

---

# 风险因素的识别和结局的预测大不同

作者：曾琳 赵一鸣 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/1467.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！*

临床研究中我们常常探究的是某种疾病(结局)发生有哪些可能的影响因素，对风险因素进行控制从而避免或减少疾病和不良结局的发生;此外，通过对风险因素的认识，个体发生疾病(结局)的可能性进行预估，达到风险评估的目标。以上都是我们临床研究中常见的研究目标，细细读下来，其实这里面有两类看起来很相像，但实际大不同的目的，就是风险因素的识别和结局发生的预测。

这两者看起来十分相像主要有以下两个原因

：第一，  
风险因素  
的识别和结局发生  
的预测都涉及潜在的风险因素和研究  
结局的测量和数据收集  
，在病因学研究中，就是潜在的危险因素和疾病的发生;在预后研究中，就是预后相关因素和疾病的转归(预后)。第二，  
无  
论是  
风险因素  
的识别还是结局发  
生的预测都需要在风险因素和研究结  
局之间构建统计学关联  
，并进行因果推断。所以很多研究者容易把这两者混为一谈，认为在一个研究中，只有通过统计模型的构建，就既可以识别结局的风险因素，同时可以进行结局发生可能性的预测。但实际上这两者不能简单的认为是同一个工作，应该分别进行研究。

对于风险因素的识别的研究来说，着重于试图发现新的可能的风险因素并判断其与结局发生之间关联的方向。说人话就是：对于某个临床结局，经过学术界的探索可能有某些已知的风险因素了，这时我们的研究目标就是看能不能通过自己的研究，来探索发现新的风险因素;并判断这个或这些因素的存在(及其水平的提高)会增加临床结局的发生还是会减少临床结局的发生。在这种情况下，我们往往在设计阶段就会采取匹配的方法，针对出现结局者的特征选择已知风险因素和出现结局的人相似的对照。比如心梗的病因学研究，已知血压水平是心梗发生的独立风险因素，此时，如果我想探索血脂水平是否心梗发生的独立风险因素时，我会根据收集到的心梗患者的特征寻找血压水平和病例组一致的对照来开展研究。

接着上面这个例子，如果是希望对心梗发生进行风险预测的话，如果还采用上述的研究设计方案

(匹配血压), 那么得到的预测模型必然会缺少血压这项指标, 这显然是不可接受的。也就是说我们在试图构建结局的风险预测模型时, 不应采用上述的病例-对照设计(尤其是匹配的病例-对照设计), 而应采用队列设计, 在一个有同质性的人群中构建模型探索各风险因素对结局发生的贡献, 同时进行合理预测。

另外, 对于风险因素的识别和结局的预测两类目标来说, 还有一个显而易见的差别。我们来看一个Logistic回归的结果:

方程中的变量

	B	标准误差	瓦尔德	自由度	显著性	Exp(B)	EXP(B) 的 95% 置信区间	
							下限	上限
步骤 1 <sup>a</sup>			8.721	3	.033			
Age category								
Age category(1)	-.154	.240	.410	1	.522	.857	.535	1.373
Age category(2)	.251	.243	1.062	1	.303	1.285	.798	2.071
Age category(3)	.609	.301	4.095	1	.043	1.839	1.019	3.318
Gender	.271	.173	2.453	1	.117	1.311	.934	1.839
Blood pressure	.188	.142	1.738	1	.187	1.206	.913	1.595
Atrial fibrillation	.350	.262	1.779	1	.182	1.419	.848	2.373
Smoker	.678	.203	11.142	1	.001	1.970	1.323	2.934
Cholesterol	.228	.178	1.645	1	.200	1.257	.886	1.782
History of angina	-.045	.212	.045	1	.832	.956	.630	1.449
History of myocardial infarction	.088	.218	.163	1	.687	1.092	.712	1.674
常量	-2.307	.288	64.146	1	.000	.100		

这个Logistic回归的结果是我们在研究中常常会得到的结局(脑卒中院内死亡)和可能的风险因素(年龄段、性别、血压、房颤、吸烟、血脂、心绞痛、心梗)之间统计关联的结果。如果我们在进行风险因素的识别时, 我们的这个表中关注的重点会是“显著性”和“EXP(B)”以及“EXP(B)的95%置信区间”, 这三项足以帮助我们来识别某个因素是否结局发生的独立影响因素了。而如果我们要做结局发生可能性的预测呢?这是我们的关注焦点会变化, 变成上面表中的“B”和“常量”, 其中, “B”是回归系数是风险预测计算中每个影响因素的权重;而“常量”则是在风险因素识别的时候不需要, 但在进行结局发生可能性预测时必不可少的。“常量”其实就是截距, 指示的是这个人群的基础风险;也就是当各风险因素都不存在时, 结局发生的可能性高低。显然在进行风险预测时, 我们必须把基础风险考虑在内。

其实风险因素识别和结局发生的预测还有其它的不同之处, 但由于篇幅所限, 今天就不展开讨论了, 如果大家对这部分内容感兴趣, 我们可以在以后的推送中继续聊这个话题。通过今天的文章阅读希望大家明白, 虽然风险因素识别和结局发生的预测存在很多相似点, 但是两者要解决的问题和着重关注的方向是不同的, 在研究设计的时候需要对研究有合适的定位, 在撰写论文的时候需要区分自己研究到底能达成哪个目标, 这两者千万不要混为一谈。

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](http://iikx.com)转发