
SPSS：Kaplan-Meier法中的假设检验方法

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/6288.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！

SPSS：Kaplan-Meier法中的假设检验方法

。临床研究的预后研究中，我们常常根据可能影响预后的因素进行分组，比较不同组间预后事件发生的差异来推论研究关心的因素是否是和预后有关的因素。预后常常不仅仅和是否发生预后事件有关，还与事件发生的时间有关。比如死亡最为预后，那么我们不仅仅关心不同组间死亡事件发生的频率，

同时也关心死亡事件发生的时间。因为，死亡是必然会发生的，但是**生存期**很重要。

在我们做统计分析时，这类预后研究我们往往会采用生存分析的模型进行假设检验。一般单变量生存分析常常会用**Kaplan-Meier法**

来比较组间差异。可是当我们要做组间比较时，会发现统计软件中会提供Log Rank检验、Breslow检验和Tarone-Ware检验这三种组间比较的方法;其中Log Rank检验在研究报告中最为常见。那么我们到底应该用什么方法来进行组间比较呢?我们今天来聊聊这三种假设检验方法的差别。

总的来说，这三种假设检验的方法都和属于卡方检验的方法，都需要计算各观察时间的实际死亡数和预计死亡数，并套用卡方统计量计算的公式。其计算所得统计量同样符合自由度=组数-1的卡方分布。但不同的是，每种方法的统计量具体算法不一样。Kaplan-Meier法会根据观察时点(每个病例对应随访时间)顺序，把生存资料从小到大排列来进行分析，根据时间顺序计算实际死亡数和预计死亡数。

LogRank检验

各时点的权重均为“1”。就是不考虑各观察时点开始时存活的人数对统计模型的影响。也就是每个时点死亡情况的变化对整个模型的贡献是一样的。

Breslow检验则在Log Rank检验的基础上增加了权重，并设置权重为各时点开始时存活的人数。也就是开始存活人数多的时点死亡情况的变化对整个模型的贡献较大，而开始存活人数少的时点死亡情况的变化对整个模型的贡献较小。

Tarone-Ware检验

是权重的取值方法介于以上两种方法之间，设置权重为各时点开始时存活的人数的平方根。同样是开始存活人数多的时点死亡情况的变化对整个模型的贡献较大，而开始存活人数少的时点死亡情况的变化对整个模型的贡献较小。只是开始存活人数多的时点对整个模型的贡献不如Breslow检验大。

