相当近似。

归因危险度(AR)

暴露人群中某疾病或其他结局的率指标中可归因于暴露的部分。本测量值表达为暴露组发病率与非暴露组发病率之差。

人群归因危险度 (PAR)

某人群中某疾病的发病率中归因于某危险因素暴露的部分。反映了人群中所有病例中由于暴露所导致的比例。

危险度的计算

		病例		合计
		是	不是	
暴露	是	a	b	a+b
	不是	c	d	c+d
合计		a+c	b+d	N

$$\text{绝对危险度} = \frac{a}{(a+b)}$$

$$\text{相对危险度} = \frac{\frac{a}{(a+b)}}{\frac{c}{(c+d)}}$$

$$\text{危险度差} = \frac{a}{(a+b)} - \frac{c}{(c+d)}$$

$$\text{比数比} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

$$\text{归因危险度} = \frac{a}{(a+b)} - \frac{c}{(c+d)}$$

$$\text{人群归因危险度} = \frac{a+b (AR)}{N}$$

第七章

7.研究实施的实际步骤

研究项目可分三个阶段。第一阶段本教程中已有介绍，即*研究方案的制定*：在研究开始之前必须详尽地制定研究方法和研究计划，并将它准确描述出来。第二阶段是*项目的执行*，必须采取某些措施来保证项目按时按预算完成，收集到的资料质量可靠，能够实现研究目的。这些步骤就叫作研究实施计划。最后阶段是对结果的解释和报告。

7.1 项目执行：假设你的研究方案获得了资助—现在怎么办？

7.1.1 研究实施计划

研究实施计划首先应把所有的研究任务罗列出来，再将各个任务的完成责任分配至研究组的某个成员。根据对完成研究任务所需要的时间和培训水平的估计来确定实施研究所需要的人员。

然后，制定完成上述活动的时间表，包括研究的所有关键阶段：

- 召集研究成员；
- 培训；
- 完善研究技术和研究程序的预试验（必要的话）；
- 召集研究对象；
- 实施测量；
- 数据核对；
- 结果分析；
- 准备发送给研究对象和相关社区的研究报告，
- 准备最终研究结果的科学报告。

在确定了研究成员的类型和数目后，应将研究组每个成员所承担的责任罗列出来。主要研究成员的责任包括：

- 项目管理；
- 质量控制；

- 对外联系；
- 伦理学问题的处理；
- 研究结果的分析以及向研究对象、卫生当局及科委报告研究结果。

7.1.2 研究实施过程和研究技术的质量控制

研究实施必须遵守研究方案中所制定的规程。

需要规定评价信息和测量值质量的精确指标以及对这些信息进行记录和解释的技术。

可以采取多种措施以保证研究质量的高水平：

■ 新研究方法的预试验

在研究开始前必须对操作程序进行测试，以确保其正常运行。例如，研究方案计划每天召集 10 名新病人，则必须核实这是否可行（可以通过现有记录核实，也可以通过一个小规模的预试验来核实）。同样，所有新设备均应测试校正，调查表也应经过预调查以保证应答者能够如你所期望地完成表格填写。许多研究者在制定主要研究方案前，已经完成了一项有资助的预试验。

■ 操作程序手册

- 对于数据收集每一步骤的操作说明；
- 对于完成数据收集后保持对信息追踪的操作说明。

必须制定操作程序手册以详尽描述这些研究技术的操作方法。在某些情形下，已有现成的操作说明；否则就需要为本研究专门制定。另外，必须制定一项详尽的数据管理计划，要特别关注所获数据的行踪、保证数据的质量控制、误差核实并保证结果的安全保存。

■ 研究人员的培训

负责信息收集的研究人员必须接受培训，即使是那些临床专家型（如临床医生、呼吸系技师、放射学家）调查人员，研究者也需确保他们理解了本研究目的。例如：询问研究对象疾病史的操作程序不同于完成调查表的操作程序；由于技师在医院里面对的大多数病人是病情严重者，所以可能很难意识到在对待几乎完全健康的研究人群时需要遵循完全不同的研究程序。

■ 对不同中心、不同调查员的数据质量进行定期比较考核

由多个调查员进行变量测量或者在多个中心进行数据收集时，对不同调查员收集的数据以及同一调查员在不同条件下收集的数据进行比较非常重要。这将保证结果间的一致性（调查员外部和调查员内部比较），或者至少测量了不同中心或测量人员间的测量差异。

■确定每个研究成员的责任

保证研究项目特别是大型项目能够收集到高质量的数据，最后的方法就是将责任直接分配给课题小组或个人，并使其认可所分配的责任。许多流行病学研究项目都会产生大量的数据，这些数据可能不能在一部科学出版物或科学报告中完全反映。因此，很必要事先决定根据研究结果可以写出哪些报告（或大报告中可以包括哪些部分），以及研究组成员如何承担这些报告。

7.1.3 保证研究效率

下列措施有助于保证在规定的的时间和预算内完成研究：

- 写出时间进度表；
- 定期的进展会议或进展报告；
- 确定保证时间进度和资金预算的负责人；
- 所有研究成员共享资金预算信息，该预算不包括研究成员的工资，但应包括每个成员需要投入项目的时间；
- 对于多中心或多研究组的研究项目，应考虑以研究组或中心为单位进行独立预算（但应留下用于数据分析汇总的经费）。

7.1.4 严格控制项目的研究内容

研究者在研究项目进行中试图改变研究方案中某些内容的现象并不少见。其理由包括：

- 原定研究方法不奏效；

例如：

召集不到预期的病人；用来测量呼出气的仪器性能不可靠或者经常出毛病；研究对象对调查表中的许多问题都不能正确理解作答等。

- 发生了一些科学事件（出版了一篇新的科学论文，上市了一种新药），使你对本研究的假设或研究目的产生疑问；
- 初步研究结果提示初始假设是错误的（或提示该假设是正确的）。

对于这些情况的正确反应是暂停项目，召集研究组讨论下列问题：

- 全面停止研究（直到有了新的研究方案、时间表以及预算时再继续研究）。这种处理方法适合于前两种情形。这也是在开始项目前进行预试验的主要理由。
- 决定继续执行现有研究方案
这种处理方法适合于最后一种情形（即你认为自己的研究结果以后的走向可能会发生改变）。记住，人群的信息易呈簇状分布，这很重要。

例如：

由于未知的原因，你可能会发现研究对象中的哮喘患者主要出现在研究的头 1 到 2 个月，在后面的 6 个月里则很难遇到有哮喘症状的个体。如果你在 2 个月时停止研究的话，那么就很容易得到一个错误的哮喘患病率。

奇怪的是即使利用现有数据进行科学研究，也会出现数据成簇现象（如结核病登记数据）。

中途改变研究方案不可取。为什么？

几乎所有的研究项目都是匮乏的卫生资源与研究效力（即研究人群的规模）之间的妥协结果。如在本教程中所提到的，每个研究方案都包括了一个研究样本/研究效力的计算过程。如果中途改变研究方案的话，实际上是将当初设计的较大规模研究分成了两个小规模的研究（二者的研究方案略有不同）。这两个小规模研究说明研究目的的效力也就降低了。因为目前基金会的资助趋势是不资助超过必要规模的研究，所以可以认为，无论多么仔细地收集余下部分的数据，新研究都将无力检验研究假设。

但临床试验则可以根据其原始研究方案中的“中止准则”进行期间分析，即可以在特定时间点上（有足够的统计效力）进行期间数据分析。如果在该时点发现治疗措施的效果（阳性或阴性）很强，就可以停止研究（只要满足“中止准则”的标准）。注意，即使在这种情况下也只是决定停止研究，而不是改变研究方案，原方案要继续执行。

7.2 收集信息的处理

当你在某项研究中收集到了信息，这些信息绝不可能让你立即就可以看到结果。如你到目前为止所学，数据有多种形式。

其中包括：

- 从临床记录以及论文表格中提取的信息；
- 死亡证明书的复印件；
- 系列病人肺功能检测结果的计算机文件；
- 手填完成的调查表；
- 抄录的医院或诊所记录。

有经验的研究人员认识到收集数据将占用项目所需时间的 1/2。而涉及数据核对、组织、分析以及结果报告的步骤，一般至少要花费和数据收集一样多的时间。

将粗数据转换成“可分析”数据的步骤。

7.2.1 原始数据收集表和信息记录的核对

数据处理的第一步就是查阅原始数据的完整性和准确性。

文字记录的初始核对

- 小规模项目：在计算机录入信息前，对每页记录进行核对和修正；
- 大规模项目：对随机抽取的部分记录资料进行核对，找出常见错误。如果可能的话核对所有记录并修正正常出错的地方。

修正原始数据的一般原则

对原始数据的修正只限于明显的错误，不要对需要进行数值判断的地方进行修正。例如，对于从临床文件中提取出来的数据，如能证实原记录有误，则可以修改原来的诊断日期；如果只是觉得不可能是那个日期，则不要进行修改。

7.2.2 数据的编码

有些原始数据可以直接录入计算机（如出生日期、调查表的是/否型答案、吸烟的支数/天、结核菌素皮肤试验的反应测量值）。而对于其他类型的信息（如药名、死亡原因、临床诊断、工作名称），则需将这些信息分类成可供处理的组别，然后再对每个组赋予编码值，这个编码值即可录入计算机了。

数据编码的标准方法是建立书面编码手册，规定将文本信息转换成数字编码的准则。编码体系可以很复杂（如用几页篇幅叙述治疗方案—每种方案都有自己的编码），也可以很简单（如女性为“1”；男性为“2”）。

请将你的编码手册多备份几份！

预编码：

许多情况下，应该在收集数据前就选择编码方法。例如，如果准备收集关于处方药的信息，最好事先就决定好药物名称如何分类以及如何编码，然后将编码信息提供给数据收集人员。这样就能保证调查员（或进行编码转换的人员）能按照编码方式记录信息。

另一种方法是将编码值（以及编码原则）直接写到数据录入表格上。

可以选择如下编码方法：

1. 只记录编码值，不记录文本内容
 - 优点：减少了一步操作（少花时间、少犯错误），
 - 缺点：以后不能对编码工作进行查证。

2. 记录文本内容，但只录入编码

—优点：减少了一步操作（少花时间），

—缺点：需要返回到文件记录中去查证编码。

3. 直接录入文本型信息，然后用计算机进行编码

—优点：减少了一步操作（少犯错误），可以较容易地查证编码，

—缺点：对于非标准拼写或非标准格式文本，不能进行计算机化编码。

过去由于计算机内存的限制选项3是不可行的。如今已可把文本型信息直接录入计算机，但编码工作仍要通过非自动形式完成（即研究者或助手直接将文本信息编码到计算机）。但其后的编码查证却大为方便。

