

心血管病调研方法

世界卫生组织

H. Blackburn, G. A. Rose
R. Gillum, R. Prineas 著

北京心肺血管中心流研室

吴兆苏 姚崇华 等译

洪昭光 吴兆苏 审校

2880k 27

一九八五年

2880k 27
吴兆苏
姚崇华
洪昭光

译 者 的 话

世界卫生组织 1982 年第二版“心血管病调研方法”(Cardiovascular Survey Method)是由国际知名心血管病专家编写的一本标准参考书籍,目前在国际上被广泛应用于心血管病防治和老年医学的研究工作。这本书介绍了国际上最先进的研究内容、科研成就和方法学,对国际心血管病防治及老年医学的研究起到了很大的推动作用。

心血管病是严重危害人民健康的疾病。许多国家(包括我国在内)心血管病死亡已占人口总死亡的首位或第二位。我国解放后,随着人民生活水平的提高和医疗卫生事业的发展,人口平均寿命明显延长。因此,老年人的保健,尤其是老年人心血管病的防治已成为我国卫生工作的一个重要方面。多年来,世界各国以及我国各地已广泛开展了心血管病流行病学和人群防治的研究工作并取得了很大成绩。但过去由于研究方法不统一,缺乏统一标准,致使研究工作受到一定限制。各地心血管病和老年医学工作者迫切需要有一本标准的参考书籍(或工作手册)。本书在很大程度上能满足这方面的要求。本书也可作为医学其它研究领域的参考书及作为医学院校师生从事教学研究的参考书籍。鉴于本书对国内心血管病和老年医学研究有一定参考价值,在我中心主任吴英恺教授倡议下我们组织翻译了这本书。

由于译者水平有限,译文难免有缺点和错误,欢迎批评指正。

第二版序言

“心血管病调研方法”第一版于 1968 年出版，它对世界各地流行病学研究方法的标准化作出了重要贡献。它被用于不同研究中制订研究方案和工作手册并用作为各种训练班的教科书。许多从事于心血管病流行病学和预防的学者都熟悉这本书。

现在，在工业化国家中心血管病在早期死亡的原因中居于首位，这些死亡发生于创造力最高并对社会和家庭的作用最大的年令组。发展中国家也面临着类似情况，因为这些情况的发生先于社会经济的发展。但是有办法来预防和控制这些疾病。因此，工业化国家中心脏病的明显流行不是经济发展的不可避免的结果。它能够而且应该通过确定和及时的预防措施来避免，从而加强而不是妨碍经济和技术的发展，对“2000 年人人健康”的目标作出肯定的贡献。

这个挑战在许多国家中正被心脏病学专家和临床工作者同时也为有关公共卫生当局所认识。对预防和流行病学方法的知识需要日益增长，受训的机会正在产生。

考虑到这些发展，世界卫生组织和国际心脏病学家协会和联盟的科学委员会认为本书第二版的出版是及时的。本书的目的是根据近十年中进行的许多流行病学调查得到的经验提供最新的资料和技术。为达到此目的，有必要扩大编委会，除了原有作者外还邀请了另外二位专家担任。世界卫生组织感谢 H. Blackburn 医生，G. A. Rose 医生，R. Gillum 医生和 R. Prineas 医生，感谢他们热心地接受了本书的修订和出版任务。

第一版序言

过去 15 年的流行病学研究对实验和临床——病理研究给予了新的重视。与分析性研究相比，实验性方法——它常导致整体系统和局部功能的研究——着重于疾病发展中多种因素重要性的人群研究。这需要同时测量许多变量，例如性格特征，生理功能和环境因素。这种测量只有通过现代的资料收集、分析和改正的方法才能有效地进行。但是，在流行病学中——广义地讲，在医学中——人们不愿应用这些技术。只是在近几年来，这些技术才被应用于人群研究中。

目前，最大的问题是这些技术应用得不恰当。因此，国际心脏病学家协会研究委员会在 1963 年建议世界卫生组织撰写一本心血管病流行病学应用方法的手册。本书就是这个建议的结果。本书初步涉及到一些技术问题，并试图区别什么是已经了解了的，什么是还需要进一步研究的。

引言

被心血管病夺去的生命是临床学家同时也是公共卫生工作者严重关注的问题。在许多国家心脏和动脉的疾病占男性死亡的三分之一以上。在其它心血管病死亡率还不高的地区,由于传染病死亡率的下降和生活水平的提高会使情况发生改变,因此在将来这个问题会成为世界范围的问题。

在大多数心血管病中,疾病的发生过程是潜在的,只是在其自然史的晚期,临床病症才表现出来。有时病人毫无警觉而突然死亡;在所有心血管病发作中,临床医生很少有机会在心血管系统受到重大损伤以前检查病人。因此,治疗医学不能解决心血管病死亡的一般问题。

另一方面,有充分的理由说明环境因素对这些疾病来说是非常重要的,因此可通过改变环境来预防和减少疾病的发生。这就导致人们对于病因学,识别易感人和亚临床病例以及建立预防措施的兴趣日益增长。所有这些都涉及一种流行病学方法。