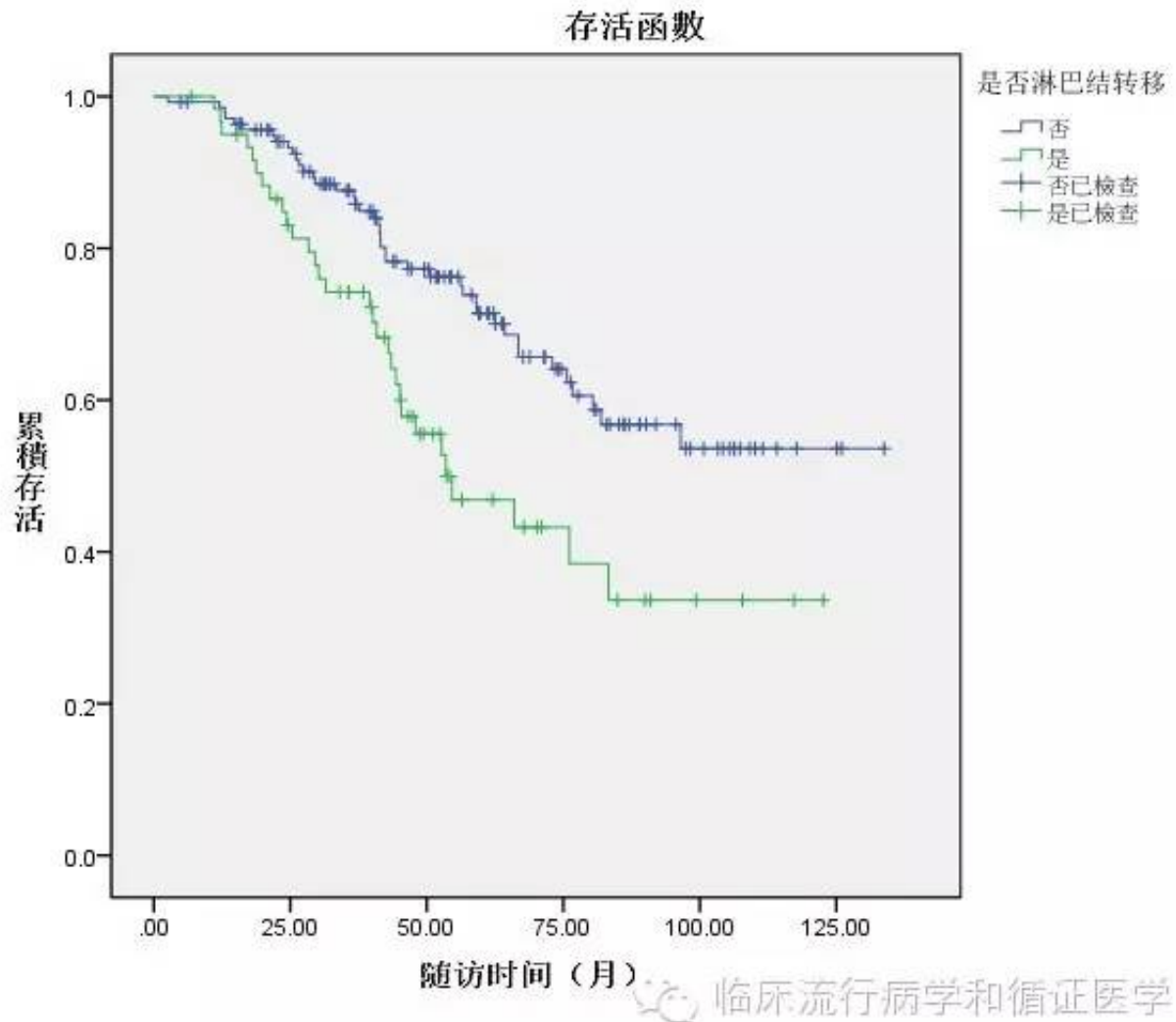
上面都看不懂?没关系，我们都知道在生存分析里随着观察时间或随访时间的推移，观察时点开始时尚存活的人数会越来越少。因此，相对而言，Breslow检验研究开始时(开始存活人数多)组间差异对卡方值的影响更大，而Log Rank检验相对Breslow检验和Tarone-Ware检验，则研究后期组间差异对卡方值影响更大。也就是说，一开始粘在一起随时间推移越来越开的生存曲线Log Rank检验要比Breslow检验更容易得到差异有统计学意义的结果;而开始相差较大，随着时间推移越来越接近的生存曲线则是Breslow检验比Log Rank检验更容易得到差异有统计学意义的结果。

下面我们用一个模拟数据来比较一下：是否淋巴结转移对肿瘤患者生存情况的比较。原始数据是比较出来Log Rank检验、Breslow检验和Tarone-Ware检验的结果比较一致，都是差异有统计学意义。

整體比較

	卡方	df	顯著性
Log Rank (Mantel-Cox)	8.922	1	.003
Breslow (Generalized Wilcoxon)	8.798	1	.003
Tarone-Ware	9.208	1	.002