一般而言，在原始数据与计算机储存数据之间的重写操作次数越少犯的错误也就越少。

编码核对

应该对编码工作进行查对（就象处理原始数据那样）。查对编码结果的标准方法是在原始资料中抽取一个随机样本进行重编码（不知道初始编码）。如果发现两次编码的差异很明显，则需要对所有资料进行重编码再检查其差异。如果在选择编码选项时需要作出判断（如诊断、死亡原因、工作、胸部放射学检查的异常表现），那么编码核对就特别重要了。

7.2.3 资料的计算机录入

将数据录入成计算机文件的方法很多：

- 按照数据收集格式直接将其录入计算机（如计算机化的调查表）；
- 计算机打孔编码法；
- 光学扫描或条码。

记录和变量

一条记录即所收集到的关于某个体的全部信息（假定观察单位是人）—就象电子表格中的一行。

一个变量即所收集到的全部研究对象的某特征信息，如年龄、身高、诊断、目前咳嗽情况等—就象电子表格中的一列。

一些计算机的数据录入程序可以进行自动错误核对（如 Epi Info）。计算机不可能发现所有的错误，但可以通过设置数据录入格式来检查超出容许值范围的或不合理的数值。

7.2.4 计算机文件的核查

在所有数据录入为计算机文件后，必须对计算机文件进行核查。

a) 核查超出容许值范围的数值以及无效值

对于分类变量（如是/否型，编码变量），让计算机给出每个变量的频数分布（即每个取值的出现频度列表），并检查该结果。对于连续性测量值变量（如年龄、身高、FEV₁），指令计算机给出每个变量的最大值和最小值，并检查该结果。

b) 非显值（如某个体无结核病诊断，其治疗变量就不应有记录值）的核查：

指令计算机给出两个变量的交叉列表，检查逻辑上不一致的结果（如某个体的教育编码为未读完中学但其职业编码却为医生；一个 12 岁的小孩却有三个孩子；没有吸烟时间信息记录的吸烟者等）。

c) 核查看起来不可能的结果

仔细检查每个变量的频数分布（或均数以及标准差），并仔细考虑频数结果是否合理。例如，已完成了一项对某地区全部居民的调查，则应核对其年龄和性别分布。如结果与已了解的信息不一致，那么可能在数据录入方法上有重大错误。再如某研究的对象是护士，但发现 80% 为男性，那么可能是编码员把编码编反了。

“每个有经验的研究者都知道，即使是最认真的数据收集工作，也会在细致的数据编辑过程中出现错误。如果数据编辑过程是研究计划中数据处理的一个常规过程，那么此类错误就不会引起严重问题。但如果忽略此类编辑过程的话，问题就发生了。”

摘自 Rothman KJ, Greenland S. 著现代流行病学, Lippencott-Raven. 1998

7.2.5 缺失数据的处理

无论多么仔细地进行数据的收集和核对，资料仍然会存在缺失数据。你必须决定如何处理缺失数据。这是为什么？

例如:

假定你完成了一项小规模研究(100例病人):调查病人对于结核病治疗效果的知识和态度。在研究人群中有12个病人属于某特殊社会群组,其中10人为耐多药结核病人。这时也许你想检验属于该社会群组是否与耐多药结核病的发生有关联,或这种显而易见的联系是否由于其它原因(如教育、HIV感染状态、年龄等)造成。然而这12人中8人的HIV感染状态信息缺失。

分析结果时,如果想同时运用该因素(即属于该社会群组)和HIV感染状态信息,就只能包括4个病人。但进行单因素分析时,则可以包括12个病人。如果不对缺失值进行核对,则结果肯定令人混淆并且难以解释。

缺失信息的类型

■ 随机缺失

不是指真的随机,只是说这些数据的缺失与研究中的其它重要因素无关。

例如:

因为某一天肺活量计坏了,所以一小部分个体的肺功能结果缺失。这些数据缺失非随机发生,它与某检测日期相关联。但只要那一天没有发生什么对研究很重要的事件,这些结果的缺失就可以看作是随机的。

■ 系统缺失

这是缺失数据的最常见类型。它意味着这些数据的缺失与研究中的其它重要因素有关。

例如:

如果利用某州收集的死亡记录来研究死因,可能会发现时间越早,记录越不完整。这就是说研究中最年老个体的出生日期信息可能缺失。若研究疾病与年龄(或与历史上的某期间)有关,则会因为研究排除了许多早期出生的个体(至少在此分析中如此),而使研究存在偏倚。

如何处理缺失值或存在明显错误的信息

1. 删除

在卫生研究中最常用的方法是(从分析数据中)将带有缺失数据的研究对象删除。如果选择了这一处理方法,那么就得考虑剩余的研究人群是一个修正过的人群。也可以只删除那些缺失最重要变量数据的个体,但当利用其它变量分析结果时必须十分当心了。

2. 替代

a) 以均数替代:

如果能够证明缺失信息是“随机缺失”的话，就可以用其它所有个体的均数来替代缺失值。

例如:

当研究一个成年工作人群时，发现 2% 的个体缺失出生日期。首先要进行核查来确定这种缺失是否与研究中的其它因素无关。如果能肯定这一点，就可以简单地将这些个体的出生日期赋值成研究人群的平均出生日期。

这种处理方法常用于社会科学研究——如需要大量个体对调查表进行回答的研究。

b) 用估计值替代 (要特别当心)

有时可以对缺失值或错误值进行一定准确度的估计。

例如:

在墨西哥城进行了一项空气污染和死亡率关系的研究。研究者发现许多官方记录中登记的死亡日期与医院记录不一致。他们有两种选择：赋予这些个体一个死亡日期（选择一种“更可信”的来源，或取两种来源取值的均数），或者将其从研究中删除。在该研究中，研究者对死亡日期与该日期相应期间空气污染程度之间的关系感兴趣，因此必须删除这些缺失值个体。然而，如果他们的目的只是描述月或年基础的死亡率模式，那么就可以根据数据记录中的其它信息赋予这些个体一个死亡日期。

8.研究结果的解释

8.1 保证结果的有效性

一项科学研究的价值在于根据其研究结果得出推论的程度，特别是其研究样本的结果可以外推到人群的保证程度。除了研究方法内在的局限性会影响结论的外推外，样本的代表性、总体人群的特征等也会损害研究的有效性。在本章节，将对其中一些问题进行讲解，并讨论如何避免这些问题；或者如果这些问题不可避免的话又该如何处理。其中最重要的问题是误差和偏倚。

8.1.1 定义

误差：在流行病学研究中可能会产生两类误差：

随机误差：指测量值的无规律不准确性，随机误差是因机会、检测方法的不精确或测量技术的不精确造成的。

系统误差：测量值偏向同一方向的不准确性，系统误差是由于仪器出错、测量技术不当或研究设计的缺陷造成的。

偏倚：偏倚导致变量的测量值和真实值之间出现同一方向的差异。偏倚可以产生于资料收集、数据分析、结果解释、数据的发表或评阅等任何阶段。它能导致：

- 错误估计某种研究条件的发生频率；
- 对某病的患病影响因素作出错误的结论；
- 关于某疾病与影响因素之间的联系作出错误的结论。

要想降低发生误差和偏倚的机会，就必须对其来源以及在研究设计和实施阶段应该保持的严密性有清晰的认识。尽管可以对偏倚（不是误差）进行一定程度的校正，同时/或者可以采用适当的分析手段测量其后果，但误差和偏倚还是在所难免。

流行病学研究的资料分析涉及各种率或平均值的计算和解释以及组间率或平均值差异的比较。样本的率或某变量平均值的计算很容易发生误差和偏倚，而这些误差和偏倚会使组间比较结果失去意义，从而得出错误的结论。误差可能产生于研究样本的抽样方法，也可能产生于研究变量的测量方法。本章节将对此进行详细讨论。

8.1.2 研究人群抽样过程中的误差和偏倚（选择偏倚）

抽样方法不当

目标人群必须适合于研究目的。如果所选择的目标人群不适合于研究目的，研究结果就可能不具有普遍适用性。

例如：

如果选择了一个有毒因子暴露水平非常低的工作场所进行研究，可能就发现不了与暴露有关的负面健康效果。再如，如果存在健康工人效应（即工人对研究工种的选择取决于其是否出现了与影响因素暴露有关的症状或效果）的话，就可能发现不了影响因素的作用。

可疑因子的暴露频率必须足以产生可检测到的负面效应，而且研究人群的工种选择也不应该与其是否出现相关症状有关，因为这将使暴露因素的效应难以检出。

其他常见问题：

- 未能准确规定研究人群或者病例抽样人群；
- 未能调查目标人群或样本人群中的所有合格对象；
- 偏离了样本抽取规则，导致所选择的样本不能代表目标人群；
- 遗漏“很难找到”的个体；
- 无应答率高（由于参与研究的动机不同，如参与研究者可能会认为参加研究更容易接受治疗服务，而无应答者则考虑参加研究会影响到其工作等，因此与应答者相比，无应答者可能包括较多或较少的暴露或患病个体）。追踪研究样本中的所有个体使其参加研究是非常重要的；
- “中途退出率”高，特别是当这种退出与可疑因素或某干预措施的暴露有关时；
- 无应答者、样本中难以找到的个体或中途退出者的替代。只有在此类人是错误纳入研究样本时，这种替代才是合理的。例如利用一个过时的名单进行抽样。不可使用志愿者来替代上述个体；
- 开始时的暴露者在研究开始后改变了其暴露状态（“污染”）。

避免样本选择中的误差和偏倚。通过采取下列措施可以避免此类误差和偏倚或者使其最小化：

- 利用最新的完整的人群名册进行抽样；

- 努力避免给参与者造成不便或不适，鼓励所有样本个体参加研究；
- 追踪并说服无应答者参与研究调查；
- 检查应答者与无应答者某些特征的相似性，以保证二者在诸如年龄、性别、婚姻状况以及职业等关键变量方面无系统差异存在。

在分析过程中降低由于研究人群丢失造成的误差或偏倚效应：

在分析阶段有时可以利用下列方法之一对由于无应答以及“失访”所造成的偏倚进行校正：

- 从分子分母中剔除此类个体；
- 在“暴露时间”内包括所有个体；
- 在一半的“暴露时间”内包括所有“丢失”个体，就是假定这些丢失个体平均暴露了一半的研究时间；
- 假定所有“失访”个体患病或发生最坏结局，然后再做相反假设，这样就可以计算出两个危险度比，也就得到了一个包含“真实”值的区间。

8.1.3 变量测量中的误差和偏倚（信息偏倚）

在数据测量或数据收集中产生的误差和偏倚来源，可以分为以下几类：

- 研究对象变异；
- 观察者变异；
- 技术方法的局限性。

*研究对象变异*指对同一对象在不同条件下测量值间的差异，其产生原因可能是：

- 生理性变化；例如：哮喘病人的肺功能；
- 影响研究对象对研究作答的因素发生变化，包括对过去事件回忆的准确性、动机、情绪、对调查环境的反应以及对调查员的信任程度等；
- 作为研究对象而导致个体产生的变化（霍桑效应）。