死亡率和发病率统计应用在这个新方法中起着重要作用,它能指出慢性病的主要问题所在并为进一步研究提出假设。这种调查统计将继续有用,但它们的结论至多是推测性的;很有必要检验和扩展这些发现,并通过对合适人群的个体检查和随访收集更直接的病因证据。心血管病的问题需要协同作战,流行病学方法,临床和实验室研究各起必要的作用。

这种需要已被广泛地认识到。世界上许多地方的临床学家已把他们的研究从医院中扩大到一般人群中去。这就涉及到很不相同的目的和研究设计。一些技术,虽然是根据临床方法来的,一般也需要作很大的修改以使之适用于大数量人群和统计分析要求。这也导致临床学家,流行病学家和其它医学工作者对交换资料日益增长的需要,看看为什么要进行这些调查。很清楚,如果不同研究组能同意用一些相同的方法,这些研究的价值会得到很大提高,因为这样的话不同人群中的发现可以相比较。本书的目的就是帮助建立这二种需要:即提供心血管病人研究方法的资料和提高不同研究的可比性。

全书的重点是实际问题。但这并不意味着想要进行人群研究的人不需要熟悉流行病学的基本原理;而这些材料在教科书中已有介绍(见 MacMahon & Pugh, 1970; Barker, 1976; Rose & Barker, 1979; Lilienfeld & Lilienfeld, 1980)。这里我们只谈及那些直接与心血管调查有关的设计和计划的原理。参加这类研究的研究人员首先应该清楚地确定他们的目的,然后和一名流行病学家或统计学家讨论他们的计划。

目 录

第二版序言	
第一版序言	
引 言	
第一部分 总原则	(1)
第一章 人群研究：总目的和类型	(1)
第二章 测量的原则	(11)
第三章 计划和准备	(23)
第二部分 检查技术	(33)
第二部分 序	(33)
第四章 调查询问	(34)
第五章 物理检查	(45)
第六章 特殊调查研究	(53)
参考书目	(74)
附 录 1. 人群研究的心电图分类	(80)
附 录 2. 重复性检验的报告	(102)
附 录 3. 结果列表方法	(104)
附 录 4. 吸烟询问	(108)
附 录 5. 病史格式举例	(111)
附 录 6. 伦敦卫生学校心血管病询问	(119)
附 录 7. 呼吸困难询问	(123)
附 录 8. 呼吸询问	(124)
附 录 9. 体格检查记录格式	(125)
附 录 10. 心脏杂音列表	(128)
附 录 11. 专业词汇	(130)

第一部分 总 原 则

第一章 人群研究：总目的和类型

人群研究的必要性

临床研究和流行病学二者在用途和局限性方面有所不同。在医院中能深入地研究病人的特征，但在人群中很少有可能这么做。大多数病因假设来源于临床和实验室观察。但同时从临床观察引出的推理有其局限性，表现在许多重要方面：

(1) 病人通常只是在疾病表现明显时才去就医。此时，病理过程也许已进展到不可逆程度。而且，疾病发生的本身会‘改变’病人。例如，在一项关于心肌梗塞情绪因素的研究中，病人对以前情绪经历的记忆可以明显地和不可预测地发生偏差，这是因为他现在在医院中病得很厉害。企图在这样的病人身上研究血压也许同样会失败，因为这也会因疾病而发生变化。

(2) 由于已决定了病人中某些特征——如风湿性心脏病病人家庭过度拥挤——的频率，临床观察不能说明它是如何与特定人群中的频率不同，但病人正是从该人群中来的。其它住院病人可以作为对照组，但这些对照本身也许不具备有代表性的家庭条件，或者他们是从人群的不同部分被选择性地收入院的。

(3) 心脏病的住院病例由于其疾病类型和因素与社会经济状况有关而易于被高度选择。这样就不能将医院经验（不管是临床的还是病理的）推广到实际上存在于一般人群的疾病中去。例如，在一次人群调查中曾估计到约 75% 心绞痛病例以前不为医生所知，并且其中大部分人没有明显丧失活动能力的表现。因此，医院和人群中的疾病表现可以是非常不同的。

(4) 通常不可能确定一个人群，住院病人来源于这样一个人群。并且这会明显影响到疾病率的计算。只考虑医院中病人的数目会导致错误。例如，常常认会缺血性心脏病发病率在中年后期达到高峰，因为由这种病而住院的人数在随后的年龄组下降。但事实上，发病率在一生中是逐渐上升的。

生命统计和其他统计研究的局限性是不同的。这里，基本人群一般是已知的，疾病率可以计算出来。但信息局限于常规收集的关于死亡和致残疾病的资料。即使如此，它的准确性常常不知道。如果某人想在二个国家中根据死亡率来比较疾病的数量时，这种局限性就特别严重。不仅在每个国家中率的准确度不知道，而且不同准确度既可产生假的死亡率差别，也可使真正的差异被掩盖起来。在考虑某一具体疾病时这种危险最大，而对大的疾病组（例如“所有心血管病”）的率这种危险较小（Reid & Rose, 1964）。