是否淋巴结转移 不同層次的存活分配相等檢定。

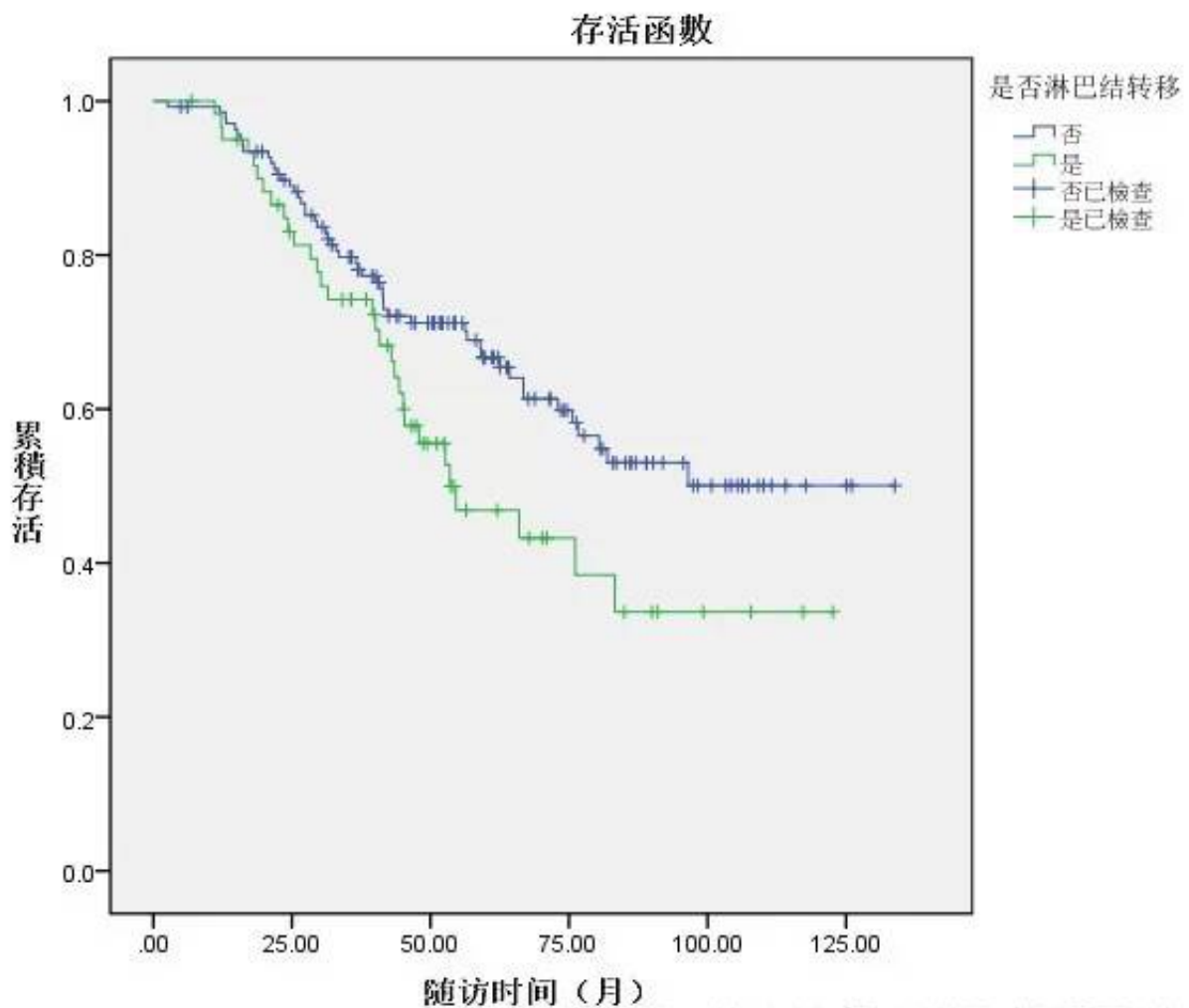


可是，小编对原始数据做了一点点改动，把随访时间较短的病例中没有淋巴结转移的患者死亡率提高了一点，从生存曲线上看就是两条曲线在观察开始时的死亡情况更相似了。这时，Breslow检验的结果是两组差异没有统计学意义，而Log Rank检验还是两组差异有统计学意义。

整體比較

	卡方	df	顯著性
Log Rank (Mantel-Cox)	4.489	1	.034
Breslow (Generalized Wilcoxon)	3.367	1	.067
Tarone-Ware	4.000	1	.045

