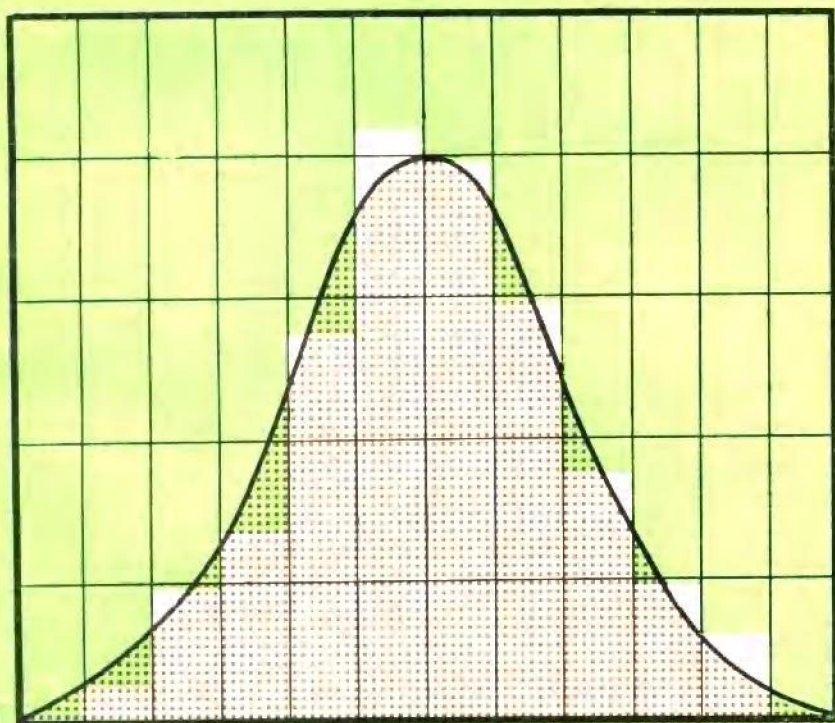


〔英〕奥斯丁·比·希尔 著
施 正 信 译

简明 医学统计学



人民卫生出版社

R195.1

11

3

简明医学统计学

〔英〕 奥斯丁·比·希尔 著

施 正 信 译

田 凤 调 校

(1980年版)

B731/26

人 民 卫 生 出 版 社

B027834

A Short Textbook of Medical Statistics

Sir Austin Bradford Hill

First Published 1977

Reprinted 1978, 1979, 1980

HODDER AND STOUGHTON

简明医学统计学

[英] 奥斯丁·比·希尔 著

施正信 译

人民卫生出版社出版

(北京市崇文区天坛西里10号)

北京印刷二厂印刷

新华书店北京发行所发行

850×1168毫米32开本 $9\frac{7}{8}$ 印张 270千字

1983年7月第1版 1983年7月第1版第1次印刷

印数: 00,001—13,600

统一书号: 14048·4364 定价: 1.20元

[科技新书目46—33]

译者的话

今日的医学卫生——临床医学与预防医学，在评价防治工作的效果方面要求有准确的数据，已成为公认的准则。在推广应用一种新的疗法或预防措施（如疫苗）之前，作出比较严格的、合乎统计学原理和实践的试验尤为必要。当前有些临床和预防医学工作者对基本统计学原理与实践尚缺少必要的知识，甚至有轻视或厌恶统计数字者，以致他们在安排实验研究和解释统计数据时常犯可以避免的错误或得出无效的推论。本书作者是针对这类工作者的需要而撰写这本简明实用的医用统计学教科书的。

本书的特点有：（1）作者用非数学性词语说明、解释了统计学原理，避免或少用数学公式以及方程来阐述计算方法，使不熟悉或害怕数字的医学工作者易于接受和领会；（2）作者在说明问题、解释和分析计算方法时引用的实例大部分摘自医学卫生文献中发表的文章，这对医学卫生工作者特别具有启发性；（3）本书作者用较大篇幅详细阐述了“临床试验”（第二十章）和“常见的谬误和困难”（第二十一，二十二，二十三章）这两个课题。前者对临床医学工作者在研究疗效问题时有特殊的实用价值，而后者强调在解释医学卫生领域内通常遇见的统计数据时要持小心谨慎态度，以免误解而作出错误的推论；（4）作者在介绍计算方法时始终用普通词语逐步进行计算，使读者易于领会和学习算法，对初学者（包括自学者）帮助较大，这增加了本书的实用性。

