

所有队列研究都能计算RR值么？

作者：李楠 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/8263.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

队列研究是我们常用的一类研究方案，在研究中暴露在前、结局在后，未发生某一结局(未患病)的研究对象因为是否存在暴露(某潜在危险因素)被自然分成两组，经过一段时间的随访观察结局(是否患病)出现的情况。由于队列研究这种从“因”到“果”的特点，其在建立因果关系、并合理估计暴露因素的效应过程中至关重要。

队列研究中，最常见的模式是这样的：

	D	D	Total
E	a	b	a+b
E	c	d	c+d
Total	a+c	b+d	N

D	患病组	P	暴露组患病概率
D	非患病组	1-P	暴露组非患病概率
E	暴露组	$\bar{P}$	非暴露组患病概率
E	非暴露组	1- $\bar{P}$	非暴露组非患病概率

RR: 暴露人群与非暴露人群患病概率之比

$$RR = \frac{P}{\bar{P}} = \frac{a/a+b}{c/c+d}$$

在良好的队列研究中，我们能够直接得到暴露与分暴露人群最终的患病率，因此在队列研究中，我们常使用RR值来估计暴露因素对患病风险增加的实际效应。说起来，其实使用RR值还有个非常重要的前提条件，就是我们能准确计算暴露组的患病率P和非暴露组的患病率。那么问题来了，队列研究都能计算出这两个值，进而得到RR值么?对此不同的专家都曾在论著中发表过意见，让我们一起来看看。(由于意见不一致，在此我们不一一指出出处了，真感兴趣的话，可以私下讨论~)

观点1：回顾性队列研究不能计算RR值?

有专家在著作中提出，是否为“回顾性队列”是判断能否使用RR值评价暴露效应的条件。的确，在前瞻性队列研究中，我们能更完善的记录患者的信息，研究对象的观察起点、终点也很明确，得到准确的P和并不困难。但是在回顾性队列中，则可能出现一些情况导致我们对两个概率的估计出现偏差。但是也并非所有的回顾性队列都无法准确计算这两个概率，比如基于某医院的孕妇产检队列完成的研究，各项信息和最终生产结局记录完整，当然可以准确计算两个概率和RR值。

观点2：基于医院患者建立的队列不能计算RR值?

基于社区人群的队列，其人群代表性比较可靠。而基于医院的队列并不能完整的代表社区人群，毕竟其中非健康状态人群比例较大。此时用医院队列得到的某危险因素RR值来估计该因素的实际效应，就有可能存在错误估计。但是如果我们的外推仅仅是对医院的患者呢?当我们仅外推到医院这个大人群时，并不存在无法计算RR值的问题。

观点3：当潜在偏倚存在的可能性较大时，队列研究的RR值可能存在偏差!

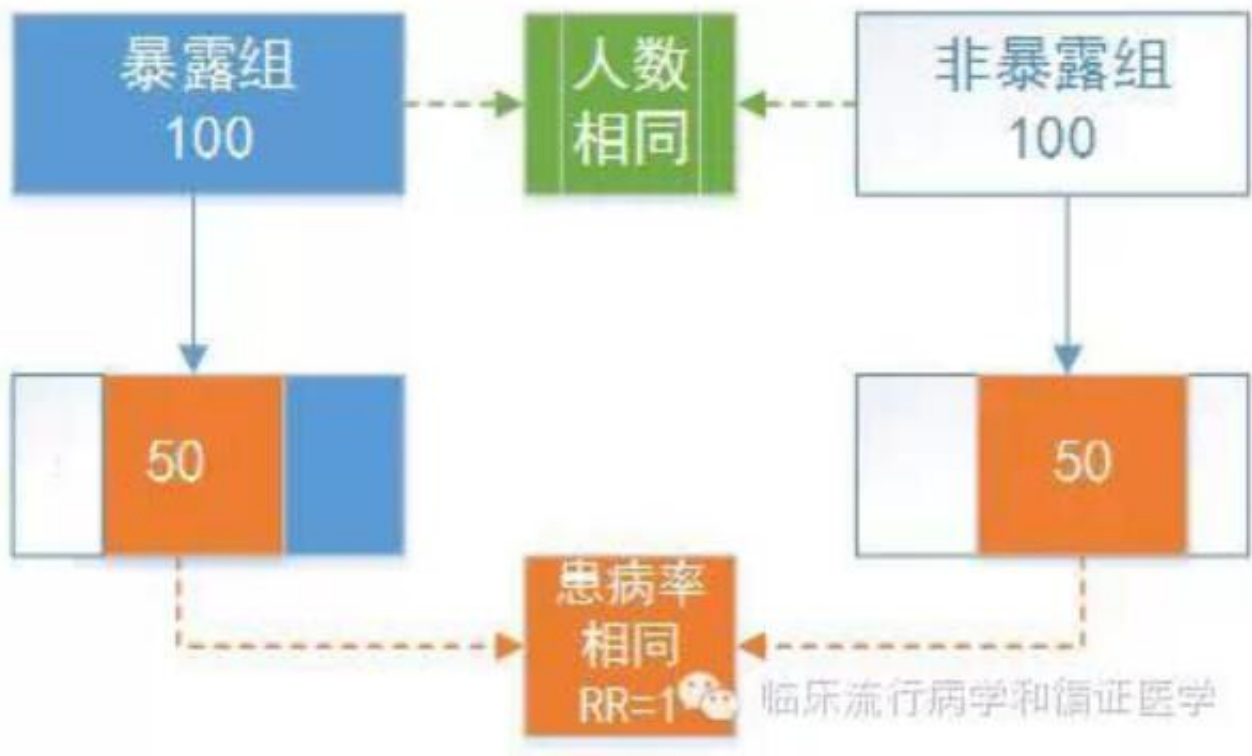
其实从上两个观点各有道理，却又都有局限性。但是他们背后的规律都是潜在偏倚导致了对效应的错误估计。基于回顾性队列计算RR值时，其RR值可能受到信息偏倚(比如失访不均)的影响而发生偏离;基于医院人群的队列计算RR值时，其RR值可能受到选择偏倚的影响而发生偏离。总之，偏倚本身的效应就是导致我们对危险因素效应的估计存在偏差，当然也会影响到RR值本身。