*观察者变异*指同一调查员对同一研究对象在不同条件下进行测量时的测量值差异（调查员内部误差），或指不同调查员对同一研究对象在同一条件下进行测量时的测量值差异（调查员间误差），产生此类误差的原因可能是：

- 调查员知道了研究假设，如知道吸烟与呼吸道症状有关；

- 检测试验操作误差或问题表达中的语法错误，例如，某调查员通过对问题语法的安排提示“可接受”答案；
- 调查员粗心、操作技巧或经验不足，如遗漏了调查表中的一些问题或在测量肺功能时未能让研究对象使出最大力量呼气；
- 检测试验操作过程中的偏倚，例如，在结核菌素皮肤试验中阅读结果时发生的“末尾数字偏好”。

技术性误差是由于测量工具失效而导致结果不准确或不正确，其原因有：

- 检测试验不适合于本研究目的，例如：运用胸部放射学检查来确定结核病的活动状态；
- 测量仪器不可靠或不准确，导致测量结果无重复性或测量结果与测量疾病的严重程度无关，例如：“小气道功能”检测试验用于哮喘调查；
- 检测体系存在错误，例如：肺活量仪漏气或结核菌素成分不当；
- 所采用的数据收集方法有利于某种暴露或结局信息的收集而不利于其他信息的收集，例如：煤矿的粉尘测量只记录了整个班次的平均粉尘水平，而未反映高峰暴露频率；

例如：

如果未对测量（例如肺活量测量值）进行标准化，那么研究结果就可能没有什么意义。一个经典示例是佛蒙特州花岗岩工人研究，该研究中肺活量仪在某测量现场出现了微弱的漏气，从而导致了肺活量随时间降低的错误结论。注意：如果肺活量仪的读数降低幅度在整个研究中都一样，研究中肺活量值在不同组别以及不同时间上的比较是有意义的，但与外部人群的比较就没有意义了。

随机误差模糊组别间的真实差异，从而导致研究措施无效的错误结论；而系统误差（偏倚）则可能导致实际上不存在但表面上明显的差异。

如何减少信息收集中的误差和偏倚

下列措施将有助于避免上述误差：

- 必须明确制定诊断标准并严格执行，即使有漏诊少数病人的风险也要如此；
- 对于严重程度或临床分级的分类，尽可能使用定量指标，而且必须覆盖所有的可能性；
- 必须对所有研究对象进行调查观察，并在相似的条件对研究对象进行检测，要尽量让他们感觉放松；

- 问题越简单越好；可以运用核对性问题来检测答案的一致性；（如“你用力时感到气短吗？”和“你匆忙时感到气短吗？”）
- 调查员数目应该最少；在接受培训后，应该通过不同调查员对同一研究对象或同一调查员对同一对象在不同条件下进行重复测量来检测调查员内部和外部变异；
- 研究对象和调查员应该不了解研究假设，以避免个人期望对测量产生影响从而使结果发生偏倚；
- 检测试验应适合于研究疾病的诊断，而且要准确、可重复以及研究对象乐于接受；
- 检测设备要牢靠、简单、可靠并易于操作；
- 检测方法和调查表应该标准化并且要实施定期的质量控制程序，以保证结果的准确性未发生“偏离”。

8.1.4 检测试验或检测工具内在误差的评价

考虑到检测试验固有的错误倾向性，可以用几个指标来反映一个检测试验给出有效信息的效力。

辨别力指检测试验能够将某病患者（或有某特征者）与非患者（或不具有这些特征者）区分开来的能力，或者指检测试验对研究对象的某种特征严重程度准确分级能力（例如，哮喘病例的高峰流量测定、结核病的痰涂片显微镜检查）。

可再现性（可重复性或精确性）是衡量一致性的指标，指某检测试验或某问题对同一对象在相似条件下进行连续测量给出同样结果或同样答案的能力。可以从以下几个方面评价检测试验的可重复性：

- 同一或不同调查员利用同一仪器对同一组对象（或一套标本）重复检测试验或重复问题询问；
- 将正在使用的检测试验系统与其它测量工具或测量系统进行比较；
- 利用核对性问题进行可重复性评价（关于相似特征的问题应该得出类似的答案一如年龄和出生日期）；
- 将研究对象随机分配给调查员进行重复观察。

可重复性的测量指标是：

$$\text{符合率}(\%) = \frac{\text{获一致性结果数}}{\text{检测总次数}} \times 100$$

效度（准确性）某项测试给出真实结果的能力。也就是检测试验能够正确检测出某条件的有无或者将个体正确分类的能力。

效度有两个组分：

- **灵敏度**，指检测试验能够正确鉴别某病患者的能力；
- **特异度**，指检测试验能够正确鉴别某病非患者的能力。

在分析连续性变量时，确定何值为阳性和阴性的判别切点常常涉及到对灵敏度和特异度之间的平衡。

预测值是检测试验对阳性或阴性结果正确预测能力的测量指标。预测值可分为阳性预测值和阴性预测值：

$$\text{阳性预测值} = \frac{\text{检测试验阳性病例数}}{\text{所有病例数}}$$

$$\text{阴性预测值} = \frac{\text{检测试验阴性非病例数}}{\text{所有非病例数}}$$

在灵敏度和特异度不变的条件下，预测值将随着患病率的增加而升高。在给定患病率和灵敏度的条件下，预测值则随着特异度的增加而升高。

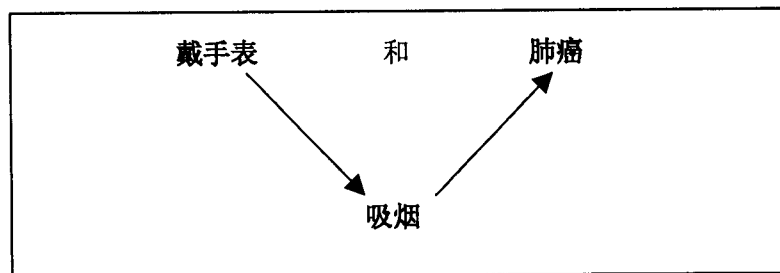
8.1.5 混杂（选择偏倚的一种特殊类型）

研究结果中观察到的联系，有可能受到与影响因素（危险因素）和研究结局都有关联的外部变量的影响——这就是所谓的混杂联系。

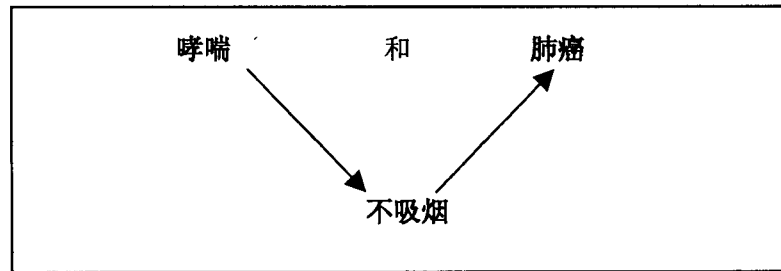
定义

某危险因素与某疾病之间的阳性或阴性关联，是由于该因素与研究疾病其他病因的联系造成的，而其本身非该疾病之病因。

如假阳性：戴手表引起肺癌



如假阴性：患哮喘可以预防肺癌



混杂的控制

研究设计阶段的控制措施

- 随机化：如果样本很大，该措施可以保证重要的混杂因素在比较组间分布相同；
- 限制：将进入研究样本的个体限制在重要混杂因素规定的取值范围内（如规定研究个体为某个年龄组或某个种族）；
- 匹配：保证重要混杂因素在研究组间的分布比例相同（如在上述两个示例中按照吸烟习惯进行匹配）。注意，这种情况下存在发生“匹配过头”的危险性，亦即一旦作为匹配变量，就无法评价该变量作为危险因素的作用。

在资料分析阶段的控制措施

■ 分层分析

在混杂因素的分层内计算反映研究因素与研究疾病间联系的危险度估计值；在后面的统计学分析章节中，列举了一个利用分层分析控制混杂的示例。

例如：

如果对两个年龄结构差异很大的人群进行死亡率的比较，可以首先将该人群分成不同的年龄组（层），然后在每一层内计算死亡率，并对两个人群进行比较。

■ 标准化

另一种控制年龄混杂作用的方法是计算一个考虑了两个人群年龄结构差异的综合值。这就是所谓年龄标化。尽管理论上也可以对年龄以外的因素进行标准化，但实际上绝大多数的标准化都是为了控制年龄的混杂作用。

间接标化

本方法是将某一般人群（或标准人群）在某层（按照年龄、性别等分层）内特定条件下的死亡率运用于研究人群，获得研究人群的期望死亡数（按照一般人群或标准人群的水平）。在每一层内或在整体上将该死亡数与观察死亡数进行比较，得到一个比例（即标准化死亡率比 SMR）。

直接标化

本方法是将研究人群某层（如年龄组）的死亡率运用于标准人群的年龄结构上。在本方法中，通过将不同比较组的年龄别死亡率运用于同一标准人群，分别计算每个研究组的期望死亡数，从而调整了年龄结构差异的影响，便可对两个年龄结构差异很大的人群进行死亡率比较了。

参见关于偏倚、标准化和混杂的相关章节练习内容。

8.2 了解结果的含义—数据分析

作研究就象作雕刻家。从一块普通的花岗岩开始，在这块花岗岩里是一个精美的雕像。雕刻家的任务是将这个内部的雕像发掘出来。

同样，研究者需要处理的也是这么一块石头（信息集—通常是收集并抄录下来的一套数据）。在这块石头里藏着我们必须发现的图案。就象没有使用合适的工具或者在雕刻过程中不细心就不可能发掘出那个雕像一样，如果研究者在分析这“块”信息时不细心，就可能发现不了存在于其中的新知识。

参见关于统计学分析在研究中的作用相关章节练习内容。

对统计学分析方法进行概括介绍的参考文献有：

McNamee R, Cockcroft A. Statistical and epidemiology reviewing. Occup Environ Med. 1994; 54:721

Greenhalgh T. How to read a paper. (a series of articles from the British Medical Journal;1997:315):

The Medline database. p180-3

Getting your bearings (deciding what the paper is about). p243-6

Assessing the methodological quality of published papers.p305-8

Statistics for non-statistician. I: Different types of data need different statistical tests.p364-6

Statistics for non-statistician. II: "Significant" relation and their pitfalls. P422-5

Papers that report drug trials. P480-3

Papers that report diagnostic or screening tests. P540-3

Papers that tell you what things cost (economic analyses). P596-9

Papers that summarize other papers (systematic reviews and meta-analyses). P672-5

Papers that go beyond numbers (qualitative research). P740-3

Motulsky H. Intuitive Biostatistics. New York: Oxford University Press,1995.