临床和生命统计研究二者对我们了解心血管病的原因和分布都有卓越的贡献。但它们的结论需通过在确定人群中用直接的测量来澄清。此外，某类问题只有通过这个方法才能得到回答（例如对于吸烟和非吸烟者或者不同血清胆固醇水平的人患心脏病危险的估计。）

人群研究的目的

心血管病人群调查几乎和临床研究一样，其目的很不相同，合理的计划取决于对特定目的形成一个清楚的构思。这里我们将讨论一般目的和调查类型，详细的计划以后再谈。

有五个总目的，当然，在同一研究中可以有二个以上的目的相结合。

1. 人群中疾病数量的测量

疾病率的资料可以用来和其他人群的率作比较，我们将重复地强调运用可和其他任何地方得到的结果相比较的普查方法的优点。无论如何，一个很好地开展的调查也应该能够证明它本身是对知识的独立贡献：可比性的想法很少能予先得到保证。在许多研究中，如果研究者只是把它看成是一个有价值的奖励而不是计划的唯一正确方法时，情况也是如此。同时，对于研究之间可比性的考虑也就是那些能改进研究内可比性和质量的要素。

2. 人群中疾病分布的研究

这意味着除了疾病是否存在外人的其他特征也被研究，并且不同阶层的人其疾病的率要分开计算。

这又是“研究内”和“研究间”二者的重要性。例如，在吸烟数量和缺血性心脏病频率之间常有一强的联系（特别是在年轻男性中）。这就是说为得到两组人群的更有意义的比较应该看相似吸烟组的疾病率。一般说即使已知某特征与疾病存在之间有强相关，特征的水平也应该记录。在分析时研究者或者按主要个体类别（年龄、性别、吸烟组等等）表达不同疾病率，或者换一种方法，在考虑了其它变量时计算标化率。另外，一个特异的联系如果在不同的研究组和不同类人群研究中是恒定的，那么这特异联系是真正因果联系的可能性较大。

3. 疾病自然病史的研究

了解疾病自然过程对制订防治措施很重要。这需要有关于疾病在早期是如何被识别的知识，它的发展速度如何，什么样的人进展较快或发生并发症等等。这些资料为冠状动脉外科，先心病或瓣膜病外科，高血压的药物治疗等提供了有根据的基础。

计划研究这种问题的人必须根据普通临床经验小心地定义疾病，因为通常易于根据在高度选择组中的严重疾病的表现来定义疾病。因此，一个35岁男性的血压是140/90 mmHg (19/12 kPa) 可被看作是临床意义很小，因为这不太会被认为是高的血压水平；

但在流行病学中这就有很大意义，因为即使是一个很小的升高（如果是真的）也意味着晚年严重心血管病危险的明显增加。

4. 健康人特征的描述

如同人群研究描写病人特征一样，通过对“健康人”的测量及其在整个人群中的分布也能进行人类生物学的基本描写。

“正常”值可以估计，因为（a）它在通常的范围内（“统计学正常”），（b）它出现在无明显疾病的人中（临床“正常范围”），或（c）它们是“理想”的，因为在随诊研究中它们的长远健康情况最好。只要合适，后二组中的人可以用来和病人作对照比较。

5. 病因假设的形成和检验

这是人群研究中最重要，同时，通常也是最复杂的目的。在某组个体中观察到疾病率的上升可提示有可能病因重要性的个人或环境因素存在。病因假设可以通过进一步研究来验证，这些研究是设计用来看联系是否是恒定的，以及疾病频率增加是否与暴露于有关危险因素的水平成比例。

任何一个特殊范畴的流行病学发展的第一阶段是涉及到描述疾病的分布及形成病因假设；当在一个通过减少暴露于可疑因素来影响疾病的发生的对照研究中假设能充分得到证明时问题就更深入一步。只有通过这种方法，病因才能被最后证实，控制方法的有效性得到显示。

研究类型

人群研究可以是观察性的或实验性的，前者或是横（断）面的或是纵向的（前瞻性的）。

横（断）面研究

横（断）面调查的目的是提供在某一特定时点某一人群的疾病图象。一般只作一次检查。虽然有时可作二次检查，其中间隔很短的时间——例如，最初对所有的人作一次简单检查，随后对那些有可能是不正常的人作一次更详细的检查。

从横面调查中得到的信息有两种类型，或者是（a）与疾病的状况有关，或是（b）描述其它特征。如果疾病已定量地给予定义（如心绞痛”，可以有或无），那么它的频率可估计为患病率：

$$\text{患病率} = \frac{\text{有疾病（按定义）的人数}}{\text{被检查的人数}} \times 10^n$$

如果样本中某些人不能作检查，有时估计一个最小患病率较为适当：

$$\text{最小患病率} = \frac{\text{有疾病（按定义）的人数}}{\text{样本总人数}} \times 10^n$$

这里指出了在医疗服务中会遇到的问题的最小范围。又例如，一组应答率为 80% 的风心病研究估计最小患病率为 4% 可能比一组应答率为 100% 而患病率为 2% 的病人数要多。

