

---

# 利用SPSS软件K-M曲线计算随访时间

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/9493.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！**

## 利用SPSS软件K-M曲线计算随访时间

。真实世界临床研究风头正盛，在真实世界临床研究中，队列研究可以说是异军突起，独领风骚。临床专病队列可以充分利用临床的患者资源，结合中长期的随访来产出真实世界临床证据。对于队列研究来说，其实施难点在于随访，如何很好的管理研究对象，如何减少失访已经在之前的文章中和大家聊过了，感兴趣的小伙伴可以检索“失访”找到相应的文章来学习一下哈。

那么当我们开展了一项队列研究，在撰写论文或者撰写研究报告的时候都需要报告一下随访时间

。不同文章的报告队列随访时间的形式也不太一样，有的文章会介绍研究对象至少随访了多长时间;有的文章会说明在队列中的研究对象随访时间的范围;更多的文章会告诉读者中位随访时间或平均随访时间。那么这个中位随访时间和平均随访时间怎么计算出来的呢?直接把每个研究对象的数据list出来计算均数和中位数么?大家想过没有，如果这样直接计算，那么在随访过程中某些研究对象出现了结局，而随访时间只能截至到事件发生时间的情况我们并没有考虑。这就像我们做生存分析的时候，如果不考虑失访事件而直接计算生存时间会有偏差一样，在计算随访时间的时候如果不考虑出现了结局事件而无法继续随访的情况，则我们计算的随访时间也是会存在一定的偏差。那么怎么更准确的计算随访时间呢?

今天推荐大家一个计算中位或平均随访时间的方法——Reverse Kaplan-Meier法。这是考虑了出现了结局事件的研究对象权重后计算的随访时间。这个这是考虑了出现了结局事件的研究对象权重后计算的随访时间。这个方法其实和我们做生存分析是一样的。有所不同的是我们计算的是随访时间而非生存时间。在我们计算生存时间的时候，把失访当作删失处理，把死亡当作结局;而在计算随访时间时，则把出现了事件(如：死亡)当作删失处理，而把失访作为结局。看到这儿，聪明的读者就知道其实Reverse Kaplan-Meier法计算失访需要数据的格式其实和我们做生存分析是一样一样滴，不同的是我们对事件状态(status)的编码不同，一般在生存分析中状态变量应1为出现了结局事件，0为未发生结局事件。而计算随访时间时则需要反转一下，1为出现了失访，0为出现了结局事件。然后再用一般的K-M法计算随访时间。过程如下图所示：

通过和K-M生存分析同样的操作就可以计算随访时间了。结果如下：

| Means and Medians for Survival Time |            |             |             |                         |            |             |             |
|-------------------------------------|------------|-------------|-------------|-------------------------|------------|-------------|-------------|
| Mean <sup>a</sup>                   |            |             |             | Median                  |            |             |             |
| 95% Confidence Interval             |            |             |             | 95% Confidence Interval |            |             |             |
| Estimate                            | Std. Error | Lower Bound | Upper Bound | Estimate                | Std. Error | Lower Bound | Upper Bound |
| 21.154                              | 1.617      | 17.985      | 24.322      | 17.000                  | 1.080      | 14.884      | 19.116      |

而如果只计算我们记录下来的随访时间会是什么结果呢?我们直接对刚才那个数据库中的Duration变量进行了描述，结果如下：

| Descriptives |                                  |             |            |
|--------------|----------------------------------|-------------|------------|
|              |                                  | Statistic   | Std. Error |
| Duration     | Mean                             | 13.0663     | .48397     |
|              | 95% Confidence Interval for Mean | Lower Bound | 12.1113    |
|              |                                  | Upper Bound | 14.0213    |
|              | 5% Trimmed Mean                  | 12.4279     |            |
|              | Median                           | 12.0000     |            |
|              | Variance                         | 42.396      |            |
|              | Std. Deviation                   | 6.51119     |            |
|              | Minimum                          | 3.00        |            |
|              | Maximum                          | 46.00       |            |
|              | Range                            | 43.00       |            |
|              | Interquartile Range              | 7.00        |            |
|              | Skewness                         | 2.028       | .181       |
|              | Kurtosis                         | 6.903       | .359       |

可见由于没有考虑出现了结局事件研究对象无法再进行随访的情况，直接计算我们对研究对象的观察时间是会低估随访时间(无论平均随访时间还是中位随访时间都较Reverse Kaplan-Meier法的计算结果短)。

总结一下，由于Reverse Kaplan-Meier法考虑了那些出现了结局事件研究对象对我们计算随访时间的影响，因此我们会认为Reverse Kaplan-Meier法得到的随访时间更准确。而且我们也发现，Reverse Kaplan-Meier法实施起来很简单，与一般的K-M生存分析法唯一不同的是我们需要改变结局编码即可。下次写队列研究论文的时候大家不妨采用一下这个方法哦。

---

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