8.2.1 熟悉数据

组织数据的第一步是熟悉数据。进行研究设计、数据收集需要大量的时间，检查和总结数据同样需要大量的努力，否则就不可能很精确地了解这些资料。

在没有进行数据检查之前，不要急于着手分析结果。

步骤 1—检查变量

准备一系列的表格来显示每个变量的数据分布。检查每个变量的分布—检查每个变量的含义并确定其分布形状。

提示：为每种变量类型制作一个表格（即一个表格显示“是/否”型或离散型变量；一个表格显示所有具有少量答案分类的离散型变量；另一个表格则用以显示所有的连续性变量）。

步骤 2—对重要信息进行归纳总结

离散型（或分类）变量：

- 计算具有某特征人群的率、构成比或百分比（频数分布）；
- 给出相关图（条图/饼图）表。

连续性变量：

- 画出直方图；
- 给出最小值、最大值、均数（或中位数）以及变异的测量值（如标准差）。

EPI-INFO 注释：

- 在 EPI-INFO 中要想得到一个连续性变量的频数直方图，需要将该变量分类化（转变成一个新变量）
- 如：定义年龄组变量_____
- 将年龄按 5 岁一组年龄组化
- 给出年龄组的直方图

步骤 3—非统计学检验比较

离散型变量：

- 对相关的率（或构成比）进行简单的“肩并肩”的比较；
- 交叉制表；
- 计算率的差值（即一个构成比减去另一个构成比）；
- 计算率比或比数比（一个比除以另一个比）。

连续性变量：

- 对相关均数（或中位数）进行“肩并肩”的比较；
- 将两个频数直方图重叠起来进行比较。

参见关于非统计学检验条件下的研究结果相关章节练习内容。

8.2.2 统计学分析方法

为了完成对数据的统计学分析，首先必须确定：

- 数据或测量值类型（离散型或连续性）；
- 希望进行何种比较或描述。

统计学分析中最困难的任务是决定进行何种比较。一旦决定了如何列示结果以及进行哪些比较，就可以对这些比较进行统计学检验了。

回顾—仔细想想：

- 1.研究问题是什么？
- 2.这是一个不需要进行比较的简单描述性研究吗？
- 3.回答研究问题需要进行哪些比较（回顾研究假设）？
- 4.需要对两组或多组（或亚组）进行比较吗？
- 5.只是需要在同一组内进行两个或多个测量值的比较吗？
- 6.要把研究因素表达为一个比例（或率）还是表达为一个连续性值？

8.2.3 用于不进行比较条件下描述结果的统计量

a) **离散型数据**：计算相关率及其 95%可信区间

如在一个 2000 名儿童的样本中，发现了 123 例哮喘病例
患病率： $123/2000=6.15\%$
95%CI: 5.1%-7.5%

b) **连续性数据**：计算相关平均值及其 95%可信区间

如调查一个 165 名成人样本的年龄
平均年龄（岁）：42.7（标准差=18.5）
95%CI=42.5-42.9

EPI-INFO 注释:

- 要想计算一个率的 95%CI, 命令是: **FREQ 变量名/C**
- EPI-INFO 中没有对均数 (或平均值) 求解 95%CI 的命令, 这里给出其计算公式:

$$95\%CI = \text{均数} \pm 1.96 \times \frac{\text{标准差}}{n-1}$$

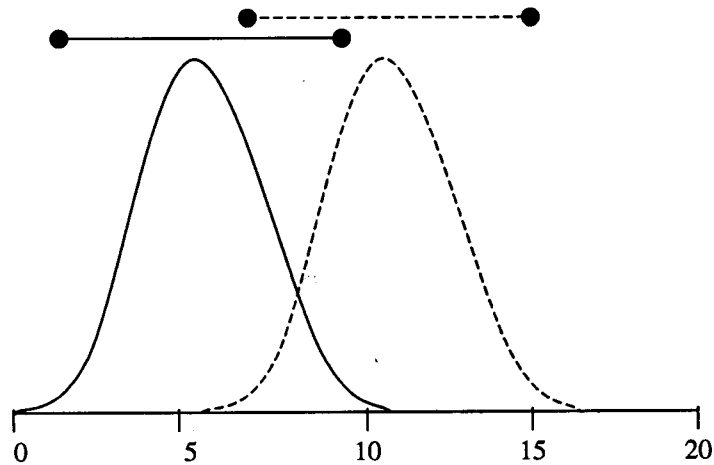
什么是 95%可信区间?

当你选择了一个人群样本 (从目标人群中) 进行研究, 那么研究结果 (如哮喘患病率) 对于样本来说是准确的。然而研究的目的是要了解目标人群 (不仅仅是样本) 的 “真实值”, 因此弄明白样本研究结果是否能够反映目标人群的真实患病率就非常关键。

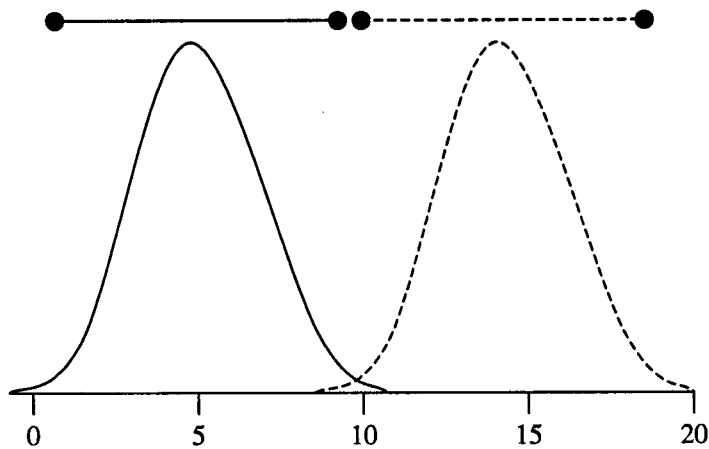
可信区间就是给出目标人群真实患病率分布于其中的一个区间范围。它根据研究观察到的率 (就你的研究而言)、样本量以及根据概率论推导出来的公式进行计算。

为了理解 95%可信区间的意义, 想象在同一目标人群中, 利用相同方法和同样的样本量重复了 100 次同一研究, 只是每次选择的对象不同。如果真这么作了, 就可以计算每次研究的哮喘患病率, 这样将获得 100 个略有不同的患病率。绝大多数患病率是很接近的, 另一些则可能表现出较大差异 (仅仅由于机会或随机变异造成)。

所谓 95%可信区间就是指其中 95 次研究结果所在的区间。可信区间越窄, 越有理由认为研究结果是目标人群真实率值的良好反映。



然而，如果这 100 个样本是从另外一个（不同）总体中抽取的，并与以实线表示的样本人群进行比较的话，结果就可能如下图，其 95% 可信区间不重叠。



8.2.4 离散型变量两组间比较的统计学检验

a) 两组（或两个变量）间的比较—离散型变量，各有两个类别

两分类取值离散型变量示例：

- 性别（男/女）；
- 居住条件是否拥挤（是/否）；
- 有无接受适当的病例管理（是/否）；
- 是否接受教育性干预措施（是/否）；
- 是否在工作场所暴露于高浓度的刺激性气体（是/否）；

- 结核菌素反应性是否阳性（是/否）；
- 有无患哮喘（是/否）。

例如，需要回答这样一个研究问题：哮喘与刺激性气体的职业暴露有无联系？

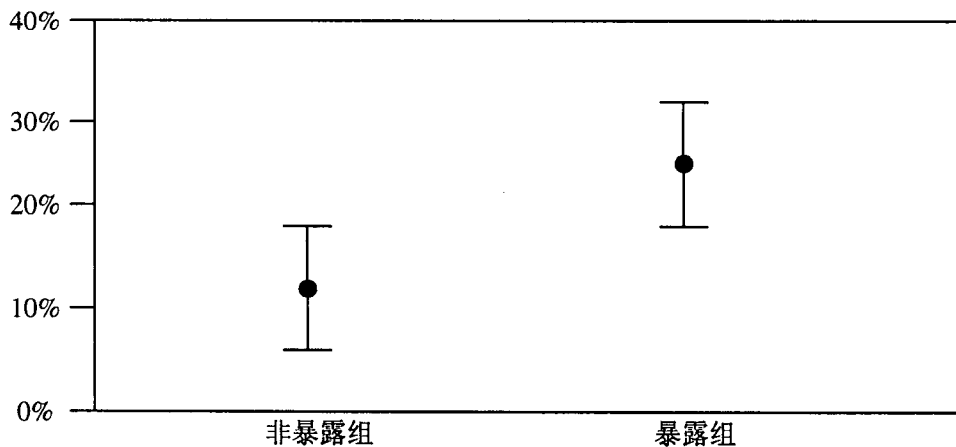
可能的解决方法有：

(i) 计算两个率并计算每个率的 95% 可信区间（并进行目视比较）

例如：

暴露于高浓度刺激性气体的工人中哮喘样症状患病率是： $48/199=24.1\%$ ，95% CI: 17.1-30.6 非暴露工人中哮喘样症状患病率是： $16/130=12.3\%$ ，95% CI: 6.9-19.3
--

哮喘样症状的患病率



这种目视方法可以良好地反映两组工人间哮喘样症状患病率的差异幅度（效应大小的反映）以及该患病率真正不同的可能性。如果 95%CI 大部分重叠，就不可能得出两组患病率不同的结论。

为了计算实际的统计学概率，则需应用卡方检验。

(ii) 卡方检验

利用卡方检验检验两组工人（即暴露于刺激性气体的工人和非暴露工人）的哮喘样症状患病率实际上相同的统计学概率。本检验方法实际上是检查这两个率来自同一总体或同一分布（由于机会造成其表面的差异）的概率。

卡方检验（比较 48/199 和 16/130 这两个率）：P=0.008

P 值的含义是什么？

对于本结果的解释是这两组工人的哮喘样症状患病率真正相同（即来自于同一总体）的概率大约是 0.008。

EPI-INFO 注释：

●要想获得两个率比较的卡方检验结果，命令是：

SET PERCENT=ON

TABLES 变量 1 名 变量 2 名

这里变量 1 指影响因素，变量 2 指结局或疾病

这些数据也可以用其它方法来说明（2×2 表）：

	有哮喘样症状	无哮喘样症状	合计
暴露	48	151	199
非暴露	16	114	130

也可以利用 EPI-INFO 中的 STATCALC 程序（在该程序中可以按照上表格式录入数据）直接计算卡方值。

如果研究规模较小，并且 2×2 表的一个或几个格子中的人数小于 5，就得应用适用于这种条件的特殊形式的卡方检验（如 Fisher 精确概率法或“Yates corrected”卡方检验）。EPI-INFO 同时给出一般卡方检验和特殊形式卡方检验的结果。一般而言特殊形式卡方检验的结果与普通卡方检验的结果相似。