如果疾病已定量地表示出来（如血压），其结果最好用频率分布来作初步检查；如果它接近正态，就可以用平均值和标准差（S. D.）来描述。

个体的其它特征可用同样方法描述，这些数据可和有病或无病的人进行对照。或者，如果疾病已进行了定量的描述，则其它特征的数值可根据疾病的不同水平分别予以分类（例如，根据血压水平分类的组其血清胆固醇均值和标准差的分类）。这样就可能对易于发生疾病的人进行描述。

由于通常只进行一次简单检查，横面调查相对经济而迅速。这也许是为什么极大多数慢性病的流行病学研究是这种类型的原因。遗憾的是它与临床研究有同样的缺点（见 1 页）。横面调查的一些局限性如下：

（1）疾病频率的估计是根据时间上一特定点受影响的人数作出的。该数目取决于新病例产生的率（发病率）和已病者离去率之间的平衡（康复率 + 死亡率 + 由于其它原因离去的率——例如在职业人群中由于健康原因所致的退休）。一种死亡率高的病其患病率要比一种有同样发病率但死亡率较低的病的患病率来得低。换句话说，一个患病估计的本身不能肯定疾病真正的频率。

庆幸的是某一疾病的死亡率和康复率在不同人群中常是相似的，在这种情况下应用相应的患病率作为疾病的比较指标是合理的。但重要的是要记住在进行这种对比时所作的假设，因为有时它是不正确的。一个很好的例子是男性和女性心血管病的对比。某些调查指出，中老年人心绞痛的年龄性别专率男性和女性相似，虽然心电图 ST 和 T 改变率实际上是女性高。这个未予料到的发现也许和缺血性心脏病与心肌梗塞的相关性在妇女中比男性中小有关。因此，患病妇女的大部分存活并可在患病调查中被识别出来。故横面调查本身提示妇女患病程度更严重，但这和真实情况相反。

（2）由于大部分慢性病只影响一般人群中的小部分，为了发现确当数量的病例，患病调查一般要检查大量的人。

（3）在职业人群中进行横面调查是流行病调查的一种普通形式。在这里，研究者只研究继续在工作的人，而大部分有病的人因健康原因而过早地放弃了职业（Roberts & Reid, 1954），丢失这些不健康的人会引起对疾病率的低估。

在比较不同人群中做相似工作的人的患病率时必须记住这个选择性丢失因素；因为根据养老金权利，就业标准和是否能做其他工作，有同等残废的人在一个人群中比在另一个人群中容易离职。其他偏差可由于导致选择性进入某一职业组中的因素引起——例如，就职前体检和工作的体力要求。

（4）在某种程度上，每一种诊断方法在每一新的人群中应用时需要进行评价，最有意义的评价常常是通过把测量结果和预后意义相关联后得到的。在一个真正纯横面调查中这是做不到的。

（5）调查工作的一种重要用途是对比病人和健康人。由于某些原因，在一横面调查

中在病人中观察到的特征可以不是那些在疾病形成时期的特征。如同临床研究一样，特征可以因为疾病的结果或治疗而改变（如抗高血压治疗）；或者它可以随着时间的流逝而改变。例如，粥样硬化的主要形成时期似乎在成人的早期：因此，对有缺血性心脏病的中年人的饮食调查只有局限意义，这或是因为时间的流逝，或是因为某些人因为生病的缘故改变了他们的饮食。

另外，那些在发病后很快死亡者不能通过横面调查确定，他们的情况可能和存活者或仅有心绞痛者不同。

最后，缺血性心脏病的发展可不仅简单地取决于某一特定时间的个人习惯，而且取决于习惯的改变。在一些经历了比上一代人的生活水平有所提高的人群中疾病似乎更普遍。因此，研究一些因素，如社会状态的改变，食物，或一段时期的体力活动等是重要的。

可以通过收集回顾性资料来弥补横面调查的不足。在对一组职业人群进行研究时有可能通过分析退休资料来估计选择丢失（Taylor 等，1964）。在任何调查中被查者可能被问及他过去的经历——例如，症状持续时间或过去的吸烟习惯。但必须指出，在重要的研究中这种从个人或家庭中收集的资料常常被证明是不可靠的。他们对一些简单问题回答也许是可靠的（如某人是否吸烟），但对详细的项目则不然（如吸多少烟）。

纵向（前瞻性）研究

前瞻研究在开始时和横面研究相同，但在初次检查以后的某一或某些日子后要对同一人群从新检查。疾病的复查计划可以和初查时基本上一样，但如复查不可能做，随诊资料可以从病休记录（对职业人群），入院资料或生命保险和死亡记录中得到。但遗憾的是这些资料通常只是诊断性的，不完整而且质量未知或不稳定。其结果可能对当地来说意义很大，但常不能用来和其他地方的研究相比较。二种方法（复查和常规病例报告记录的应用）不是相互排斥的，常常可以根据各自的优点结合应用。

