

40.65-6

病例——对照研究 现场主管人手册

刘 韵 源 编写

黎 钧 耀 审定

中国医学科学院肿瘤研究所
流行病学研究室
一九八三年三月

目 录

第一章: 病例—对照研究概要.....1—16

何谓病例—对照研究	2
目的	3
优点	5
局限性	7
病例系列	8
对照系列	10
配对	11
资料收集	14
信息处理和结果解释	15

第二章: 调查前的准备.....17—21

访问对象名单及编号表	18
病例名单及医生意见编码表	18
不宜选用之病例名单及原因编码表	19
需要联系的访问对象名单	19
无需联系的访问对象名单及原因	20
调查过程概述	20
保密性教育与措施	21

第三章: 准备与访问对象联系.....22—27

追踪方法	23
填写《与访问对象联系信息表》	24
着手与访问对象联系	24

第四章: 分派任务和验收结果.....28—40

调查员任务分派表	29
----------	----

联系记录表	30
追踪报告	31
调查失败报告	33
审核调查结果	35
重新分派任务	39
第五章:编辑调查资料	41—46
编辑说明书	42
对错漏信息的处理	44
编辑备忘录	45
第六章:调查资料的核实和补救	47—54
调查资料的质量抽查	48
核实调查的准备工作	49
重新联系与资料核实及错漏补救	49
调查核实报告及有关处理措施	52
第七章:主管人工作日记和报告	
总结	55—60
主管人工作日记	56
工作报告	58
调查总结	59
第八章:资料数量化和变量均匀	
化方法	61—66
调查因素的分类	62
资料的数量化技术	63
变量均匀化方法	65

第一章：病例—对照研究概要

内容提要：

何谓病例 对照研究；

目的；

优点；

局限性；

病例系列；

对照系列；

配对；

资料收集；

信息处理和结果解释。

第一章：病例——对照研究概要

我们主要参考美国Westat公司的《Field Supervisor's Manual》编写了这本现场主管人手册。本手册为病例——对照研究现场主管人提供了所需的基础知识和必要的方法与技术。我们希望经过实践考验和修改充实后，本手册能成为癌流行病学研究现场负责人的有益指南。

何为病例——对照研究

病例——对照研究又称为回顾性研究，乃是寻找一批确诊为某种疾病（例如食管癌），并满足一定研究条件（例如年龄、性别等限制）的病人作为指示病例，又从一般人群中选出未得此病的可比个体作为对照，调查他们对于该疾病各种可疑危险因素的暴露史，以便通过统计分析，确定各因素与疾病间的关联，以及诸因素的交互作用效应，进而评价相应的病因学假说。

病例——对照研究的主要困难表现在两个方面：对照人群的选择，以及对于作为整个分析组成部分之混杂和交互作用效应的处理。本章将扼要介绍病例——对照研究的目的、优缺点，并描述从设计到实施全过程中的某些重要环节。

病例——对照研究通常仅有一病例组和一对照组，在上述情况下，也可包括第二个对照组：

(1) 对于所研究的某种癌症了解甚少，以致很难确定应优先选用何类对照组；(2) 当所希望的对照组存在某些缺陷，

而能用另一合乎要求的对照组予以克服时。但增添第二个对照组会导致一项研究的经费和周期增加约50%，有时可能得不偿失，并且难以断定两种比较结果是否能相互印证。此外，感兴趣的暴露也不限于环境因素，病例——对照设计还适于用来研究基因型和内源性因素。类似，除了可疑的病因因素外，病例——对照研究还能涉及其它因素，其中包括可能的保护性因素。现在许多研究还要求用复杂的生化检验，或其它各类生物学样品的分析数据来补充访问资料。尽管增加了代价，却能改善我们对癌症病因的理解。但须注意，在病例——对照研究中采用生物学样品分析技术，有时可能并无收获，甚或适得其反，因为所发现的生化指标变化，也许只是反映了癌瘤的病理生理学效应，而非病因因素所致。

对于病例——对照研究来说，至关重要的是，全部病例都必须通过组织学方法确诊为所规定的疾病患者，而对照与他们是可比的。可比性意味着不存在偏倚，特别是选择性偏倚和回顾性偏倚。

目的

一项病例——对照研究的主要目标在于，至少为某种假说的因果关联强度提供明确合理的估计。在实践中，这一目标还往往为其它要求所补充，最常见的有：对若干病因学假说的评价；因果关联强度随之而改变的环境条件，即生物学交互作用的描述。这与追踪研究，乃至实验病因学研究的目标并无不同。需强调指出的是，病例——对照研究的首要目的不在于研究结论的普适性，关键是它的真实可信程度。真实性目标要求，病例——对照研究应建立在比较严格定义的病例系