(iii) 计算两个率，率比（相对危险度）及率比的 95%可信区间

利用同一示例，通过计算暴露于高峰刺激性气体的工人和非暴露者的哮喘样症状患病率的比值，对二者进行比较：

$$\text{患病率比} = \frac{24.1\%}{12.3\%} = 1.96$$

并可以计算该比值的 95%可信区间：在本例中，95%CI=1.2-3.3

EPI-INFO 注释：

●要想求得率比及其 95%可信区间，使用的命令同计算卡方检验统计量：

SET PERCENT=ON

TABLES 变量 1 名 变量 2 名

这里变量 1 指影响因素，变量 2 指结局或疾病

如何解释率比及其 95%可信区间：

仔细想一想，如果两个患病率相同，其比值将等于（或接近于）1.0。因此，如果了解这两个患病率是否相等，只要知道该率比是否不等于 1.0（即大于或小于 1）即可。

在本例中，率比约等于 2（为 1.96），说明暴露组与非暴露组相比，前者的患病率大约是后者的两倍（1.96 被称为“点估计值”）。

95%可信区间提示我们，如果再从总体中重复抽样 100 次，并比较其哮喘样症状患病率，那么其中 95 次的率比将在 1.2-3.3 之间。相反，该结果也显示率比等于 1.0 的概率小于 5%（或 $p < 0.05$ ）。

率比的卡方检验及其可信区间提示的是同一件事，即“统计学显著性”。可信区间方法的优点是它同时给出了组间的差异幅度，而卡方统计量却作不到这一点。

要点：

在解释率比时，我们的兴趣在于：

- 率比自身的大小；
- 其 95%CI 是否包括（或排除）1.0。

b)两个以上组别间的比较—离散变量或两个以上分类的组别间比较

具有三个及以上类别的变量示例如下：

- 吸烟状态：从不吸烟、从前吸烟、现在吸烟；
- 哮喘严重程度：轻度、中度和重度；
- 居住地区：大都市、中小城市、农村；
- 种族。

(i) 计算每个组的率及其 95%CI（然后进行比较）。这与前述两个率的处理方法相同。

(ii) 卡方检验

本检验也与两个率的卡方检验相同，只是这里要对所有三个（或更多）率进行比较。其目的是确定这些率是否来自同一总体，是否有其中一个或几个率来自其它总体。

例如：

	非吸烟者	从前吸烟者	目前吸烟者	P 值（卡方检验）
研究总人数	259	267	196	
有气道阻塞 症状的人数 （%）	20 7.4%	34 13.2%	26 13.3%	0.05

(iii) 率比

有两个以上组别时，便无法只计算一个率比，因此必须靠卡方检验对两个以上组别进行比较。

在某些情况下可以计算三个或三个以上率中任两个率的比值。这只有在卡方检验提示组别间有差异，并且研究设计中就包括了对某两个组进行比较的条件下才合理。

例如，如果研究假设涉及对不吸烟者和目前吸烟者的比较，就得在上述示例中排除那些从前吸烟者，然后按照下式就气道阻塞和吸烟之间的关联计算相对危险度：

$$\frac{\text{吸烟者中气道阻塞的患病率}}{\text{非吸烟者中气道阻塞的患病率}} = \frac{13.3\%}{7.4\%}$$

相对危险度=1.78；95%CI：1.03-3.10

为什么要作此类限制？

当有三个及三个以上组别时，通常有一个最高率和一个最低率。如果只是简单地检验这两组之间是否不同，那么统计学检验的结果可能倾向于支持存在差异的结论。但这种方法是错误的。应该首先确定所有组别是否来自同一总体。只有完成了这一步，才可以进行两两比较（所谓双向比较）。只有进行了双向比较才能回答你所研究的问题。

8.2.5 连续性测量值的统计学检验：两组或多组间的比较

服从正态分布（或近似正态分布）测量值的检验。

a) 两组间的比较

(i) 计算均数及其 95% 可信区间（并进行比较/看二者是否有重叠）

例如：在一项评价一种新联合治疗方案的随机对照试验中，要比较试验组病人与传统治疗组病人的年龄：

	试验治疗组	传统治疗组
年龄均数（岁）（标准差）	33.5（17.4）	36.8（16.2）
95%CI（n=120/组）	33.2-33.8	36.5-37.1
95%CI（n=15/组）	31.1-35.9	34.5-39.1

EPI-INFO 注释:

●使用以下命令计算均数:

MEANS 变量 1 名 变量 2 名/N

这里变量 1 指连续性变量, 变量 2 指分组变量。

命令中的“/N”阻止全部数据的打印显示。

本命令并不给出 95%CI。但可利用下列公式计算:

$$95\%CI = \text{均数} \pm 1.96 \times \frac{\text{标准差}}{n-1}$$

(ii) 方差分析 (ANOVA)

Z-分检验 (大样本)

t-检验 (样本量 < 30)

这些检验方法在检验两个组别是否来自同一总体的研究假设时, 检验效果等价于对该二组可信区间的比较。结果以 P 值形式给出。P 值指该两样本来自同一总体的概率 (即比较组别间无差异)。

续上例:

	试验治疗组	传统治疗组	P 值*
年龄均数 (岁) (标准差)	33.5 (17.4)	36.8 (16.2)	
95%CI (n=120/组)	33.2-33.8	36.5-37.1	<0.01
95%CI (n=15/组)	31.1-35.9	34.5-39.1	>0.10

*t 检验或方差分析: 比较组别来自同一总体的概率

b) 三个或三个以上比较组

方差分析

该统计学检验方法是用以检验三个或三个以上组别是否来自同一总体的研究假设的 Z-分检验。上述关于多个率比较的要点同样适用于此处, 即必须首先检查所有组别是否来自同一总体, 然后才能进行两两比较。只有进行了双向比较才可以回答你的研究问题。

不服从正态分布 (即偏态) 测量值的检验

- 不能进行 t 检验, 但可以运用 Mann-Whitney U 检验 (非匹配资料);
- 不能进行方差分析, 但可以运用 Kruskal-Wallis 方差分析 (计算秩和)。

EPI-INFO 注释:

- 要求得出方差分析或 Mann-Whitney 以及 Kruskal Wallis 的结果, 使用同一个命令:

MEANS 变量 1 名 变量 2 名/N

这里变量 1 指连续性变量, 变量 2 指分组变量。

8.2.6 连续性测量值的统计学检验: 两测量值间的关联

服从正态分布 (或近似正态) 测量值的检验

a) 两个不同测量值间的关联 (同一组别内)

i) 以该两测量值为坐标作散点图—观察二者间的联系

EPI-INFO 注释:

- 以两个连续性变量为坐标作散点图, 命令格式为:

SCATTER 变量 1 名 变量 2 名

(变量 1 作为 x 轴, 变量 2 作为 y 轴)

- 如果在变量名后面加了 “/R”, Epi-info 将给出最佳直线

ii) Pearson 相关系数 (“r”)

- 相关系数 “r” 提供关于两个连续性变量间线性联系的信息;
- “r” 既可以是正值, 取值在 0 (零相关) 和 1 (完全相关) 之间, 也可以取负值 (即在 0 和 -1 之间取值);
- “r²”: 指总变异中用该两个变量间的线性联系可以解释的比例 (或粗略地说, 一个变量的方差能被另一个变量解释的部分); 注意, “r²” 介于 0 和 1 之间。

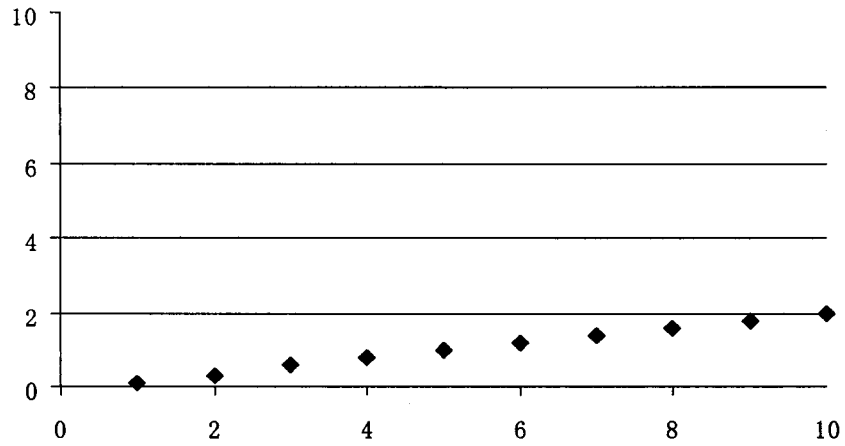
b) 某变量是否能解释或预测另一变量

- 相关系数提示两变量间的联系密切程度, 而线性回归系数则表明 x 变量的一定变化值使 y 变量的相应变化值;
- 数学含义: 变量间最佳直线相关关系的直线斜率;
- 回归系数可以取任何值 (正数或负数);
- 回归系数的取值与两变量的单位有关—可以简单地理解为 x 每变化一个单位, y 的变化量;
- 线性回归系数 (“β”): 指一个测量值解释另一个测量值方差的程度 (不能反过来)。

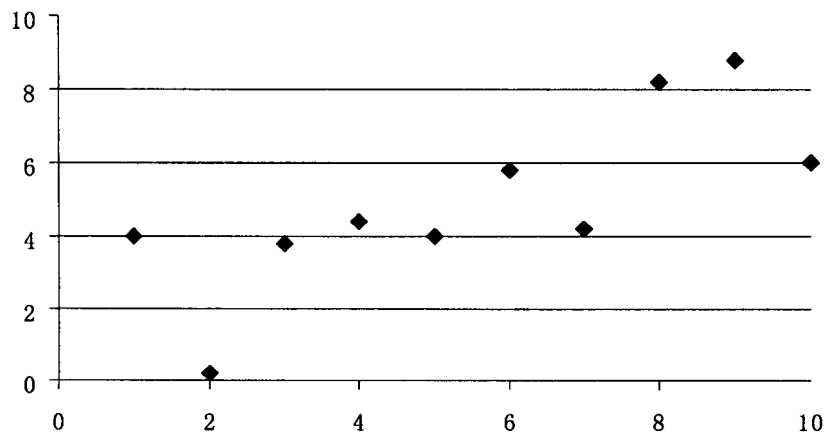
相关系数以及回归系数的 “P 值”

相关系数和回归系数的 P 值只提示该系数等于零的概率, 不提供该系数与零之间差异程度方面的信息。

例如:



其回归系数(斜率)是 0.2, $P < 0.001$; 而下图为:



其回归系数是 1.0, 但 $P > 0.05$ 。

EPI-INFO 注释:

●求解两个连续性变量间的相关系数和回归系数, 命令格式为:

REGRESS 变量 1 名=变量 2 名

这里变量 1 是结局变量或者要解释的变量; 变量 2 是影响因素或者希望用来解释或预测变量 1 变化的变量

●输出结果包括 “r”、“ r^2 ”、 β 值以及 “ r^2 ” 和 β 的 95%CI

8.2.7 在控制其它变量条件下，进行组间比较的统计学检验方法

一个无偏倚的研究设计要求两个研究组除了研究因素外其它方面要相似。研究设计应尽力作到这一点。然而在流行病学研究中，实际上很少能满足此要求。

因此，需要考虑研究组（在基线）的比较结果如何以及如何“控制”这些组间存在的基线差异。

例如：

研究两组工人中气道阻塞患病率，一组暴露于高浓度木尘（n=174），另一组为非暴露的对照组（n=126）。但这两组工人的年龄结构不同。

问题：在本研究中年龄是（气道阻塞和木尘暴露之间联系的）潜在的混杂因素吗？

答案：混杂因素是与健康结局变量（本例中指气道阻塞）以及暴露都有关联的变量。

注意：根据定义，如果某因素是暴露和疾病因果链上的中间环节而与暴露发生联系，则该因素就不能称为混杂因素。

研究获得如下数据：

	年龄≤40	年龄>40	RR (95%CI)
样本量	137	163	/
有气道阻塞*的%	2.9%	16.6%	5.7(2.4, 13.7)
暴露于木尘的%	70.1%	47.9%	1.7 (1.3, 2.3)

*FEV₁<预测值的80%

如上所示，在本研究中年龄与健康结局变量和暴露都有关联，因此它是潜在的混杂因素。

控制混杂的统计学方法

a) 分层

如果只是简单地研究气道阻塞与木尘暴露之间的联系，那么将得到以下结果：

	木尘暴露者	对照组	RR (95%CI)
样本量	174	126	/
有气道阻塞*者的%	12.1%	7.9%	1.5 (0.7, 3.1)

*FEV₁<预测值的80%

结果提示木尘暴露组的气道阻塞症状患病率有轻微的升高，但没有统计学显著性。

然而，将研究对象按照年龄分组后（如≤40；>40），再在同一层内比较与暴露相关的气道阻塞率的话，则得到如下结果：

	木尘暴露组	对照组	RR (95%CI)
样本量	174.8%	126.8%	
年龄≤40 有气道阻塞者的%	4.1%	0	***3.9
年龄>40 有气道阻塞者的%	21.8%	11.8%	1.9 (0.9, 3.8)
Mantel-Haenszel 综合相对危险度	/	/	2.1 (1.1, 4.2)

***logit 近似值

Mantel-Haenszel 综合相对危险度（以及病例-对照研究中的 MH 综合比数比），是一个加权相对危险度，用于检验控制了年龄影响后危险因素（暴露于木尘）与疾病（气道阻塞）间的普遍性联系。

在本例中，我们看到，一旦考虑了比较组间的年龄差异，木尘暴露者发生气道阻塞的可能性几乎是对照组的 2 倍，而且这种差异具有统计学显著性。

EPI-INFO 注释：

●获得分层分析结果的命令格式为：

SET PERCENTS=ON

TABLES 变量 1 名 变量 2 名 变量 3 名

这里变量 1 是影响因素或预测因素，变量 2 是健康结局变量或疾病，变量 3 是潜在的混杂变量或想控制的变量

通过分层分析发现交互作用

绝大多数能够计算 MH 综合相对危险度的统计学软件包，都可同时进行交互作用的显著性检验（Epi-Info 中称之为“比数比或相对危险度的同源性检验”）。

本方法用以检验这种假设，即在其它因素不同取值水平的层内，暴露与疾病间的联系强度（如相对危险度或比数比）不同（即气道阻塞与木尘暴露间的联系在年轻人中和在年长者中是不同的）。交互作用的另一个名称是“效应修饰”。

在上例中，两层的相对危险度不同（即在年轻者中 RR 是 1.9，而在年长者中则为 3.9），但这 2 个比数比的同源性检验结果却无统计学显著性。

因此，在本研究中存在混杂作用（因为很不走运，研究组别的年龄构成不同），但不存在显著性的交互作用（也就是说，一旦控制了年龄差异，就会发现两个年龄组暴露者的气道阻塞率均增加了）。

例如:

下面的示例是从一项以一般人群中的成人为研究对象的病例-对照研究中摘录的, 结果显示在控制了吸烟状态的差异后, 哮喘与工作场所可疑哮喘激发物暴露间的联系。

	哮喘症状	无哮喘症状	OR (95%CI)
样本量	173	285	
吸烟者%	19.6%	27.7%	
工作场所的高危暴露者%:			
吸烟者和非吸烟者合计	24.7%	16.4%	1.7 (1.1, 2.6)
工作场所的高危暴露者%:			
非吸烟者	25.9%	13.6%	2.1 (1.1, 4.2)
吸烟者	17.7%	24.0%	0.7 (0.2, 1.8)
MH 综合 OR			1.7 (1.0, 2.7)

Breslow-day OR 同源检验 (即检验交互作用) $p=0.04$

在本例中, 描述哮喘与职业暴露间联系的“粗”OR (即分层前比数比), 与考虑了吸烟差异后的MH综合性比数比是相同的。这是因为尽管吸烟与哮喘有关联 (即非哮喘患者中的吸烟者比例高于哮喘患者), 但吸烟与暴露却不存在关联 (即吸烟者比例在暴露者中 (27%) 和非暴露者中 (26%) 相似)。因此, 吸烟对于哮喘与职业暴露之间的总体联系没有造成混杂影响。但是, 分层分析揭示, 暴露与哮喘间的联系在吸烟者和非吸烟者中却存在差异, 即吸烟对这种联系具有修饰作用。如果不对资料进行分层, 我们就不能发现这种交互作用 (或效应修饰作用)。

b) 标准化—在上述关于混杂的章节中已经讨论过这个问题。

c) 模型分析—多元回归分析

利用分层分析控制其它变量影响的局限性之一是必须将变量分类化。如果只对一个变量分层是很可行的, 但要同时控制多个变量, 就会发现每层的个体数变得非常之少。

多元回归模型则可以同时分析几个因素 (对结局) 的效应。很重要的一点是, 这些模型可以在控制模型中其它因素的条件下评价某个因素的效应。

多元回归模型可以作哪些分析?

多元回归分析具备简单线性回归分析的功能 (以前讲述过的) —即可以提供这样的信息, 在控制了模型中其它因素的影响后, “x” (回归模型中的重要预报因素) 变量的一定变化值所引起的 “y” 变量 (健康结局测量值) 的变化幅度。

回归模型实际上就是一个如下形式的方程式:

结局变量 (“y”) = 截距 (a) + (系数1, $\beta_1 \times$ 预报变量 x_1)

+ (系数2, $\beta_2 \times$ 预报变量 x_2)

+ (系数3, $\beta_3 \times$ 预报变量 x_3) 等

虽然许多现代统计学分析软件都可以比较容易地完成多元回归模型分析,但决定模型引入哪些变量以及如何解释结果等却不那么简单。不熟悉模型分析的研究者应该寻求那些接受过相关训练、具有相关知识的同事的帮助。

多元回归分析克服了分层分析的一些问题,但也带来一些与试图同时控制许多变量相关的新问题。因为你仍然受到样本量的限制,你得当心,不能引入太多的变量,也不能引入高度相关的变量。

例如,如果要研究结核病的感染危险性,你可能希望考虑社会经济状况、地理地区、家庭居住条件、接触史以及收入水平等的影响。然而事实是,这些预报因素之间存在关联,这将使得多元回归分析的结果难以解释。

■对于连续性结局变量(如 FEV₁)—运用线性回归分析

正如简单线性回归那样,线性回归系数的数学意义是指在控制了所有其它变量后结局与预报变量之间最优直线的斜率。

■对于二分结局变量(如患病:是/否)—运用 logistic 回归分析

从数学意义上说,logistic 回归模型中的回归系数(β)可以用来计算比数比(Odds Ratio),而比数比反映的是健康结局与预报变量增加 1 个单位间的联系,二者间的关系为:

$$OR = e^{\beta}$$

一些统计学软件可以进行此操作,并给出比数比的 95%CI。

EPI-INFO 注释:

●完成多元线性回归分析的命令为:

REGRESS 变量 1 名=变量 2 名 变量 3 名…等

这里变量 1 是结局变量或者是想解释的变量,变量 2、变量 3 等则是“影响因素”或是想用来预报、解释变量 1 变化的因素。

输出结果包括 r^2 (整个模型)、每个预报变量的 β 系数及其 95%CI。

提示: 如果想在回归模型中纳入离散变量,需要对此类变量按照“1”代表“是”,“0”代表“否”的形式进行编码。回归系数(数学意义)指预报变量增加 1 单位,结局变量的增加数量。因此可以将回归系数解释成当预报因素从“是”变成“否”时,“y”变量的变化幅度。

例如:

在上述关于木尘暴露工人的研究中，利用多元回归模型分析 FEV₁（预测值的%）与年龄（岁）、目前的吸烟状态（0=不吸烟，1=吸烟）、哮喘（0=非患者，1=患者）以及木尘暴露（0=非暴露，1=暴露）间的联系，其结果如下：

REGRESS

FEV ₁ （预测值的%）=年龄 吸烟状态 哮喘 木尘暴露			
	系数	95%CI	p
截距	111.6	107.1, 116.1	0.001
年龄（岁）	-0.31	-0.41, -0.22	0.001
目前的吸烟状态	-2.12	-4.27, 0.03	0.05
哮喘	-5.84	-9.09, -2.58	0.005
木尘暴露	-2.98	-5.07, -0.88	0.005

这些结果说明：

- 年龄每增加 1 岁，FEV₁（预测值的%）降低 0.3%；
- （与非吸烟者相比）目前吸烟者的 FEV₁（预测值的%）降低 2.1%；
- （与非哮喘患者相比）哮喘患者的 FEV₁（预测值的%）降低 5.8%；
- （与非暴露者相比）木尘暴露者的 FEV₁（预测值的%）降低 2.9%。

这里的 95%CI 和 P 值是什么意思？

如果结局变量与任何预报变量均无关联的话，那么其回归系数应该为 0（或接近 0），也就是说无论该因素如何变化结局变量都不会发生改变。换言之，如果 95%CI 不包括 0，则说明结局变量与影响因素之间存在显著性关联。因此，在回归分析中 P 值是反映回归系数为 0 的概率指标。

