# 回复—20201020工作汇报

**1.**  **ratetoprob 函数**

首先，我们的模型不是基于10年队列建立的，利用的不是这个人群10年的数据，而是我们通过马尔科夫模型观察这个人群在10后的情况，只用到了2010年这一年的数据，在这年的数据里人群的年龄为40-74岁，按每5年进行年龄分组，估计每个年龄分组的人群这一年的发病率和死亡率，所以t不用设置为5，而是1。

**2.**  **cycles**

这个数据不是5年一个数据，而是每5岁一组计算发病率和死亡率，举个例子，比如表里的40-44岁的发病率和死亡率，指的是2010年这一年40-44岁的人的发病率和死亡率。因为40-74岁的人按照每5年分一个组，所以就有7个年龄组。cycles是根据研究期限确定的，每个cycle的长度也是研究者根据研究目的确定的，比如，Cohort State-Transition Models in R: A Tutorial这篇文章里，研究对象基线的年龄是25岁，因为研究者想要研究这群人终生的情况，所以就将研究期限设置为75年，研究者将1个cycle的长度设置为1年，所以就有75个cycles（In the Sick-Sicker model, we simulate a hypothetical cohort of 25-year-old individuals over their lifetime (until a maximum age of 100 years) who all start in the “Healthy” state (denoted “H”). This means that we will simulate the cohort for 75 cycles）。同理，在房山人群的这个研究里，我们不观察终生的效果，而是10年的效果，我们将每个cycle长度设置为1年，所有就有10个cycles。但是与Cohort State-Transition Models in R: A Tutorial这篇文章里提到的Sick-Sicker model这个例子不同的是，房山这个人群并不是所有人的基线年龄都一样。如果所有年龄段的人的发病率和死亡率都一样，比如，如果40-44岁这个年龄组的发病率和45-49岁这个年龄组的发病率一样，那么，其实不需要基线时所有人的年龄都一样。也就是说，如果状态转换概率与年龄无关，是不需要基线时候所有人的年龄都一样的。但是由于状态转换概率和年龄是有关的，就需要针对基线时不同年龄段的人分别构建马尔科夫模型。

**3. transition probability matrix**

文章的状态转换概率不完全是time-independent的，不同年龄段的人会对应不同的心血管疾病发病和死亡率，比如，40-44岁这一组的的人会对应一个心血管疾病发病率和死亡率，也就是说，如果一个人的年龄在这个范围，他在40岁时候的心血管疾病发病率和死亡率和44岁时的是一样的，但是跨越了这个年龄段，比如变成了45岁，那就对应45-49岁这组的发病率和死亡率，这时候就是time-dependent的了。我整理的表格有不同年龄段对应的心血管病发病率和死亡率，比如基线时40-44岁的人如果之后的年龄不在这个范围了就可以用下一个年龄段的来计算他的状态转换概率，也就是说，我们需要考虑到人群在基线时年龄的差异，也要考虑随着cycles的增加年龄的变化，对于基线时的不同年龄段需要根据cycles的往后推每个年龄段之后的年龄及其对应的状态转换概率。

我是为了简化状态转换矩阵，才使用了p1, p2的表达方式，这是对同类的状态转换概率的简称，比如p1，并不能用一个p1的具体数值来赋给所有的“无心血管病人发生心血管病的概率”，而是需要根据所对应的策略来计算，由于篇幅的限制，学报文章里没有直接给出数值，但是状态转换矩阵里p1, p2等参数是可以通过我给你发的马尔科夫模型参数的表格里计算的，比如，p1（无心血管病人发生心血管病的概率）可以通过”模型参数“这个表里参数来计算的，如果这个人是一般人群，就可以在“一般人群心血管病发病率”这个表里用这个人对应的年龄段的参数来计算；如果这个人不是一般人群，比如，他是根据策略1划分出的中危人群，就在”模型参数“这个表里的p\_live\_cvd\_m这个变量这里找到它的计算方式。

另外，为了帮助你更好地理解我整理的excel表格里参数的含义，附上单个马尔科夫模型的Treeage框架，可以结合“状态转换图”这个表里的图理解状态转换过程。

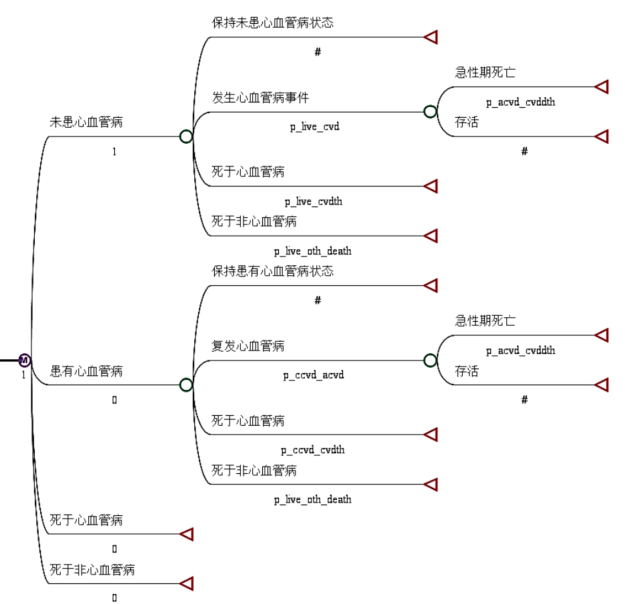


图1 马尔科夫模型