列和与其高度可比的对照基础之上。研究对象的这类限制，很象实验中选用同种动物，以便提高效率，减少可能存在的混杂效应。而普适性目标却要求将一规定人群中出现某种疾病的所有病例视为相同，并取该人群的一随机(或分层随机)样本作为对照。在寻求普适性之前，需记住两个问题。首先，倘若研究对象齐性很差，因为随机变差，由其中任何并组得出的结果都可能不甚明确。此种含糊性将使我们对结论的真实程度缺乏信心，它反过来又排除了结论的普适性。其次，将结论外推，使其具有普遍意义虽是最终所求，但这有赖于相应课题的专门知识，特别是在所研究的人群与欲将结论外推于它的人群之间，对原因的敏感性是否可能不同的正确认识。众所周知，癌症的病因是复杂的，包括多因素的综合作用；而同一因素的致癌、促癌或抑癌作用，系指其在某些确定条件下所致的效应，因而是相对的，条件变了，作用的性质也可能随之而不同。因此，研究某种或某些可疑危险因素在癌瘤病因中的作用时，切莫忽视了当时当地的其它具体条件，以及它们之间的相互影响和综合效应。在复杂的人类条件下，孤立地把某种因素作为癌瘤的单一病因来探讨，或者将某特定地区人群调查结果中的个别认识单独抽取出来，作为普遍结论外推，甚或拿去否定其它地区人群的调查结果，势必引出盲人摸象，各持己“见”的争论不休局面。事实上，在未能从不同人口学环境下的至少两次以上研究中获得类似发现之前，很少有人会将病因学研究的某些结论轻率外推。这两点提醒我们，片面追求结论的普遍意义，往往难以兼顾到真实性；普适性来自于一种，特别是一系列真实的研究

究结果。

需指出的另一件事是，选用一些具有生物学意义和解释的术语，以便满意地表述病例——对照研究结果。在实践中，这意味着为暴露与未暴露人员之疾病频率间的差异，提供一种适宜测度；倘若可能，还应包括剂量——反应关系的描述。此种测度已经找到，就是相对危险度。此外，暴露和未暴露个体发病率或患病率之差也是一项有用指标。只是提供病例与对照的暴露频率及相应显著性检验水平（P值），或者仅仅提供群体比较所得出的相关系数及P值，都是不够的。

优点

与实验研究相比，流行病学的主要优点在于：它的结论直接适用于人类。在癌流行病学研究中，不存在推断研究结论是否适用于人类时所需克服的种属障碍。而与流行病学中的其它研究形式相比，病例——对照研究的主要优点是其所含的信息丰富。一项病例——对照研究能够同时评价多种病因学假说，还允许我们探究各种类型的交互作用——疾病的两种（或多种）病因因素彼此协同或拮抗，因而改变对方效应的强度与方式。此类设计也使我们能估计和控制混杂（confounding），即由于所研究的因素与癌症的某种已知或可疑病因伴随出现而导致的连带（association）关系或关联。丰富信息来自于病例——对照研究所观察的大量病人，而在追踪研究中，仅有少数对象最终会发生癌症，其他对象所能提供的信息甚少。

病例——对照研究还能以另种方式获得大量信息。如果已经从某一基本人群中收集到一系列发病病例，就能描述疾

病在此人群中的流行情况。即是说，人们能够按照年令、性别和某个时点（实际上是在一短暂时期内）的各种危险因素暴露水平来描述发病率。这在追踪研究（前瞻性观察）中是无法实现的，除非在所观察的对象中，定期地增加新的亚队列。原因在于追踪研究人群只在起始时点给出癌症的概括图象。此后，研究组中的对象将发生变化。某些亚组（如年青对象变老）将消失；另一亚组会逐渐脱离研究——失访。但对于基本人群的病例——对照研究，却能为我们提供一个洞察某种癌症全貌的“窗口”。

病例——对照研究的第二个优点是效率高，从它所提供的丰富信息来看，尤其给人以深刻印象。倘若采用已有的资料，数周内便可完成这样一些研究，但通常情况下需一、二年，特别是需要对各研究对象进行访问调查时。此外，这类研究的费用一般较低，因为仅需收集较少对象现有或既往病史资料。病例——对照研究通常包括数百名对象，而在追踪观察中，需有数以千计的个例。

