

应用统计学的基本概念（一）

临床研究中统计学的作用是什么？我们所做的就是区分事实和偶然性。我们需要比较组间差异, 并检验干预的效应。

那么，临床研究中生物统计学家的作用是什么？虽然我们希望能得到令人满意数据并知道如何设计试验及进行分析数据，但通常我们需要依靠生物统计学家的专业知识来选择适当的试验设计和计算适合的样本量。我们都应认识到样本量是由对结果的测定决定的，涉及其精确度、准确度、可重复性和可行性。此外，在进行数据分析时，生物统计学家帮助我们决定使用何种分析工具。在试验开始之前就应确定所使用的统计方法，非常重要的是，分析方法决不能在试验完成之后加以改变。

在对试验进行分析时，统计学的作用是什么？我们应当记住，统计方法仅仅是一种帮助我们解释试验中所获得的数据的工具。它们是一种工具而不是试验的最终结果。而且像任何工具一样，使用统计工具必须小心。计算机可以产生一些或有统计学意义的数据，但是只有研究者才知道该使用何种统计学检验来进行统计学分析。已参加培训的 researcher 可以

很容易地选择统计学检验方法，必须记住的很重要的一点是，对于没有足够知识的人而言，有强大功能的统计软件包可能导致致命性的错误。

生物统计学的重要概念之一是其正确性(validity)。对于关键性的数据分析、试验的结果尤其是结果的发表，正确性都是其核心。有两种正确性：内部的和外部的（可推广性）。

内部的正确性就是在设定的试验范围内结果是准确的，使用的方法和分析经受得住检验，数据和相关的医学文献均支持研究者对试验结果的解释和结论。

外部正确性或可推广性决定了试验设计是否能够允许所做的观察和所得的结论推广到整个人群。试验人群的选择决定了最大可推广范围，这个概念我们在这个讲座的其它部分已经谈到过。如果研究对象包括男性、女性、不同的种族、不同的年龄分层，那么就有更多的机会将临床试验的结果应用于普通人群。另一方面，受试者的选择也将决定研究和结论可应用的人群范围。例如，如果在临床试验中选择年龄介于 5-10 岁的儿童，那么该试验的结果就仅能应用于该人群。如果选择 45 岁以上的亚洲男性作为受试者，那么试验结果

就只能应用于这个人群。

在正确性的概念中，应该认识到须有足够的样本量以支持所得出的结论，同时要选择适当的对照人群，特别是强调随机双盲对照这一临床研究的根本的科学方法。著名科学家 Isaac Asimov 引用过这样一句话“科学工作的可敬之处就在于，任何科学信仰，虽然已具有坚实的基础，仍要不断地被检验看它是否真实，是否普遍地正确”。这是他对正确性之重要性的看法，泛指任何科学范畴也包括临床试验。

现在我们将要讨论如何看待一些类型的数据。首先是相对危险度和比值比。这是评价后果的指标，当比较暴露因素对结果的影响时是非常有价值的。比值比主要用于病例对照研究。相对危险度主要用于队列研究。这两类研究设计在前面的讲座中已讲过。

让我们首先看一下相对危险度。这张表显示是如何得到一个相对危险因素的。表格被分为两行两列，第一列是发病，第二列是未发病。我们看一下发病是否是暴露于危险因素的结果或者未暴露于该危险因素。暴露组发病数被标为 A，暴露组未发病数被标为 B，非暴露组的发病数被标为 C，非暴露组未发病数被标为 D。从这张表格中我们可以得到相对危

险度，相对危险度是暴露组的发病率除以非暴露组的发病率，即 $(A/(A+B)) / (C/(C+D))$ 。这就是相对危险度。

用来计算比值比的表格结构与上表相似，但是计算方法不同。仍具有这样的自变量，暴露于危险因素或未暴露于危险因素与发病或未发病比较。即 A 和 B，与 C 和 D。但是比值比与相对危险度不同，它是由 $(A \times D)$ ，即病例组有暴露史 $\times$ 对照组无暴露史除以 $(B \times C)$ ，即病例组无暴露史 $\times$ 对照组有暴露史。比值比即 $(A \times D) / (B \times C)$ 。

在解释关联性检验时，我们如何使用比值比和相对危险度？实际上非常简单。当比值比或相对危险度小于 1 时，这种危险因素与疾病呈负相关或该因素是保护因素。比值比和相对危险度等 1 时二者无关联性，如果大于 1 时，二者均证明为正相关。

以上介绍了临床试验中生物统计学应用中的一些概念。下面我们将举一些例子，看看在分析临床研究数据时如何使用以上及其它的检验方法。