8.2.8 统计学分析中的其它问题

a. 配对与非配对

如果对同一研究对象（或组别）就某变量进行了 2 次测量或者研究对象与比较个体是以个体形式匹配的话，就应该进行匹配分析。

例如：

- 配对 t 检验（替代一般的非配对 t 检验）；
- Wilcoxon 配对分析（替代 Mann-whitney U 检验—用于偏态分布资料）。

b. 异常值

在描述性分析阶段，可能会在数据中发现一些极端值或异常值。这些数据可能是错误所致，但也可能是一个真实值（异常值）。

首先应考虑该数据代表某个体来自另一总体的可能性。

例如：

在一项关于肺功能和工作场所暴露间关联的研究中，遇到一个非吸烟、非哮喘的个体，其肺功能特别低下。后来发现该个体只有一侧肺。

如果该数值是错误所致或代表另外一个人群，则可以从研究中将其删除。然而如无理由删除此个体，就必须警惕这些极端值对某些比较结果产生强大影响的可能性。

c. 多重比较

如果统计学检验结果的 P 值为 0.05，则说明单就机会的影响而言，得到本结果的概率是 5/100（或 1/20）。

因此，如果在某项研究中进行了 20 次比较，单就机会而言，就有很大的机会得到至少一次比较结果的 P 值为 0.05。这就是“多重比较的问题”。

d. “阴性结果”的解释

当研究结果无“统计学显著性”时，应该考虑由于研究样本太小而不足以验证假设的概率。虽然不可能得到肯定的答案，但可以检查可信限的上限（如比数比的可信限上限），看它是否界于临床相关值范围之内。

e. 参数检验与非参数检验

“参数”检验指统计学检验方法假设待分析数据近似服从正态分布，并且数据在所有水平的变异近似相等。在描述性分析阶段需要决定准备采用何种类型的检验方法。

f. 大样本研究结果

对于大样本研究而言，常会发现比较结果或相关分析结果具有“统计学显著性”，即使是从生物学或卫生观点来看意义不大的比较结果。

g. 相关/因果联系

研究中发现某危险因素与某疾病之间存在统计学显著性联系，只能说明两因素间存在联系。在作出此联系为因果关联的结论之前，尚应考虑其它一些证据。这些内容已在本教程的其它地方讨论过了。

8.3 结果的报告

到现在为止研究还不完全，必须把研究结果提交给研究组成员，送给外部专家进行独立评审，并送交发表供同行评议。

8.3.1 研究报告的准备

应该以正式报告的形式将研究结果提交给研究参与者（在下一节进行讨论）。该报告应根据研究设计的轮廓，并利用为该设计准备的材料来完成。内容应包括专业知识背景、研究人群和研究方法的描述、结果陈述及其意义的讨论以及根据这些结果提出的一系列建议。

8.3.2 科学论文的撰写

撰写科学论文无固定格式，但有一些一般性原则需要遵守。科学论文包括题目和以下部分：摘要、引言、材料（研究对象）与方法、结果、讨论以及参考文献。

在开始撰写论文之前，应该确定哪种杂志最适合发表你的研究。然后去查阅该杂志的稿约。你得严格遵守稿约中的要求（偏离这些要求可能导致退稿）。应该仔细阅读该杂志上发表的相同主题的论文，以对杂志录用论文的格式获得较深了解。

题目

论文题目应该简洁并能引起读者的兴趣。将题目写成问题形式有时不失为好主意。例如，已经研究了社会经济状况与结核病治疗失败之间的联系，那么论文题目就可以写成：

“社会经济状况与结核病人治疗失败之间的联系”或者“社会经济状况与结核病的治疗失败有关吗？”

摘要

绝大多数杂志在稿约中告诉作者如何书写摘要以及摘要的字数要求。须仔细遵守。摘要通常是在完成了论文的其余部分后书写的。因此要保证摘要内容与论文内容一致。

前言

在前言的开头可以一般性地介绍论文的主题及其意义，然后集中介绍研究主题。为了使读者了解最新信息，要简短地综述有关该主题的文献资料以及既往研究尚未解决的问题。例如，如果想调查巴基斯坦私立全科医生对结核病管理规定的执行程度，就可以这

样开始你的前言：

“全球每年约有八百万的结核病新发病例。为了保证对这些病人实施合理的治疗和随访，已经建立了结核病人管理的指导方针。关于对这些指导方针执行程度的了解，是开展相关干预项目以保证遵守和执行这些指导方针的先决条件。在 A、B 和 C 国，有 X-Y% 的医生了解这些方针，但在巴基斯坦尚无此类数据。”

然后你就可以叙述研究假设，最后是你的研究目的。叙述的研究目的不要超过两到三个。

前言部分的长度应该是用 12 号字隔行打印的半页纸到一页纸。

材料与方**法**

叙述研究设计以及研究人群的选择。如果是病例对照研究，就需叙述从哪里抽取病例和对照，病例和对照的应答率及失访率。如果是干预研究，也需给出干预组的此类数据。有时将研究设计以图形表达更有助于说明。

规定结局变量（应变量）和影响因素（自变量），进一步叙述如何测量这些变量。本部分的描述应该很详细，以便其它研究者评价或重复试验。对于以前已经公开发表的方法，则只需简短叙述并给出参考文献即可。如果涉及药物的话，就得给出其一般名称、剂量以及使用方法。

如果在研究中应用了调查表，则需对其进行简短描述，包括问题的数目及其主题。接着要介绍调查问题的措辞，包括备选答案。如果应用的是标准化调查表，只要给出其参考文献即可。

应该提供资料的统计学分析方法。描述每项统计学分析所应用的分析方法。如果某检验方法不常见，还得给出其参考文献。交代样本量的估计方法，并按照标准格式给出样本量的估计结果。例如，要检验某干预措施对结核病治疗失败率的作用，可按照如下形式给出研究效力的计算过程：

“如果能将治疗失败率从 10%降低到 5%，就认为具有重要的临床意义。根据 80%的研究效力和 5%的显著性水平，干预组和对照组各需要 431 名研究对象。考虑到预期应答率为 70%，则每组实际需要 616 例研究对象。”

根据赫尔辛基宣言原则，所有涉及人体的研究必须取得伦理学委员会的批准书。在论文的材料与方法部分，必须明确已经获得了批准。

结果

论文中应该只包括重要结果，也就是有助于回答研究问题的结果。一般而言结果应以图表形式给出。文字部分应该用来强调最重要的结果或者图表中最重要的部分。

结果部分应该从简单描述研究样本开始，然后根据单因素分析的结果描述结局变量与影响因素间的联系，最后叙述多因素分析的结果。后者则给出每个自变量与结局变量间的关联性检验结果。

结果部分的文字说明应该简洁；以 12 号字隔行打印不超过 2 页。其中的图表数不要超过 6-8 个。用图通常比用表更容易说明信息。与使用表格相比，使用图的缺陷是不能给出具体数值。

讨论

讨论部分应该从对主要研究结果的简述开始，然后讨论所使用的方法与其它方法相比的优缺点。应对所选方法所获结果的可能含义进行评论。例如，要对结核病人进行检查，须要讨论所使用诊断方法与其它方法相比的优缺点。

接着要讨论研究的有效性，包括外部有效性和内部有效性。然后再讨论研究的主要发现，并解释矛盾之处或意外的结果，并将此次的研究发现与前人进行比较。

为了引起编辑的兴趣，还需强调研究的创新或重要方面。讨论研究发现的意义并指出将来进一步研究的主题是非常明智的。如果已经就其中某个主题开始了研究，那么强调后一方面就更有意义了。

最后，在讨论部分要给出研究结论。研究结论要与引言中提出的研究目标保持一致，这很关键。

论文作者不可能包括所有对研究作出贡献的人。因此在讨论后面的致谢部分应向那些对研究作出了贡献但又未包括在作者名单中的人表示感谢。

参考文献

参考文献应该按照稿约格式进行书写。一定要严格按照稿约的要求列出参考文献。参考文献一般按照其在论文中首次出现的顺序系列编码。绝大多数杂志建议的参考文献数不超过 30 到 40 篇。

完成论文撰写后，应当将其送给研究组外的某个人进行批评指正。稿件回修时，必须仔细考虑评阅者的意见，并对评语进行准确回答。如果论文被退稿，则应该参照评阅者的意见重新书写，然后投到另一家杂志。

9.研究中的其它问题

我们通常首先关注的是研究设计和研究执行中的技术性环节，实际上还应考虑一些其它方面的重要问题。

9.1 该项研究的拥有者是谁？

研究是由许多人共同完成的一项活动，这些人包括研究者、研究组成员、他们所代表的机构、研究资金提供者、研究对象、研究对象所在的社区以及研究完成机构。

9.1.1 构建研究责任

研究是一项合作工程，成立包括所有参与者代表的“指导委员会”就是对这种合作的最理想体现。一般情况下研究者和资助方定期会晤，共同制定研究计划、研究条件以及时间进度表。其它的重要伙伴如社区代表以及负责该社区健康服务的管理机构的代表则很少纳入该委员会。

指导委员会应该从研究的构想开始运行，并且在研究设计过程中扮演积极角色。而且应该在研究完成过程中发挥积极作用。这对于保证研究结果的交流以及根据研究结果提出的政策建议的推行具有重要意义。

9.1.2 尽量满足社区在研究过程中提出的需求

在研究执行过程中，需要考虑研究对象对研究健康相关状态以及影响因素的理解。其中部分可由指导委员会中的社区代表来解决。也许需要召开社区公众大会来解答他们的问题、获得其合作，并向他们解释研究目的和研究程序。社区也许会提出原始研究方案中没有包括的重要问题，但这些问题可以在研究文件中加以说明。有时研究者认为社区对于诸如研究之类的智能型事业不可能提出什么好建议，但许多研究者通过与社区的合作受益非浅。

9.1.3 保证研究结果的追踪

如果研究经过严谨的计划，并且成立了指导委员会，也约定了社区，那么对研究过程的追踪就非常容易了。

研究者必须记住研究的目的是为了创造知识，而知识的目的是为健康活动服务。为了实现此目标，必须通过适当的途径（见下文）将研究结果和卫生活动建议传递给社区以及负责该社区卫生服务的相关人员。

有时研究者可能会发现前所未知的疾病或对于研究对象的健康有重要影响的医学信息。这种情形下必须制定一项良好的随访计划来处理它。切记不要损害研究对象的信任。

9.2 争取有关方面对研究计划的支持

前环境与发展国际委员会主席，现任 WHO 总干事 Brundtland 说过“发展中国家必须建立自己的研究基础。只有这样才有能力在自己具体的条件下建立诊断并完成治愈。国际社会必须帮助他们完成这一过程”。

9.2.1 地方资源

如果有人认为低收入国家的所有资源（人力和物力）都应投入到科学研究上的话，那么毫无疑问是让这些国家竭泽而渔。

符合实际需要又满足成本效益原则的研究常常产生于医疗实践（或者产生于公共卫生活动，即卫生项目）中。研究和实践必须相辅相成。这就是卫生研究发展委员会提议（许多资金捐赠机构已经同意）所有公共卫生项目，其预算的一定比例（2-5%）必须运用于科学研究的理由。这是相关研究最重要的经费支持来源，因为该资金直接支持与社区卫生实践相关的研究项目。