由于译者水平有限，难免有错，请读者指正。

施正信

编者前言

“书本必须服从科学，不是科学服从书本”

弗郎西斯·培根

奥斯丁爵士的杰作——“医学统计学”——今日成为“大学医学教科书”的一种，给这套丛书增添了光彩。这本简明的统计方法入门现被重新命名为“简明医学统计学”是一九三七年问世的。四十余年来经过作者多次精心修订和再版足以说明此书的优越性和实用性。

这个新版又经奥斯丁爵士作了全部修订并增添了关于医学研究道德问题的两个文件：即世界医学大会赫尔辛基宣言及英国医学研究理事会关于用人体作研究的责任问题的声明。我赞成本书包括这两个文件，为了便于给医学院校和医院各科系和实验室的科研人员参考，这种材料对他们是很有用处的。

即使今日，医学及其有关学科的学生得到的统计方法方面的指导还是很少的。这本小书对初学者是理想的指南，对那些需要复习医学统计学知识者们也是一本极受欢迎的参考书。

塞尔温·泰勒

皇家医学进修学院
伦敦汉默史密斯医院

序 言

我乘本书改制版式的机会将全书作了广泛的审订，进行了许多修改包括文字方面的加工。我希望这些修订是有益的。

我也增加了不少内容，主要的包括：两种自由分布（或非参数）显著性检验（第十二章）；多元回归概念（第十五章）；较为详细地阐述了病人配对法在临床试验中的价值，为了鉴定多种治疗疗效的一种设计的价值以及多种治疗的相互作用（第二十章）。在进行临床试验时我要强调道德问题的重要性，故将两个文件——英国医学研究理事会关于临床医学研究的声明及世界医学大会的赫尔辛基宣言（1975年修订）——附于第二十章之末。本书也包括了英国皇家内科医师学院有关这一问题的一些意见以供参考。

关于寿命表（第十八章）我加述了“队列”（cohort）寿命表及寿命表在寿命死亡情况之外的应用。两个新的“谬误与困难”是有关患偏头痛病人的自选性（第二十一章）和家庭病史采集集中的一些问题（第二十三章）。

我故意地对说明各种统计方法的发展和实用的许多计算法保持原状不变。我当然体会这些计算今日可用电子计算器或计算机进行（在第一章里稍有述及）。但要从机器中得到正确的答复必须先知道如何向机器提出正确的问题。更重要的是如欲正确地应用一种方法，必须理解那个方法的理论基础。为了这些原因我认为了解基本的计算方法是必要的。

书末的四张显著性检验表中的 t 值表和卡方表是经过重新计算的，卡方表较前则相当简化了；新加的其他两张表是用于新介绍的两种自由分布分析方法的。最后，一套崭新的五千个随机抽样数字取代了前版的表……。

不久前我见到一个书评写道：“统计学家总是开始对他们的材料作某些假设，尤其这些材料的准确性和一致性足够应用统计方

法处理者。除非统计学家本人对医学相当熟悉,他们不能为他们自己作出的假设充当胜任的批评家。某些熟悉医学也懂统计学的人们必须参与其事”。我同意临床或预防医学医师必须“参与其事”。其实我写此书的目的就是在促进那个观点。我想趁此鼓励那些“熟悉医学”的人们获得一些统计学观念和基本的统计方法。我希望这本修订和扩大版将继续起那个作用。

奥斯丁·比·希尔

1977年

目 次

编者前言	[1]
序言	[2]
第一章 统计方法之目的	1
第二章 统计资料的收集: 抽样	12
第三章 统计资料的收集: 选择性	21
第四章 统计资料的收集: 记录和调查的表格	32
第五章 统计资料的陈述	41
第六章 平均数	56
第七章 观察值的变异度	68
第八章 标准差的计算	79
第九章 抽样问题: 平均数	88
第十章 抽样问题: 比例数	101
第十一章 抽样问题: 比例数的差值	110
第十二章 抽样问题: 平均数的差值	119
第十三章 抽样问题: 卡方 (χ^2)	129
第十四章 再论卡方 (χ^2) 并举例	136
第十五章 相关与回归	150
第十六章 相关系数的计算	160
第十七章 标准化死亡率和有关指标	167
第十八章 寿命表与治疗后的生存率	183
第十九章 疾病统计指标	197
第二十章 临床试验	203
附件: 临床研究	229
甲、用人体作研究的责任	229
乙、赫尔辛基宣言	233
第二十一章 常见的谬误和困难	236
第二十二章 常见的谬误和困难续论 (一)	247