有时可“回顾性地”进行实际上是前瞻性的研究。例如某一工厂多年来存有所有雇员的记录。让他们作常规医疗检查：于是雇员名单可用来确定研究人群，检查结果能用于最初医疗情况的分类。需做的事就是追踪这些人并估价他（或她）们的现况。这种类型的研究非常迅速而便宜；但在实际工作中由于初查得到的资料的数量和质量以及追踪初查者的困难而受到限制。

前瞻调查有许多重要优点：

（1）能计算发病率，常定义如下：

$$\text{发病率} = \frac{\text{在一定时间中发生的新病例数}}{\text{暴露人群的数目} \times \text{时间长度}} \times 10^n$$

如果研究组的死亡资料和其它伤亡资料是可用的，就可以建立人群疾病的真正数量的综合图象。

（2）通过对初查时不正常的人进行第二次检查，研究者能描写疾病的自然病史（进展和消退的频度和速度，在何种人中进展较快等等）。

(3) 初查时无疾病表现者可根据某些个体和环境特征分成数类, 在随访期每组发生疾病的危险可计算出来。这种方法有“成倍”的优点, 即初查时不会因明显疾病的存在而导致偏差以及它更接近于疾病的形成期。但有时初查时的发现可因未察觉的疾病存在而发生改变: 例如, 早期心脏病可以改变体力活动或早期癌症可影响营养状况 (Rose & Shipley, 1980)。实际危险率的计算和个人特征预测值的估计只在前瞻性研究中才有可能。

与横面调查相比, 前瞻性研究比较昂贵, 需要一个较大的研究人群, 得到结果所需时间较长。它只在随访丢失很小的相对稳定的人群中才有可能进行, 并且一旦研究开始后要用它来测验其它新的假设是有困难的。它的效率和经济效果可通过把研究人群限制在一组预计发病率高的人群 (例如老年人) 来改进。故前瞻研究仍是一种主要研究方法。无论如何, 由于横面调查的局限性相当严重, 许多研究者愿在他们的人群中进行至少是有限的随访。

前瞻研究的一个实际问题是避免随访时的丢失。如果二次检查时间间隔不太长或二次检查间继续保持某些常规接触的话, 这个问题的严重性可以减少。不适于作随访的人形成导致偏差的组。但是, 如横面调查那样, 有些资料能用来指出偏差的可能程度。

通过第二次会见或检查来估计慢性病发病率的技术有许多特殊的问题。这可以通过对比一急性传染病 (如麻疹) 和一种慢性心血管病的发病研究来说明:

(1) 急性传染病通常有一个相对分界明确的诊断依据。因此在一个学校中有明显临床表现的麻疹发病可通过对儿童观察一段时期并注意谁得了病。但在这种意义上人们不“得”冠心病、高血压或二尖瓣狭窄; 因其过程是一种慢性演变, 在这里某一个人身上可存在有疾病的不同阶段 (从轻到重) 表现。在调查期间, 疾病过程进展到晚期阶段或其它程度等, 情况变得很复杂。下一章将更多地谈及疾病的定义问题; 在现阶段人们可简单地指出, 在慢性病中“发病”是根据某些特殊表现或并发症来定义的。

(2) 麻疹的诊断错误是不常见的, 在每一病例中一旦疾病痊愈, 诊断错误就不易被发现和得到纠正。但在心血管病中, 诊断错误可以更大: 例如, 患者对描述症状的同样问题的回答可不时改变; 血压改变是恒定的; 二个医生看同一份心电图可得到矛盾的结果。因此在复查时, 被检查者情况的改变部分是真的, 部分是由于评定时随机变异的结果。

在一个研究中 (Rose, 1956a), 心绞痛存在与否通过一标准询问表来确定。一年后会见这同一些人, 在相应人群中“新”心绞痛发生率约 1.8%/年。但这个简单的方法忽视了个体对问题回答的变异性。因此, 即使在很短时间间隔后再次会见, 仍会发生一部分人从阴性变成阳性。在前瞻研究中, 那些从阳性改变为阴性者也许在开始时已作为“患病病例”而除外, 而所有从阴性转变为阳性者也许被分类为新的“发病病例”。因此, 发病估计趋向于太高, 在一定程度上取决于诊断方法的随机误差。这种过高估计的情况可以通过多次检查来减少, 因为多次检查可以减少随机变异; 但这样的方法不总是实际可行的。在一短的时间间隔后, 通过复查一样本来测量方法的变异程度是不太麻烦的; 以上这些情况在判断发病情况时要考虑。例如, 在 1000 名开始时心电图正常的

男性中有 50 人 (5%) 在 5 年后的心电图记录是不正常的。但在 3 个月后的样本检查已显示 2% 已经不正常了。这提示在主要研究中 5% 的估计发病率作为真正新病例的量度也许是太夸张了。