病例——对照研究的第三个优点，是它既适用于稀有病，也可用于常见病。对于癌症研究来说，这种设计的相对优点尤为突出。它的第四个优点，是允许我们对某种稀有暴露的致病意义作出评价。这一优点人们往往估计不足，并常常误认为病例——对照设计不适于用来研究稀有暴露。一种暴露的流行范围对病例——对照研究适用性的影响，主要取决于该种暴露与相应病例间的关联，而非它在备选对照中的一般流行情况。如果某个因素是罕见的，但仍在很大程度上解释了癌症的高发现象，即是说，倘若该因素与占特异危险人群很高百

分比的成员有关,就能采用病例——对照方法对它进行探究。实际上,这是使病例——对照设计之效率达到巅峰的情况,因为此时小规模的研究也能卓有成效。与此相反,某种常见暴露反例可能证明是不适合作病例——对照研究的,如果它所能解释的癌症病人只占很低百分比的话。在此情况下,研究规模需要扩大。

局限性

病例——对照研究的局限性之一,是它只允许估计有关疾病的相对频率,而这要求一定的技能。若是像很多情况下那样,一项病例——对照研究包括了发病率或患病率的调查,则它能为癌症频率提供一些特有危险因素的绝对测度。甚至不包括这些调查时,也能估计各类研究对象中疾病总的绝对频率,并推断特有危险因素的绝对测度。对于癌症研究来说尤其如此,因为关于发病率的资料不难从各癌症登记中心获得。

病例——对照研究的另一局限性反倒是重要和必要的,即这些研究对于偏倚极为敏感,特别是选择性偏倚会降低病例与对照间的可比性。选择偏倚问题是病例——对照研究中最严重的潜在疑难,下面将予以讨论。其它类型的偏倚,尤其是病例——对照的不可比资料引起的偏倚,也可能有严重影响,其中最常见的有回顾性偏倚,根源在于:对所提出的问题,病例往往比对照更能仔细地给予考虑和回答;或者甚至在此之前,病例就一直在思索自己之所以得癌症的原因。因此,病例——对照研究的缺点最终在于,研究者必须求助于调查对象的密切合作,或者仅能半定量地论证,从病例和对照那儿得来的资料在来源和质量上具有等效性。于

是，成功进行一项病例——对照研究的主要问题，在很大程度上取决于能否避免和尽量减少偏倚。

病例系列

当课题已定并决定采用病例——对照方法后，首先要注意的问题通常是指示病例的选择，即患有某种恶性肿瘤且病因学上尽可能同源的一组个体。每次集中探讨一种可能的主要病因较之同时考察多种病因更为容易。为探索癌症的复杂因果关系，我们应尽量利用全部现有的临床表现和流行病学知识，来选定一组合适的病例系列，他们在病因学上是尽可能同源的。多因素的综合作用提示，一种类型的癌症可以有多组独立的病因。为了鉴别可能具有同一组病因因素的病例，选择病例系列时，应该按照年令、性别、种族或某些其它适当因素，作出相应的限制。

对病例特征作出若干必要的限制，除了能获得病因学上同源的病例系列外，还可带来其它好处，其中之一是，所研究的可疑致病因素范围较窄，就更有希望把一些似是而非的“病因”排除在考虑之外，此类虚假病因与癌症并无本质联系，或者仅有由于混杂所引起的连带关系，另一特别重要的得益是能减少选择性偏倚。当所考察的因素与癌症间不存在本质关联，在一项病例——对照研究数据中，仍然存在多种途径可以显示出一定程度的连带关系，（由于类似原因，相反情况也可能出现，即实际上存在本质关联，但在病例——对照研究数据中却看不出此种联系），最重要的途径之一是选择性偏倚能够产生虚假关联。选择性偏倚同时涉及病例和对照，因为这是他们之间的可比性问题。然而，选择性偏倚问题最容易依据病例的

选择情况予以鉴别。所谓选择性偏倚,系指依据病例(或对照)显示出来与待查危险因素暴露史有关的某些特性,而将他们列入一项研究之中(或排除在此研究之外)所导致的偏差。对于病例来说,这些特性往往就是癌症患者的体征或症状,但却未必总是由此癌症本身引起的。选择性偏倚并非孤立的简单现象,它可能反映了某种选择倾向,对病例或对照的人选情况产生作用,或同时不等地影响两者,如此选择研究对象也不是有意识的,实际上,当事人往往根本未觉察到,选择特性与待查因素有关。