第二十三章 常见的谬误和困难续论 (二)	257
第二十四章 统计学的证据与推论	266
附录 1: 统计名词定义	278
附录 2: 随机抽样数字	284
附录 3: t 值表	296
附录 4: 卡方 (χ^2) 表	298
附录 5: 威尔库克森氏检验表	300
威尔库克森氏符号秩次检验为配对观察之用	300
威尔库克森氏秩和检验为两组独立性样本之用	301
索引	303

第一章 统计方法之目的

有人认为应用数值方法于医学课题上是一种烦琐和浪费时间的作法，而另一些人则宣称它是医术发展过程中的一个重要阶段。如果在1921年，一个医学杂志撰写统计论文的作者提出那样的问题，那是合乎情理的。但在今日就不能再提这样的问题了，因为在过后的岁月里接受统计方法的应用，有很明显的增加，不管医学界的读者当时对它作何反应，在医学杂志上发表的论文中用统计方法陈述和分析的文章还是日益大量地增多。不仅英国注册统计署（和其他国家类似的机关）每年出版的并加以分析的全国生命统计材料提高了统计知识及运用，而且医务工作者应用数值方法分析他们自己观察的临床试验或实验材料，也日益增多。但是许多这样的工作者和读者都没有或很少有统计方法的训练，因而他们中间许多人觉得职业统计师用的比较数理性强的方法“难解甚至讨厌”。常有这样的论调：数理方法运用于这些材料是不合理的，因为这些原始材料不够正确故不能承担要求的义务。这种说法严格说来是不合逻辑的。如果收集的数字资料值得用统计方法分析的话，那么显然它值得用最好的统计方法分析——即可从资料中得出最大信息量的方法。

数理统计方法用在特殊的事例是否是最好的方法，它们是否有必要还是被认为是不必要的详情细节？这可归纳到这样一个问题：我们能不能用比较简单的数值方法求得满意的结果？换言之，我们能不能用这样的简单方法满意地分析资料、检验假设并且得出结论？回答无疑是肯定的。医学论文中许多数据是可以用比较简单的统计方法得出有价值的结论的。有时所得的成果比用更为渊博的但作者尚未掌握的方法可能要少些，但也不应因此而忽视它，因为即使最简单的统计分析法只要逻辑地和细致地运用，也有助于清晰地了解原材料的意义和局限性。如果同意上述的论调就发生这样一个问题：解释数字用的简单方法仅是常识的同义

词呢？抑或还包含可以传授的一种技术和知识呢？熟悉医学统计者必然得出这样的结论：仅靠常识是不够的。直到受过极为基本的统计原理和方法的训练之后，很多人不能用常识来处理与解释数字资料。聪明的人常犯那些被人指出时看来是十分幼稚可笑的错误，而同样的或同一类的错误再三不断地继续重演。经常缺乏的是一种叫做“统计的机智”，这种机智比单纯、良好的辨别力要高明得多。除了少数人无疑生来就有这种机智外，大多数人务必通过学习统计基本原理与统计方法才能获得。

本书的目的是简单地讨论这些基本原理并且从医学文献里挑选出典型的范例，用以指出如何这些原理常被忘记或忽视。无疑本书内有些论述和讨论似乎是在太简单而有些被指出的错误有人认为是枉费心机的，不值得关心的。可是这种错误的屡屡出现和这些基本原理的常被忽视之多，我们就不能那样看待它。对这些情况每个专业统计师从发表的论文里或送给他审核的文稿里是熟悉的。

统计学定义

一个实验室工作者常可将他不感兴趣的变异因素排除在外，而集中注意力于一个或几个受控制的因素进行试验，一个临床或预防医学工作者通常不能这样做。他务必运用遭受他控制不了的因素的影响的记录，而对这些因素他必须加以考虑。在这种情况下统计方法实质上是为了阐明这种复杂因素的作用。所以我们可以将这样的统计定义为“受多种原因显著影响的数字资料”而统计方法为“专为解释受多种原因影响的数字资料的方法”（尤尔与肯达尔：统计原理入门）。统计学更广义地可谓“是一门关于处理由个体组成各组数字资料的学科”（阿米太奇：医学研究用的统计方法）。