(3) 麻疹发病的计算通常根据患病的易感人群部分作出, 即有麻疹病史的儿童在初查时从“易感”人群中除去——一个合理的简捷方法。遗憾的是, 在慢性病研究中相应的方法远不是简捷的。例如, 根据初查时作过心电图检查的前瞻调查比较二组人群的发病率。二组研究者会自然地“新病例”的计算中除去初查时考虑为有病的人; 但一组也许只排除有明显 Q 波的人, 而另一组还排除有 ST 和 T 波改变的人。结果, 即使二组用相同的发病诊断标准, 第二次调查会得到一个明显小于第一次检查的发病估计。

因此, 在慢性病中, “易感”和“不易感”组间无明显分界, 因为每个人对一种疾病的明显表现都是易感的。

作为一种起码的手续, 研究者应该用精确的术语来规定标准, 通过它, 患病病例在初查时得以排除而不影响以后的发病估计。但如有可能, 最好将传染病中的“新病例”的概念完全抛弃而代之以“新表现”。所指的表现应该清楚地加以定义, 它的发病率用下面的比来决定:

$$\frac{\text{发生表现的人数}}{\text{初查时无表现的人数} \times \text{随访的时间}} \times 10^n$$

在一些前瞻研究中, 有已知病史的人在初查前已被排除。初查时发现为“不正常”的人从完善的复查中被排除。这个方法不理想, 它限制了作为横断面研究的初查的价值, 它可妨碍由于后期经验而需对排除标准作修改, 并且它限制了疾病自然病史的研究。可庆的是在低患病率疾病的整个研究期间要求使所有的人 (包括患病病例) 保持不外流只需作少量额外工作。

在一特定时期, 对由于慢性病的新表现而求医的整个人群的观察已进展成为估计发病的多相检查前瞻性研究的替代性研究。(WHO Regional Office for Europe, 1976; Gillum, 1978)。

病例对照研究

病例对照研究为评价病因相关而进行的群组研究或前瞻研究提供了一种更改的方法。群组研究比较暴露于可疑因子和未暴露于可疑因子的人, 看看是否他们的发病情况不同。相反, 病例对照研究比较有病和无病的人, 看看是否他们暴露于可疑因子的情况不同。这种比较假设观察的病例能代表全部病例, 以及对照 (通常是病人) 是从同一人群中选出。一般说这二种假设是很成问题的。可比性可以通过将对照和病例在某些变量 (如年龄) 方面进行配对来改进, 已知这些变量会影响疾病, 也影响暴露机会; 换句话说, 它们的影响在分析中是允许的。如果检查者不能区别病例和对照, 如果询问的方法是很标准的, 那么偏差就不易发生。但是横断面研究固有的偏差是不可避免的并常常不易识别。由于病例对照研究非常简单, 要做得好和判断的好事实上是困难的。有关详细情况不在本书讨论范围之内 (Cole, 1979)。

实验人群研究

观察性研究可以指出特殊因素的存在和疾病发生之间的一个相关关系。如果相关是强而稳定的，就可以推测此关系是因果性的；但还不能加以证明。有时候这种假定的联系可以非常强以至根据它可以直接对公众进行宣教。例如吸烟和肺癌的联系就是如此。更常见的是联系不确定，并需要通过对照性实验来确证（例如通过治疗高血压来预防冠心病）。

控制性的治疗试验有一些实际问题：实验可能要延长，被试者不能很好遵守试验要求，以及由于发病率低，试验规模必须很大等。对第一种困难尚无解决办法。但是有时可以减少研究范围，把它限制在一些高危险人中（例如有疾病家族史或有高血脂或高血压者），甚至局限于那些无症状但心电图有异常的临床前患者。在这种情况下，选择的标准和方法必须小心加以定义，而且相应的研究组和整个人群之间的差别应当加以描述。

从本义上讲，控制性预防试验的计划和前瞻性研究相同，但另外还需一个随机选择的接受试验处理或宣教的样本，并且要采取特别的措施以保证在进行两组比较时不引起偏差。主要步骤如下（详细细则可参阅 MacMahon & Pugh, 1970）：

(1) 对需测试的假设予以精确的陈述。研究者也要确定他们想要检出的受益的程度，或认为生物学上是具有意义的，然后估计试验的规模和类型以保证能对假设作统计检验。

(2) 研究人群的选取。可选用志愿者，但每一参加者必须自愿参加对照组或实验组并遵守试验的规定。实验性研究的对象选取会存在特殊问题，因为它比观察性调查需要受试者给予更多的合作。这就是说由志愿者组成的实验组可无代表性。如果首先在人群的大部分人中进行一个简单的观察性筛选调查，这种选择性偏差的程度可估计出来。那些后来逐步减少参加实验研究的人可和坚持参加者对比。

虽然初次选择偏差的可能的重要性是不容忽视的，但在试验开始后它的严重性常常比选择性丢失要小。在这种情况下，在作为一种接受试验程度的标准用到处理组中去以前进行一个有限的研究来评价合作程度是有帮助的。这种初步研究要求参加者与以后的主要研究的对象相同。

(3) 参加者的初次检查。这应包括评价：(a) 与所研究疾病的有关情况；(b) 其它有关特征，通过这些特征可以识别一些亚组，使处理在这些亚组中更有（或无）效；以及 (c) 其它独立的特征（例如职业和过去史），它们能对处理组和对照组的可比性提供一附加的检查。