科学研究地方财政支持的第二种来源是中央政府。这种支持常常是通过国家的官方渠道诸如大学、研究机构等来实现的。要想获得此类支持，研究者必须要与上述机构或研究组建立合作关系。对于医生和项目管理者而言，这是一种富有成效的研究完成途径。如果研究获得项目本身以及特别研究基金的双重支持，那么这种根据现场需求建立研究假设，利用流行病学或统计学等专家完成研究的模式将使研究项目硕果累累。

对于科学研究地方财政支持的第三种来源是制药业。如果药厂对研究题目感兴趣，那么研究项目将获得药厂的支持。低收入国家最近进行的关于哮喘分布及其影响因素的许多研究就是靠这种支持方式完成的。

地方基金支持的最后一种途径是发展机构，包括政府性的（通过大使馆）和非政府性的。地方大使的发展部门常常拨出有限（有时也很多）的基金用以支持慈善事业和社区活动。一些服务机构如狮俱乐部和扶轮社也可能对地方水平级的研究提供支持。与此类项目的负责官员建立良好的亲善关系，并且能够说明研究项目与国家的卫生状况具有清晰的实践相关性的话，你将获得资助。

9.2.2 国际资源

支持低收入国家科学研究的国际机构数目非常少，但支持的科学研究数目却非常多。此类机构可能是多边性的（联合国体系和欧盟），也可能是双边性的（发达国家与低收入国家达成的合作协议）。

多边性机构

联合国体系是该领域最活跃的机构。联合国所提供的卫生研究支持主要通过 WHO 来实施。此预算在二十世纪八十年代快速增长，已经超过了 5 千万美元/年。此预算的绝大部分（超过 2/3）运用于疾病的预防和控制。该基金主要由 WHO 的两个机构进行管理，即人类生殖项目和热带病研究项目。后者在疟疾、血吸虫病、丝虫病、锥体虫病、利什曼病和麻风病的研究项目中投入了大量资金。有意思的是，虽然肺疾病是世界上幼儿和青年中最常见的单一死因，但在这一领域投入的支持资金却非常少。WHO 未来的卫生活动优先内容包括烟草及其预防和控制，可以预计在不久的将来，这将是获得经费支持的优先研究领域。

欧盟每年有很大一笔预算用于卫生研究，并且有专门机构负责促进与低收入国家的研究协作。然而此类研究几乎都采取国际合作的形式，并且通常要求有数个欧洲科研中心的协作参与，只有这样才符合资助要求。

双边性机构

传统上只有两家机构受命专门对低收入国家的卫生研究提供支持。这两家机构是 SAREC（这是瑞典的一家政府机构，现在与欧盟协作）和 IDRC（这是加拿大的一家政府机构），其主要权限是支持发展性研究。这家机构曾经支持过大量的卫生研究，但近来对卫生研究资助不力。目前该机构的支持重点是关于卫生服务和卫生体系的研究。

另一家双边协作研究机构也资助研究，尽管这不是其特定权限。日本国际合作机构对低收入国家与日本公民进行的大量合作研究项目给予资助。其它一些机构（如美国和英国）则常常是通过他们自己国家的研究机构予以支持。

基金会

许多国际基金会都提供研究支持经费。此类基金会支持的主要研究核心是传染病和热带疾病（1/3）、流行病学以及卫生政策和管理（1/3）。对低收入国家卫生研究的主要资助者有：

- Aga Khan 基金会，瑞士（初级卫生保健管理），
- Carnegie 公司，美国（人力资源发展），
- Edna McConnell Clark 基金会，美国（热带病研究），
- Ford 基金会，美国（生殖健康），
- Macathur 基金会，美国（热带病、妇女健康），
- Pew Charitable Trusts 基金会，美国（卫生政策和管理）
- Rockefeller 基金会，美国（人群、被忽视的疾病），
- Saskawa Memorial 健康基金会，日本（麻风病），
- Wellcome Trust 基金会，英国（医学和兽医学研究）。

9.3 研究中的伦理问题

医学研究特别是在低收入国家实施的研究，其中的伦理学问题是近年来大家争论的热点。实际上这些问题都可以归结为“赫尔辛基宣言”⁸。国际医学组织理事会在1982年的一份出版物中阐述了与低收入国家的研究相关的问题⁹。

9.3.1 基本原则

涉及人体的生物医学研究

关于涉及利用人体的生物医学研究，其基本指导原则如下：

技术方面

- 研究必须符合科学原则，必须建立在实验室研究和动物实验的基础之上，并且具有可靠的科学文献支持；
- 研究方案必须对研究程序进行详细说明，并经独立的研究伦理委员会审核；
- 研究执行者必须具备科学资格，并且接受某医学主管人员的监督。

可靠性

⁸引用于第18届世界医学大会，1964年，芬兰，赫尔辛基，修改于第19届世界医学大会，1975年，日本，东京。

⁹国际医学组织理事会。人体实验和医学伦理。日内瓦：Bankowski Z, Howard-jones N编辑。1982, 505页。

- 保证研究对象良好状态的责任在于医务人员，而非研究对象本身；
- 研究目的的重要性必须与研究程序固有的风险相称；
- 必须明确告知研究个体和社区，研究可能的风险和可预见的受益；
- 必须保证个体利益高于科学利益也高于社会利益；
- 必须保证研究个体的完整性（包括个人隐私）；
- 只有在可以预见研究风险，而且其受益超过风险的情况下才可以实施研究。

结果的可信度

- 必须保证研究结果报告中的准确性和无偏性；
- 研究者必须随时无保留地公开存在的利益冲突；
- 不符合基本原则的研究结果不应予以公开发布。

知情同意

- 必须向研究对象详细说明研究目的、方法、受益和风险；
- 必须告知研究对象可以随时抵制或退出研究而不受损害；
- 在参与研究前，研究对象必须在充分熟悉研究的基础上签署参与同意书；
- 同意书的签署必须自愿，而且应该交给研究对象与之无依赖关系的人员；
- 对于尚无法律能力签署该同意书的个体，则必须由其法律监护人签署。

研究方案中的伦理学相关程序

- 研究方案中必须包括关于伦理学问题的讨论部分，并且宣布本研究遵守赫尔辛基宣言的基本原则。

临床研究（与医学服务相结合的研究）

临床研究尚需遵守另外一些原则：

- 如果新的诊断或治疗手段提供了挽救生命、健康康复或减轻痛苦的希望，那么这些新手段应该由医生来予以实施；
- 新手段的受益、风险以及可能带来的不适必须优越于现有方法；
- 必须保证所有参与者可以接受最佳的诊断和治疗措施；
- 拒绝参加研究不妨碍卫生服务的利用和质量；
- 如果不要参与者签署参与同意书，必须说明理由并证明其合理性；
- 研究必须承诺对病人/社区具有潜在的诊断或治疗价值。

非临床生物医学研究（不涉及医疗服务）

- 负责医生必须承担保护研究对象生命和健康的责任；
- 如果研究已经引起了损害或者对某些个体具有损害危险性，必须立即中止研究。

9.3.2 在低收入国家中与赞助性研究相关的原则

所有的生物医学研究，无论在什么地方完成，无论受到什么样的资助，都必须遵守上述原则。一些低收入国家完成的研究是由国外机构赞助的，而且常常是由发达国家的捐赠机构支持的。此类研究尚涉及另外一些伦理学问题，需要特别考虑的问题包括：

- 该研究可能服务于国外利益而非国内利益；
- 国外的合作者可能不了解当地的风俗、习惯以及法律体系；
- 由于缺乏对受试者或社区长期发展的责任，导致人们对地方政府的信心幻灭；
- 由于缺乏责任机构而可能使研究对象所受的损害得不到合理补偿。

这些问题就要求研究方对东道主国家及其科学界，就整个研究过程作出详细的承诺。要符合当地优先权，为当地社区提供卫生服务或对当地科学家以及参与者进行培训。必须建立尊重当地风俗、习惯以及法律体制，并保证对在研究过程中所造成的损害给予补偿的机制。研究计划必须详细说明可预见的困难以及解决困难的步骤。

合作研究特别是在低收入国家进行的赞助性研究，会涉及不同的利益，我们必须承认这一点并进行讨论。如果对于研究对象、当地研究者和/或社区没有明确指出的利益，低收入国家的研究者和当局可以拒绝此类研究。此类研究的设计应该明确指出各参与方的受益，并评价合作本身的风险和受益。

伦理学问题的评价程序

- 所有研究必须经过科学性（技术性）评估和伦理学评估合格；
- 此类研究的伦理学原则应遵守“修订版赫尔辛基宣言”（见前述内容）；
- 单独的伦理学评价必须在研究执行社区内完成；
- 伦理学评估标准的严格性至少不能低于调查员来自的社区。

知情同意

- 必须对此类现象予以关注，即研究对象无能力了解研究目的、风险以及受益等信息，因而给不出知情同意书。这种情况下必须提出一种当地伦理委员会能够接受的机制来克服这个困难；
- 特别是当社区成员特别贫穷或不识字时，应该取得社区领袖的研究参与同意书；
- 即使社区领袖对研究计划签署了同意书，研究对象也必须具有退出研究而不受损害的权利。

法律义务

- 研究对象如果由于参与研究而受到损害，则必须予以合理赔偿；
- 研究对象或其监护人不能放弃对于研究损害应给予的合理赔偿；

- 建立损害赔偿机制是研究资助机构、组织或个人的责任；
- 能够保证损害赔偿的机制之一是介入保险计划。

9.3.3 伦理学评估的机构

研究者在将研究计划提交伦理学评估前，首先应满足的第一个要求是保证研究计划的技术性内容符合科学性标准（通过独立的同行评议）。

伦理学评估对于所有利用人体的生物医学研究都非常关键。一般由伦理学评估委员会承担此工作。该委员会的任务是确定方案中的所有干预措施对于人体是安全的，而且保证伦理学所有其它问题也得到了充分的强调。伦理学评估委员会的成员应该包括具有相关技能各类卫生专业人员以及代表社区文化和道德价值的人员。在伦理学评估前，评估小组成员必须发誓与研究方案提交者和研究对象无利益冲突。

如果已经建立了伦理学评估程序，就必须遵守该程序。如果地方伦理学评估程序尚未建立，那么研究者就有责任建立适宜的评估机制以满足伦理学评估的要求。在接受伦理学评估并获得同意之前，任何研究均不得开始。

伦理学评估必须根据研究者提交的材料进行，这些材料包括研究草案以及一份专门用以阐述伦理学问题的表格。研究者必须提供以下材料：

概述

- 明确的研究目的；
- 研究现状综述；
- 开展本研究的必要性论述；

技术性方面

- 明确的研究者资格；
- 所有干预措施包括剂量、干预时间以及已知的对研究对象存在的潜在风险的陈述；
- 包括样本量计算的资料统计分析计划；

研究对象的参与

- 研究对象纳入和剔除标准；
- 获得研究对象知情同意书的手段。

科学杂志不得发表未满足伦理学评估标准的研究论文。