例如，假定有这样一个人群，他们都与一例传染性肝炎病人接触过，我们对其中一部分人进行了丙种球蛋白预防注射，对其余的人作“对照”观察。我们要知道此注射能否预防肝炎的发生。接触患者之后，发病机会的可能性受到许多因素的影响而

异，如年龄、性别、社会经济等级、接触时间的长短和接触的疏密程度等。分析统计时要尽可能对所有上述的因素加以注意。我们必须使要作比较的组别中的各个可能有影响的方面相等，除预防注射一个因素之外。若在实验开始时不能使组别相等，那么必须竭力用分析方式来使它们相等。所以，我们应当尽一切可能在对观察值进行设计时排除或考虑到这些外来的或能干扰的因素。有这样设计我们或许不仅能够测定预防注射是否有效并且弄清在哪一种情景下比另一种情景下更为有效，或对某一年龄比对另一年龄更为有效等等。在工作结束时依靠统计方法来消除干扰因素是一个严重的错误。统计方法不能补偿在设计不良情况下得来的观察资料或一个设计拙劣的实验。但是统计知识可以大大地有助于一个实验的设计。

实验的设计与解释

对上述的那种实验统计学家可以建议采用某种统计方法进行工作。包含相当多因素的复杂的实验也可同时用统计方法考虑在设计之内（见考克伦与考克斯著的“实验设计”及考克斯著的‘实验的计划’两书）。本书不拟讨论较为复杂的设计计划和分析方法而主要限于阐述医学上时常需要的比较简单的实验安排。限制于讨论这一类实验并不一定意味着它在某一特定情况下是最好类型的实验。

前面已着重指出：一个简单实验的基本问题在于事先保证尽可能使处理（治疗）组与对照组在一切有关的方面完全相同。强调有关的理由有二：第一，显然没有一个统计学家对特殊的医学问题能够知道一切有关的因素。从一般的经验中他很可能提示某些主要的有干扰性的因素，并在计划实验时应对它加以考虑，如在上述例子中的年龄与性别；但对一个特殊问题的狭隘的和具体的有关因素未必能期望他会熟悉了解。在这种情况下只有负责实验工作者本人能够知道哪些可能是有关的因素，因为他总该通晓他那个狭隘的作业领域。因此，当要求统计学家提供建议时他有义务提出在设计实验或分析效果时可能遇到的干扰因素的处理方法，

但他不能确定哪些是有关的干扰因素。同时，没有一个明智的统计学家应对他很不熟悉的医学问题提供意见。圆满的协作要求统计学家尽其所能学习当前研究的问题，实验工作者（临床医师、公共卫生医师等）应尽量学习关于统计学的方法。两家都没有相当多的知识就等于盲人领导瞎马。

第二点必须遵守的是要求各组在一切有关因素方面相同或相等，而且在解释结果时必须小心谨慎。例如，倘若我们看到甲组与乙组在某一个特征上有所不同，比方说感染率不同，我们能否肯定说这是因为甲组接受过预防注射（以此为例）而乙组则没有。我们能否肯定甲组与乙组除预防注射外其他有关的特点相同？例如，接受注射的人通常来说可能属于较高的社会（经济）阶层，因而他们居住的环境里传染病比较少。我们永远也不能肯定说我们没有忽视某些有关因素，或说不能预见的和不能测定的某些因素不存在。这就是因为统计学家知道所涉及的因素往往很多而复杂，他有时看来似乎是一个过分谨慎的和怀疑一切的人。

临床医学中的统计学

治疗疾病的实验的本质在于比较。海尔姆霍尔兹（Helmholtz）的名言：“一切科学是测量”，此外还应加上那个伟大实验家亨利·戴尔爵士（Sir Henry Dale）所指出的那更深一层的话：“一切正确的测量基本上是比较性的”。另一方面，还有人说这样的话：人的变异性太大，以致不能容许药物对照实验所要求的那种比较。但若每个病人是“独特”的话，那就很难领会如何依靠以往观察其他病人的经验作为现在治疗的任何根据，而临床医师必须根据他们以前的经验——尽管是两、三个或一百个病例——而选择治疗方法。

虽然，广义地说，人们对某种治疗的反应不是特异的，但毫无疑问，他们很可能是有变异，有时甚至有很大的变异。所以观察两、三个病人很可能由于机遇的关系在一个医生手里得到良好的疗效但在另一个医生手里则不然。结果医学杂志变为登载互相矛盾的报告的刊物，但每一个报告的本身可能是那个医生真正看