(4) 随机地分配到处理组和对照组。初步分层常是理想的，它可保证各组中不同阶层的人具有大致相等的代表性（例如，年龄和以前的疾病；二个变量以上的分层易变得不可控制）。对于大型试验，一般不需要分层，因为机遇差别不易发生而且在分析中是允许存在的。

(5) 检查治疗组和对照组的可比性。这应在初查中指定的每一因素中进行。

(6) 保证坚持遵守处理规定的方案。这应在主要试验开始前在预初研究中建立：这

些规定必须使受试者不感到太昂贵或有严重不适，也不太干扰他们正常的生活方式。当然这也必须是安全的。为使受试者最大程度地遵守规定，需通过访问、打电话或通信等方式与之保持经常的联系。

(7) 检查遵守规定的情况。遵守规定的一个理想而客观的检查方法应包括在每一试验中，并在整个试验中定期重复执行。这种检查可以是生理学的方法（ β -阻滞剂试验中的心率改变，或食物试验中的血脂改变），或是药理学的方法（血或尿中的药物水平）。遗憾的是这种检查常常必须是间接的（如药片计数）。不少人讨论了吸烟试验遵守规定的客观检查方法（35 页）。所有这些测定遵守规定的方法应尽可能和气地进行，因为只有当人们想要作检查时他们才能更好地遵守规约。在宣教效果试验中必须注意其他生活方式或医疗保健的相应改变；例如，某人被劝告吃不同食物时他常常会改变吸烟习惯或将医疗保健作其他用处。

(8) 试验进展的评价。在随访检查中应用的原则和普通前瞻性研究相同（见 5 页），另外还要保证尽可能地在处理组和对照组中用相同的评价标准。这可通过下列几种方法得到：(a) 可用“双盲”技术，受试者和评价人都不知道处理组（这在药物试验中常常可做到）；(b) 死亡可以用作结果的指数；或 (c) 可通过标准的方法（如心电图或 X 线检查）收集资料，然后让不知道何人属于何组的其他人来复查。对于不能用这些方法来处理的资料（如症状）必须十分小心以使评价的方法非常标准。即使如此，这种资料必须认为是有明显偏差的。

另外，在整个试验中要十分注意处理的付作用或并发症。

如果试验不是“双盲”的，对处理的事先了解可影响症状被报告和研究的机。这种偏差不能简单地通过运用标准的诊断标准来去除，但如结果是在每组中在予定的时间内用同样的方法检验的话这种偏差是能避免的。

(9) 分析。首要的原则是根据在试验开始前制订的标准来比较二组的结果，该二组是用随机化方法组成的。事件率的绝对差异应加以说明而不仅仅是相应的改变：医疗政策的决定是根据效果的绝对值而不是相对大小。主要效果的统计学显著性应加以估计（即由于机遇而产生该效果的机率）。所有的检验应是“双侧”的，因为我们不能排除一个干预措施是有害的可能性。多种显著性检验不易判断，一般应避免，特别是根据自身结果得出的假设更需如此。

对那些遵守所接受的处理的人的资料加以另行分析是合适的；但判断必须小心，因为“取出”常常是偏倚的，而且在试验的两边也许是不同的。显著性检验不适用。

报告应该包括主要观察差异的统计学可信限，由此读者可估计导致过高或过低估计效果的机遇程度。这在无统计显著差异时特别重要，因为不恰当的数目可使试验不能检出重要的效果。

在预防性试验中只有一小部分人可期望得到效益。因此，识别任何相反的结果就显得特别重要，诸如干预措施引起的付作用和并发症或是一些由于其他疾病或死因引起的未预见到的效果。

一个良好设计的试验对于该特定研究人群可得出有价值的结论。但是推广到其它人

群或其它干预方法中去会产生判断的困难。

有时将整个社区作为观察单位而进行一项实验人群研究是可行的 (WHO European Collaborative Group, 1980)。虽然讨论这种研究超出了本书的范围, 但这里叙述的许多原理和方法可以说是直接和这样的研究有关的 (Gillum 等, 1980)。

(吴兆苏 译)

(2) 研究人群的选取。应使供试人群具有代表性, 即其年龄、性别、种族、社会经济状况、健康状况等应与所要研究的总人群相似。在选取研究对象时, 应尽量避免选择那些有明显偏倚的人群, 如医院病人、自愿参加者等。在选取研究对象时, 应使研究对象在研究开始时处于健康状态, 无明显的疾病或损伤。

(3) 参加者的初次检查。这应包括评价: (a) 与所研究的疾病或健康状况有关的危险因素; (b) 研究对象的健康状况; (c) 研究对象的社会经济状况; (d) 研究对象的依从性; (e) 研究对象的知情同意。在初次检查时, 应使研究对象了解研究的目的、方法、风险和利益, 并签署知情同意书。