病例有三种类型:发病、患病或死亡病例,后者只是偶然采用。发病病例往往被优先选用,也是癌症病例——对照研究中的常用病例类型,他们有患病病例所不具备的一些优点。例如,发病时间更接近于病因暴露时间,所以发病病例比之患病病例能更好回忆自己的有关经历或暴露史,并且与以往的记录案宗相比,新近的病历资料,职业暴露或其它记录更易于得到,所含信息也更为丰富。此外,不同于患病病例,一组发病病例尚未受到各种决定生存之因素的影响,即是说,任何时刻的患病病例,是以往确诊的大批发病病例中的残存者。倘若影响两者的危险因素各异,采用患病病例的研究将导致错误结论。发病病例的第三个优点是,疾病自身的效应表现要弱于病因。如果某个病例是存活多年的患病者,在此期间他可能曾以数种方式变更了其环境条件和生活习惯。因此,除非小心翼翼地详细询问发病前的情况,否则将出现虚假特征。最后,发病病例与病因学研究的通常目标关系更为直接,即是说,通过发病病例,我们能估计某种癌症的发病率,而非患病率与暴露间的

关联方式。

应用患病病例只有一点优于发病病例，那就是他们现成可得。这或许是一个重要优点，因为大多数癌症并不那么常见，以致在若干医疗中心，为收集一组足够数目的发病病例，可能需要数年时间。然而，癌症患者的病死率甚高，要收集大批患病病率也往往谈何容易，除非把很久以前就已确诊的病例也包括在内，但这样一来，又为各种影响生存的因素产生作用敞开了大门。由于这些原因，专门采用患病病例的做法，通常并无明显优点。

对照系列

对照组类型与容量的确定，以及具体对照对象的选择问题，也许是设计一项病例——对照研究时最为重要和困难的任务。选择单个对照对象的方法几乎完全取决于具体研究的细节，这儿不作深入讨论，但一个具有普遍意义的问题是，当采用医院对照系列时，若已知对象现在的身体状况与所研究的暴露有关，例如与待查暴露有关的疾病正是他们最近住院的原因，则还是不选作对照为佳。影响对照组选择的主要因素是指示病例的各种特征和来源，对有关癌症危险因素的知识，以及这些因素与所研究之暴露间可能怎样混杂或交互作用的认识。

病例系列的特征与来源，必然对所选对照类型产生重大影响，倘若要使他们满足可比性，即避免选择性偏倚的话。一般说来，如果已从基本人群中收集到了病例系列，该人群的随机或（年龄和性别）分层随机样本，将证明是一适宜的对照系列。如果病例取自于某个受到限制的来源，则从同一来源中选

择对照往往是合理的。需指出的是,有一种错误概念认为,除了所关心的疾病外,对照与病例必须在各方面都十分相似。现在已经认识到,除了非危险因素外,要求对照在与待查暴露有关的某些方面也和病例相似,即所谓“过份配对(Overmatching)”,会导致不良后果。此问题将在叙述配对时讨论。

对照选择的另一问题是系列的容量,或称为样本大小。当可供利用的病例和对照数很大,且从每一病例和对照收集资料的代价几乎相同时,则病例与对照的选择比可取为1,这就是通常所用的1:1配对。此时问题归结为如何估计可接受的最小研究容量。倘若病例或对照组的容量严重受限,或者为获得有关资料的代价对于两组相差甚远,问题将更为复杂。例如下述情况就屡见不鲜,即仅有少数病例可资利用。在此情况下,选择比应增至2或3,甚至达到每个病例取4例对照,但选择比若超过4,统计功效的增加就很小了。

对照组的来源也是需要认真考虑的问题。大多数研究利用医院病人或普通人群作为对照;加有若干限制的人群,例如病例的近邻,或某些特殊人群,诸如病例的同事及亲属,用得较少。

配对

在设计一项病例——对照研究时,必须作出决定:对照是否需要与病例配对?哪些因素要配对?以及配对的严密程度如何?这些决策将对研究产生重大影响,因此要谨慎从事。配对的主要目的在于,允许利用适当的分析方法,有效地控制配对因素的混杂效应。在比较研究中,除了所考虑的危险因素外,对于病例和对照来说,还有众多的背景因素可能不同,其中某些背景因素会混淆或掩盖危险因素的真实效