到的事实，虽然这是不够作为概括的依据。

我们非但不能说统计方法不可能解决人的变异性问题，我们必须承认：因为变异性的存在，才有必要应用统计方法。也不能象另外一个经常听见的论调说，统计总是要求巨大的数字。这样的要求是可能的，但要看什么问题。可是单独一个病人的治疗效果的反应显然是一件事实，只要观察是真实地进行和确切地记录下来。其实那个单独的病例在某种情况下可能提供极其重要的证据。

今举例来说明上述问题。如果我们用一种新药治疗一名已被证实为急性白血病的患者，结果病人获得完全的、即刻的和无可置疑的复原，我们就得到一个极其重要的和有深远意义的结果。我们接受仅仅一个病人的疗效只作为是一桩不寻常的事件——不一定是因果关系——这是因为经过长久和广泛的经验表明病人对急性白血病的反应是不变异的，即尽管用现代药物可使白血病患者病情暂时减轻，他们终究得不到完全的、即刻的和无可置疑的复原。因此，虽然以一个病人的疗效作为从特殊推论到一般的依据是极不明智的，不承认这个事实的证据是十足狂妄的。

另一方面，如果对一个患急性风湿热的病人给一种药物而患者获得完全的、即刻的和无可置疑的复原效果，我们就有很少根据可作评论。那个效果很可能因为服药所致但毫无因果关系的可能性，因为患此病者的反应是有变异性的。不管你给他们什么治疗，有的病人可能死亡，有的经长期疾患但终究恢复原状，或有或无永久性的损伤，有的可获得即刻的和无可置疑的复原效果。所以上述药物的服用和病人的复原可能一点因果关系也没有。我们需要观察更多病例才能适当地作出因果关系的推论。我们需要统计方法和一个设计好的实验（详见第二十章）。

所以虽然在很多情况下我们的确需要比较大的数据以便正确地测定一种情况，这不一定说，象有些人说的那样统计学家常因数据少而拒绝接受医学上某些首创的和基本的观察资料。且举一个具体例子作证。最初描述的骨脆症只有两例，后来的作者说统计学家因此认为是无用的证据。但是为什么这样呢？如果这两例

病症的描述和图解说明是精确的，那么它们当然成为全部科学知识的一部分。它们是不可否认的事物的证据。能够发生的和确实存在的东西不管其多少，不管什么原因或关系都可以从即使是一个例子的样本加以观察。只是关于要求从特殊概括到一般时，统计学家或任何有训练的科学工作者同样要加以反对。如果根据见到的两个病例一个在屠宰场附近的开业医师说这种病是与屠夫特别有关，那么人们要提出意见说这样的经验在数目上和地区上太局限了，所以没有足够的理由作出这样概括的结论。

简而言之。对临床医师和统计师来说没有也不可能有一个幻数。我们需要一个或一百个或一千个例子的观察资料全靠我们研究的问题和要求得出的推论是什么。

还应该清楚地看到：有时几乎完全没有统计数字或确实没有精确的测量的情况下，只要很粗略地比较一下一种治疗（或一种其他行动）与另一种治疗的效果，就可以得出真实虽然不一定是精确的答案，如果治疗的效果是实在的而且是显著的。例如，青霉素的成效之大与过去的经验形成鲜明的对比，因而虽无严格的对照试验它的优点不可能不显露出来。在这样的“冠军”面前那些对统计学的理解较为缓慢的评论家易于在事后自作聪明地批评说：在一个长久而厌烦的试验开始时已存在的一般证据足使这个试验是不必要或多余的。他们忘记了有多少次是通过试验才把自夸为有效的疗法证明为无效，尽管当时已存在很多一般的证据。如果没有一个试验，这样的情况很可能长期拖延下去遗害病人。再者要从一般的印象肯定某种药物是完全无用的，或有少量的但是确实有益的疗效的，例如将复发率从6%减少到3%，那是很困难的。要从毫无对照和序时的观察下肯定某一种有高效的药物比另一种药物在特定情况下有更高的疗效，那就更加困难了。只有经过仔细设计的临床试验才有可能解决这个问题。但这决不是说用统计学指导的实验是唯一有益的临床研究方法，或肯定是最好的增进知识的途径。这不过是一种方法而已。