(4) 检查治疗组和对照组的可比性。这应在研究中指定的每一因素来考虑。在选取研究对象时, 应使研究对象在研究开始时处于健康状态, 无明显的疾病或损伤。在选取研究对象时, 应使研究对象在研究开始时处于健康状态, 无明显的疾病或损伤。在选取研究对象时, 应使研究对象在研究开始时处于健康状态, 无明显的疾病或损伤。

(5) 检查治疗组和对照组的可比性。这应在研究中指定的每一因素来考虑。在选取研究对象时, 应使研究对象在研究开始时处于健康状态, 无明显的疾病或损伤。在选取研究对象时, 应使研究对象在研究开始时处于健康状态, 无明显的疾病或损伤。

第二章 测量的原则

必须标准化

人群的研究和临床研究一样，研究者必须描述它的结果以使其他人能对此进行评价以及与其它类似研究进行比较。此外，由于人群研究规模大，通常需要许多调查者的共同工作，只有当他们的调查结果属于同一类时才能将它们合理地综合在一起。

临床上估测观察者的变异程度时常常发现变异相当高。因此需要有标准化的方法。此外，如果研究者想通过和其他人的结果的比较而得到巨大的成果，那么为了科学的交流，对方法学的“自由”就必须加以进一步限制。当然，对于提出的标准化方法会有不同的意见。在这种情况下用标准化的方法或研究者喜用的方法来进行测量都是可以的。标准化最起码要保证基本的可比性——虽然在同时要警惕加入额外的询问或试验而影响标准化过程执行的危险。

标准化不仅有利于各种研究间的比较，而且可以允许综合不同调查的结果。在患病率和发病率低的情况下，特别是当用不止一种方法对个体进行分类时，样本数小就成为一个大问题。因此需要综合几个观察者的资料才能得出结论，这只有十分注意每个阶段的可比性才有可能。这包括：

- (1) 确定要测量什么和测量的标准；
- (2) 建立一种客观的测量技术；
- (3) 确定应用这种技术所需要的条件；
- (4) 对使用这种技术的人员进行训练以减少观察者的变异；
- (5) 对方法的执行情况和各个工作人员进行测验；
- (6) 分析和报告结果，以便与其它类似的研究进行比较。

确定需测量的现象

第一章已概述了慢性病诊断的一些困难。在那里认为考虑疾病的各种现象往往比考虑整个疾病更为恰当。如“什么是缺血性心脏病？”这样广泛的问题已经经许多专家组讨论过，但很难达到一致的意见。在计划一项普查时只需要明确与实际要测量的疾病特征和有关项目；研究者和读者可按照他们认为是最佳的想法去自由地解释结果。

当研究一种尚未精确地了解其特点的疾病时，这种方法尤其重要。例如，在某一个国家第一次作心脏病调查时最好尽可能多地将第四章所描述的各个心血管项目都记录下来。调查结束后再考虑疾病的哪些类型是突出的。

遗憾的是，即使定义一个个体表现也是不容易的。例如心肌梗塞，这是一个病理名

词，但病理学证据在活着的时候不易得到。因此只有间接地依靠一系列的症狀和特征是否存在来识别它，而这些又容易误判。疼痛可能存在且有特异性，也可能是存在但性质可疑，或疼痛不存在。同样地，Q 波可能是大的、临界的或不存在。在另一些情况，可能只存在束枝传导阻滞或一些其它的异常。因为临床医生只要对每个病人尽可能地作出正确的诊断，也许有必要将所有这些症狀认为是心肌梗塞的可能指征。

流行病学和临床医学的不同要求

在流行病学研究中，疾病的定义必须相当简明并适用于大规模研究。被检查者只做有限的检查；确定所需标准的方法必须十分客观和可重复；分析时必须基本上用描述性的词汇（即实际测量了什么）而不是判断性的。

因此，临床实践需要包罗万象的和基本的疾病描述，而流行病学则需要项目的执行定义或专门组合。内科医生的任务是即刻的诊断和采取措施，而在流行病学中时常可以等待分析和随访的结果来揭示所发现的现象的意义。

在流行病学中，开始时最好处理定量的资料，如血压、心率或心脏大小。通常这种资料与疾病内在过程的严重程度的相关性比任意分类的资料要好。如果必须要下诊断，最好在对个体的表现进行描述性分析以后。但有时候，尤其是涉及症狀时，很难避免一开始就下“诊断”。

准确性

一个试验的准确性即是测量想要测量什么的能力。例如，一个准确的测试必须是敏感的：即它能准确地确定真正的病例，不产生假阴性。它也必须是特异的：即它能够准确地确定正常人，不产生假阳性。这可以用一个联列表来表示。

参 考 试 验

		阳 性	阴 性	总 计
筛 选 试 验	阳性	真阳性 正确地确定的 = a	假阳性 = b	总的阳性 = (a + b)
	阴性	假阴性 = c	真阴性 正确地确定的 = d	总的阴性 = (c + d)
	合计	总的真阳性 = a + c	总的真阴性 = (b + d)	总计 = (a + b + c + d)

敏感性（真正的正确地确定的阳性者的比例）表示为 $a/(a + c)$

特异性（真正的正确地确定的阴性者的比例）表示为 $b/(b + d)$