RR是如何受到影响的呢?

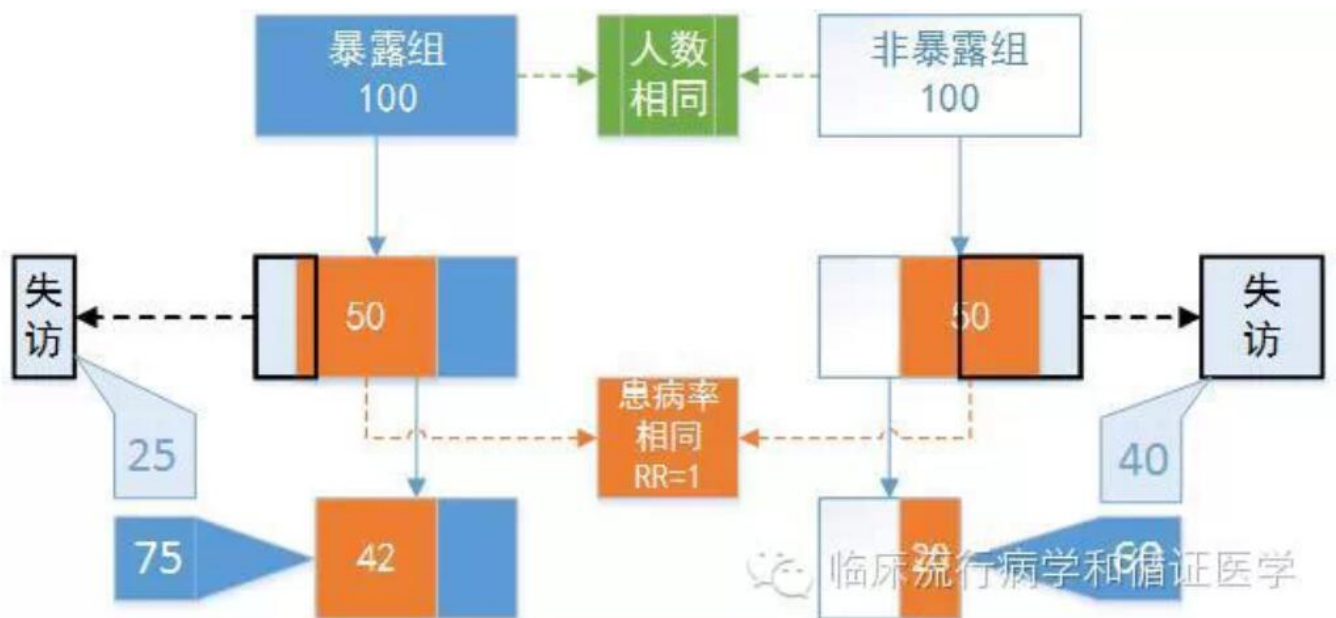
我们不妨来个糖炒栗子。假设我们有一个队列研究，研究开始选取没有发生结局的暴露组100人，非暴露组100人。