病人和病情的变异性引起的一个困难是病人分类问题，即在治疗之前和治疗之后把同类（样）的病人确实放在同一组内。除

非这个问题在具体情形下（特别是精神病方面）能得到解决，这样的基本问题如“治疗有何价值？有多少价值？对哪一类病人有价值？”就不能获得明确的回答。即使治疗没有一般的价值，但它对个别孤立的病人显然有效，这个证据还须从统计中得来，即证明这样罕见的复原在同样但未给予这种治疗的病人中不那么常见。早晚要根据那样的证据才能证实其效果，但若没有一个有计划的实验，证据要来得晚些。普拉特勋爵（Lord Platt）强调指出：“在一般的情况下，临床研究用的病历记录除非作为有计划的实验的一部分有目的地保持下来多半是使人失望的……临床实验不必意味着将无价值的或无裨益的程序加诸于病人。它欲包括计划一条行动路线，记录其观察的结果使它们经得起批评性的分析，并给一个临床问题作出答复。它是一种心理上的态度”。

再说分类问题：统计的过程必将资料的个别项目压缩和归纳到少数小组之中，并用平均数及其他数值来简单地描写这个资料，这样做我们不免牺牲了一些原有的详细材料。因此我们务必特别谨慎避免牺牲那些与问题有关的东西，不要删掉比为了清晰及易于处理、解释和表达资料所必要的简化更多的材料。实际上要把大量得到特别治疗的病人的详细病历及同样多的对照组病人的资料刊登出来是很少做得到的，即使做得到的话，单独那些资料不能为读者或作者提供他所需要的知识，直等到它被合适地加以压缩。问题在于：特殊治疗是否有效？即要求阐明其因果关系。欲达此目的必需从大量原来记载的资料中摘出材料制成简短的表格，算出统计数值，以示两组病人——特殊治疗组与传统治疗组——在治疗之前和之后的相对地位。

在临床医学领域里统计学的应用还有其他范例可引如后：
(1) 疾病自然史的观察包括调查患者初诊时的症状和体征，它的过程，病人之间症状是否因某特征如年龄性别而变异，他们是否有联系？等等；(2) 病人于特种治疗后，尤其在手术后的随访和疗效的评价；(3) “正常”的定义——人体某一特征达到怎样程度才算是病理？(4) 临床常用的实验室检验如血球计数的准确性研究。

简言之临床医学领域里许多基本问题例如诊断、治疗和愈后都要用统计学来解答。我们需要仔细地收集统计资料，用适当的方法分析它，把它发表于文献中。最初时，需要对病人作认真确切的观察，这样才能逐步建立医学知识的宝库。

公共卫生领域的统计

在公共卫生工作中我们有时也同样进行有计划的实验和实验结果的分析，例如试验疫苗对防制传染病的效能（如百日咳疫苗）。或观察给予特种补充饮食的孕妇对母儿健康的影响。通常我们还遇到一些不是从主动的实验中得来的统计数字而是从一个人人口日常生活中发生的事件的统计数字，例如人口在规定期间内的总死亡率，特殊年龄组的死亡率——婴儿死亡率，儿童死亡率，壮年死亡率，老年死亡率等；特种病症的死亡率如呼吸系结核病死亡率，癌症死亡率，暴卒死亡率等。至于法定传染病则政府常定期公布其发病率，等等。

这些统计数字及其所用的统计方法的目的有二：一方面它可以作为评价公共卫生情况的简单数据，用它来比较一个地方与另一个地方，或一段期间和另一段期间的例如人口死亡率的高低。只有根据这样的证据我们才能有效地考虑预防医学面临的问题以及在何时何地最需要采取防治措施。例如，当我们看到婴儿死亡率，即不满一岁婴儿的死亡频率，在社会阶层方面从专业人群到一般劳动者有很大的增高时，我们至少就知道这个问题的存在并需加以注意，不管它的原因何在。又如我们发现一个地区收到的伤寒报告比另一个地区多得多，虽然其原因不明确但问题已被指出了。或者有记录显示住在城区拥挤地点的儿童死于传染病者比住在不大拥挤地点者要多。我们能够排除这些过于拥挤地区的不幸遭遇吗？

这种尽可能准确地和完整地收集及编制的统计最初是用来对人口的卫生或不卫生状况引起注意。没有这种数据我们就难以知道在哪些最重要的方面应采取行动，所以收集和编制生命统计资料是公共卫生工作的基本业务。换言之，医生日常工作中要填报