应，导致后者的估计失真。同时满足下列条件的背景因素称为混杂因素：（1）在病例和对照组中其取值不同；（2）这些背景因素本身影响比较结果。如果在一项研究的设计或分析阶段忽视了对混杂的控制，将影响结果的真实性和可靠性，因为混杂因素对结果的部分乃至全部效应，可能被误认为是危险因素所致，严重时会导致似是而非的结论，给思维布下陷阱，把人引入歧途。但必须注意避免过份配对，即防止关于下述变量的配对，它们与所研究的暴露相关，却并非独立的危险因素，从而不可能是混杂因素。倘若要求对照与病例在不同变量（包括非混杂变量）上都极好地匹配，则对于所研究的暴露，病例和对照实际上也将无甚区别，结果是适得其反，“弄巧成拙”，耗费了时间和经费，却无所收获，甚至会导致有害的错误结论。区分两类过份配对是有益的，一类出现于下述情况，即配对变量本身属于待查因素致癌机理的组成部分时，这样的配对将使对照与病例在暴露史方面极为相似，即使采用适当分析方法，配对数据也将导致相对危险度估计向 1 偏倚。当取本身并非独立危险因素，也不是因果机理的组成部分，但和所查因素相关之变量进行配对时，则会出现第二类过份配对，在此情况下，采用适当的分析方法，能够给出推理上成立的相对危险度估计，但研究是低效的，即不精确，所得估计难以置信。

某项病例——对照研究需要考虑配对时，应注意下列事宜：（1）仅在已知或怀疑其对感兴趣之关联有混杂效应的因素上，配对才是适宜的；某因素可能与待查暴露“有关”的猜测，不足以证明配对的合理性；（2）若在致癌过程中，

某些因素与待查暴露可能存在交互作用效应，将它们列为配对因素是合理的，因为在对于所怀疑的交互作用因素是齐同的亚组对象中，这能提供分层相对危险度的有效估计；(3) 通常情况下，应能说明，将年龄、性别和一些对其有足够认识之名义变量(如血缘关系，周围环境)列为配对因素时，在财力和时间上的代价是值得的。然而，这样一些名义变量仅当它们满足前两条标准之一时，才应考虑配对，但这不大常见；(4) 当决定了将一个因素列为配对变量后，相应的配对本来应该尽可能严格，但在实践中，往往受经费等条件的限制而难于实现，此时可将原设计要求作适当放宽，以适应实际情况。由于配对的主要目的在于消除配对因素本身的潜在混杂效应，一般说来，为使配对层次中的剩余混杂达到最小，需要较严格的配对。

顺便再提醒几句，偏倚和混杂是两种不同概念，不可混为一谈。病例——对照研究中的偏倚，是由于病例——对照和队列研究间设计思想的不同引起的。队列研究的资料是依据疾病状况确定前的暴露得来的，对于给定时期内出现的全部病例都需进行调查，因此病例和对照的有关暴露资料是可比的，能够据以导出不同亚人群(或队列)中发病率的无偏估计。而在病例——对照研究中，有关暴露的资料是于疾病状况确定之后获得的，病例和对照代表着抽自于总人群的样本。倘若病例与对照的选择程序不同(选择性偏倚)，或者不能按可比方法从这两组人中取得暴露史资料，例如由于对同一调查表中的问题有不同回答(回顾性偏倚)，发病率的估计将是有偏的。可见，偏倚是病例——对照设计方案本身带来

的后果，也就该在设计阶段尽力加以消除。在数据分析时常常很难控制偏倚效应，虽然它们有时类似于混杂效应，并能作相应处理。混杂起因于基本研究人群中诸因素间复杂的因果网键关系，在病例——对照和队列研究中，主要的混杂成分是相同的，通常情况下，不能仅靠适当的科学设计（例如配对）予以完全排除。数据分析的重要一环就是考察各种可能的混杂效应，以及如何来控制它们。有关统计方法这里不再介绍。

资料收集

病例——对照研究资料的收集方法和问题，在很大程度上取决于研究的具体现场，以及资料的来源（走访调查，函询调查，查阅医案或其它类型记录），这里只能提出若干一般性建议供读者参考。一项病例——对照研究，往往以研究人员寻求医疗单位的合作为开端，这相当于从可获得的记录资料中鉴别出适宜的病例，或许还有对照。下步工作是从接受合作的医疗单位提取病例和备选对照的医案卷宗。如果可能，应遵照原医案资料中的顺序摘录有关项目，但若涉及到数个不同医院，或许无法统一格式。如果医案摘要涉及到与待查暴露有关的信息，对象在医例——对照研究中的地位就应对摘录者保密；还应该将医案记录中任何与病例于癌症确诊后，以及对照在同一时间内所受暴露有关的资料，删除掉或掩盖起来，以免摘录者得知。最好在摘抄过程中，至少安排两个人，以便相互监督，密切配合。

对病例和对照的访问最好同时进行，这能使下述可能性降至最小，即通过收集资料所得来的印象将会影响往后的调