理想情况下，患者不会出现失访，随访1年后两组患者的患病率相同，均为50%。此时的真实RR值应该等于1。



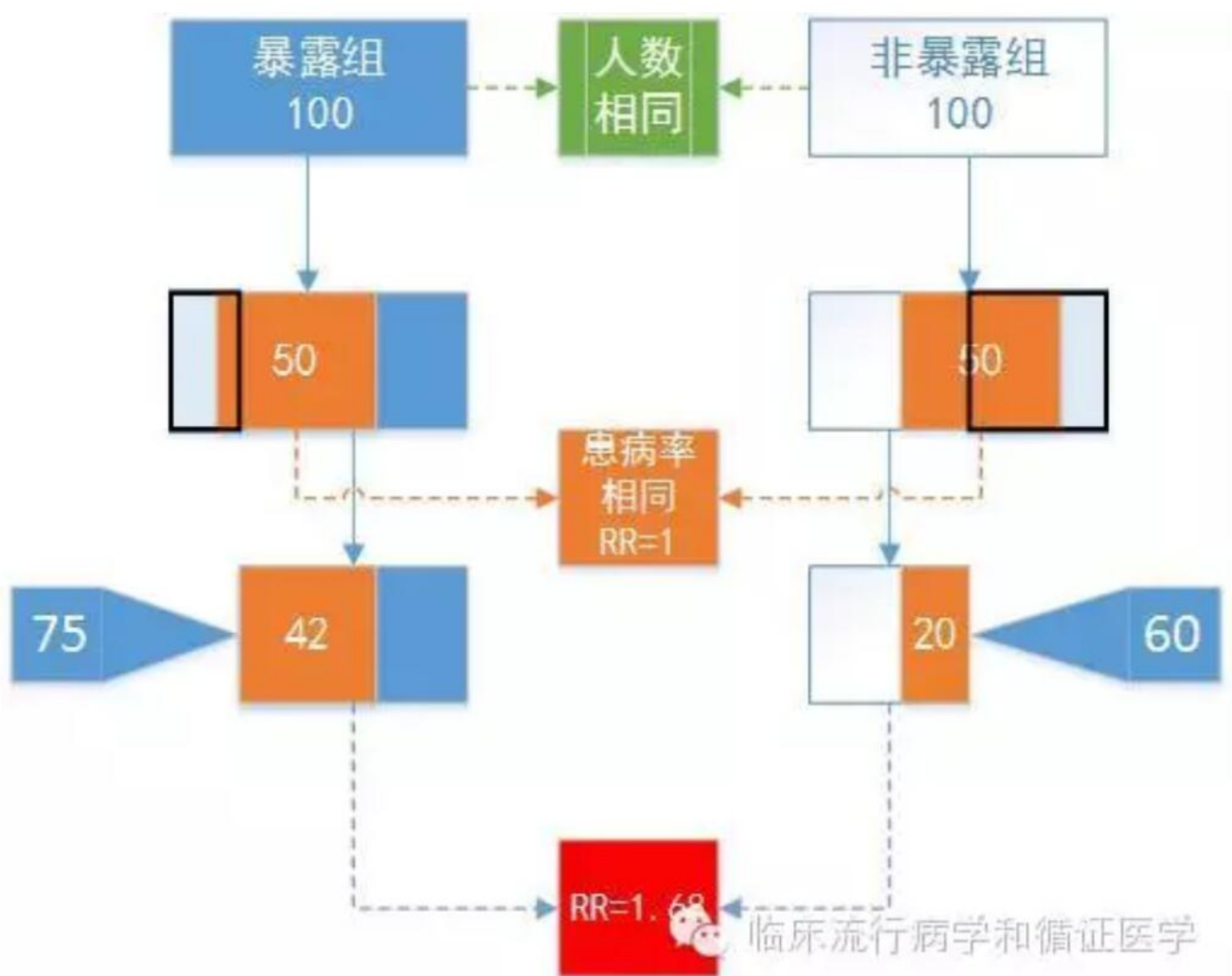
实际上失访是不可避免的，随访一段时间后一些患者陆续失访了，两组仍能随访到的人数分别为75人(暴露组)和60人(非暴露组)。



同时，暴露组与非暴露组中，失访对象中患病率并不相同。比如观察新药A和传统药物B的不良反应，因为A刚上市，因而医生和患者都更关注其不良反应，因此一旦出现不良反应的倾向都会及时随访并记录;失访的患者通常没有什么大问题。而老药B上市很多年了，即便出现了不良反应患者和医生也都习以为常了，甚至患者出现不良反应后认为医生水平不高，反而更倾向于失访。就形成了这样的尴尬局面。



此时估计的RR值已经不再是真实的1了，到底偏了多远呢?让我们来算算看。



可见，潜在的偏倚会直接影响效应的估计。因此在计算RR值之前，我们首先还是要估计一下偏倚的种类、方向和水平，从而为我们估计真实效应奠定基础。当然RR值还是可以算的，但是RR值是否是真实的RR值，还要看P和P-的估计是否准确。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发