是否淋巴结转移 不同層次的存活分配相等檢定。



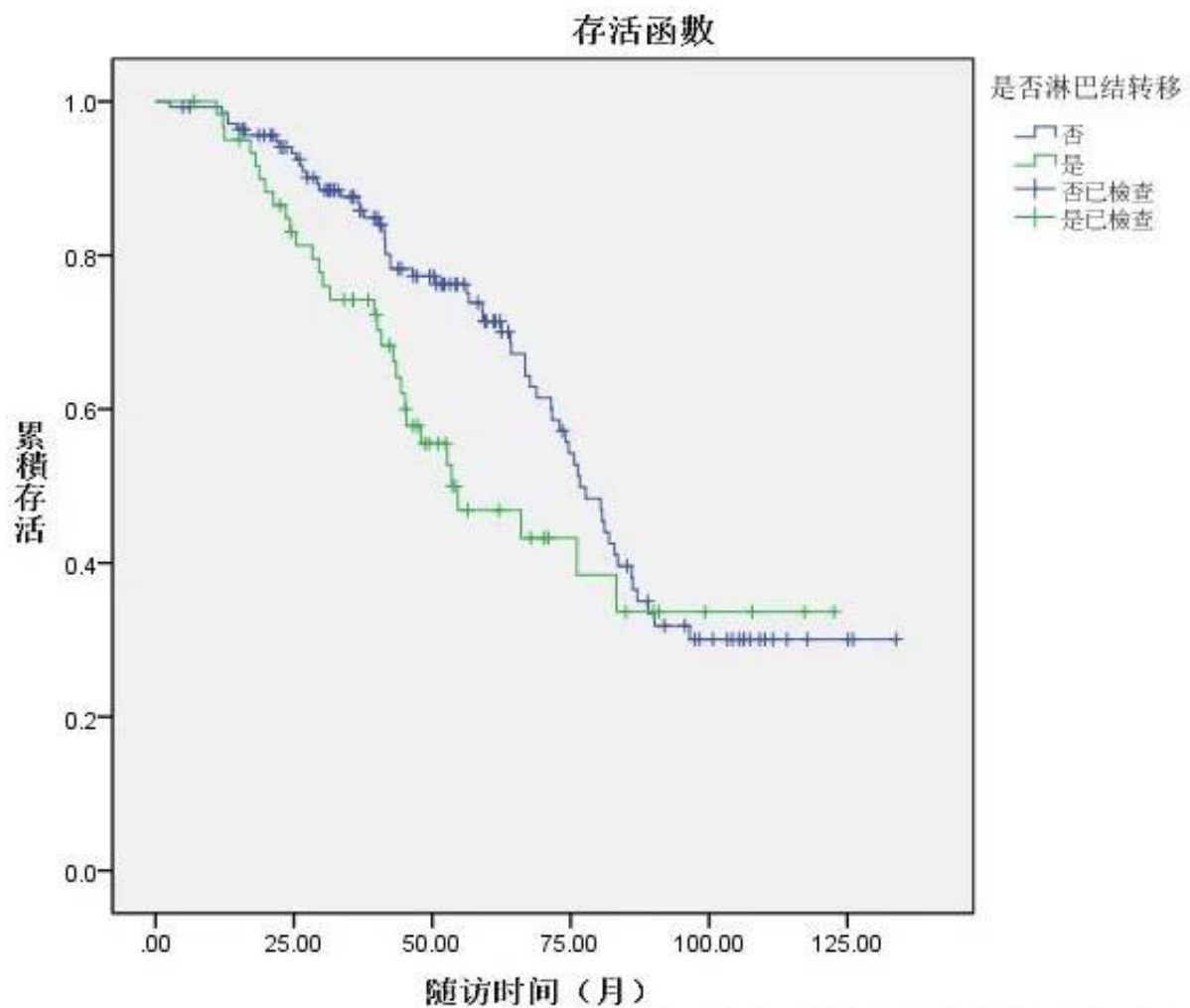
临床流行病学和循证医学

在原始数据的基础上小编这回改了随访时间较长的数据，把随访时间长的没有淋巴结转移组死亡率提高，从生存曲线上看仅是生存曲线的后面半截两组重合在一起，前面的情况和原始曲线是一样哒。这时Log Rank检验两组差异没有统计学意义，而Breslow检验两组差异有统计学意义。

整體比較

	卡方	df	顯著性
Log Rank (Mantel-Cox)	3.442	1	.064
Breslow (Generalized Wilcoxon)	7.043	1	.008
Tarone-Ware	5.867	1	.015

是否淋巴结转移 不同層次的存活分配相等檢定。



临床流行病学和循证医学

结合上面的例子，我们知道了这三种组间比较的假设检验方法其实都是卡方检验，由于Log-rank检验对所有时点都一视同仁，而生存资料一般随着时间变化生存例数是逐渐减少的，所以实质上Log-rank检验更重视远期效应。Breslow检验以观察时点存活例数作为权重，则相对重视了近期效应。而Tarone-Ware检验以观察时点存活例数平方根为权重，结果往往介于Log-rank检验和Breslow检验之间(总结可参考下表)。我们做生存分析时，应该结合这些检验的结果，统一推断组间生存情况的差异。

Kaplan-Meier 假设检验方法	同 异			
	检验类型	自由度	权重	影响
<i>Log-rank 检验</i>	卡方检验	组数-1	各时点均为 1	对远期效应敏感
<i>Breslow 检验</i>	卡方检验	组数-1	各时点存活例数	对近期效应敏感
<i>Tarone-Ware 检验</i>	卡方检验	组数-1	$\sqrt{\text{各时点存活例数}}$	介于两者之间

 临床流行病学和循证医学

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