

• 方法评价 •

队列研究中人年的分配

李克 俞顺章

队列研究中，调查对象在整个研究期间处于持续地观察（随访）之下，这是队列研究与横断面、病例-对照及其他研究相区别的基本特征。队列成员在随访过程中贡献的人年及死亡（或发病）被记录下来，为统计分析中率的计算和推断提供了基本数据。本文重点论述在有时间协变量情况下观察人年的动态分布。

一、随访定义：有两个方面：①确定研究的最终结局，可以是疾病的死亡、发生、有效或痊愈等，依研究目的而定；②确定研究起止时间，如果研究的终点定为死亡，那么假定个体从进入到退出队列的整个期间是处于死亡的“危险”，这意味着研究时期不应包括对象已知是活着的那段时期，例如，一个队列定义为某工厂中至少工作 5 年的所有工人，这开始工作的 5 年应排除在观察时间之外。随访意指有足够的方法追踪个体在研究期间的生命状态，如果个体在这之间的某个时点之后无记录，通常认为该个体离开了队列，称失访。失访的频率如较高，会引进选择偏倚，因为失访者的死亡年龄和死亡原因与未失访者可能有明显不同。

二、人年分配原则：

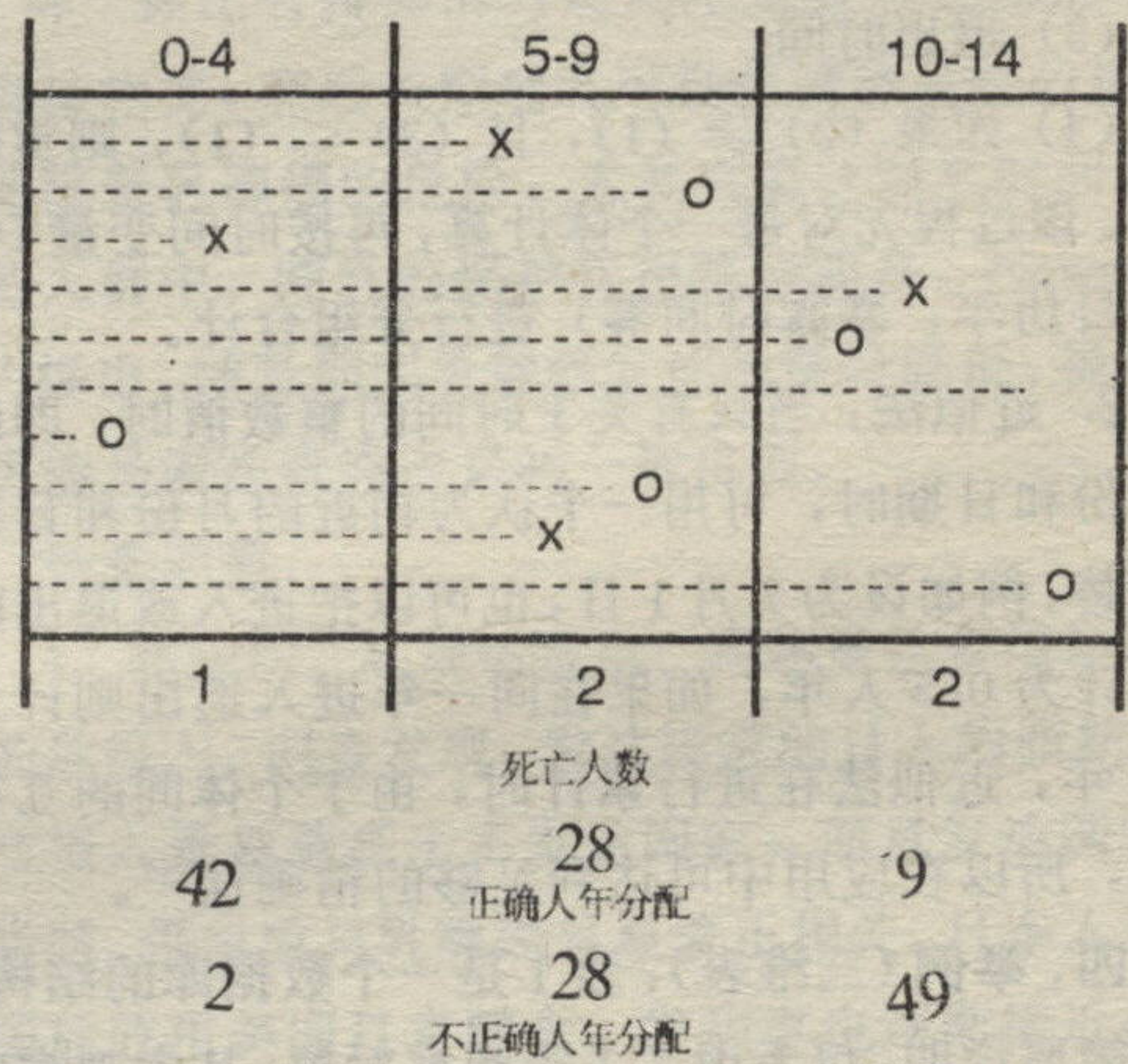
1. **固定法：**最简捷的方法是根据进入队列时的信息确定亚组，例如，根据进入队列时的年龄、日历年或暴露水平和时间来确定亚组，将各亚组作为单独的队列，然后计算个体按年龄和日历年的人年和死亡数。由于个体仅对一个亚组贡献人年，所以观察人年的分配不会混淆。

2. **动态法：**根据个体在整个研究期间的移动而导致变量的值发生变化的基础上定义亚组，例如，对象在随访的同时暴露也积累，呈连续上升分布，这样，个体的暴露分组也会改变，在不同的暴露组中贡献人年，这是较为复杂的人年分配方法，但这样分析暴露与疾病危险相互关系时，所得结果更为真实。

3. 有些研究根据个体在研究结束时的累计暴露

来确定亚组，然后进行亚组间的比较，以期建立剂量-反应趋势，然而，这种方法是错误的，它会低估最高暴露组的疾病率，高估最低暴露组的疾病率。

动态地分配人年是正确的方法^[1,2]，所反映的情况是真实的，通过附图可以更清楚。



附图 按观察时间比较固定法与动态法的人年计算和死亡分配（O 表示死亡，X 表示退出）

三、计算方法：

1. **精确法：**在实际中需要研究几个关于暴露的时间协变量，例如，暴露年数，停止暴露的年数等，另外，为了标化的目的，还需要有年龄和日历年。计算个体在这个四维表的交叉分组中贡献的观察时间的精确长度，看来是比较繁琐的。

Clayton 曾描述了人年分布的计算机算法^[3]，它要求个体具有时间协变量的进入时间和退出时间，虽然该算法思想不能解决所有的问题，例如对于中断暴露后又继续暴露的问题，但有简单的优点，例如，确定某个体在年龄 40~49 岁，日历年 1950~1954 年，及对某危险因素暴露 5~10 年格子中的观察时间，Clayton's 过程如下：

(1) 选择 3 个最早日期：出生年+40 年，1949 年 12 月 31 日，及首次暴露后时间+5 年；

(2) 选择 4 个最晚日期：出生年+50 年，1954 年 12 月 31 日，及开始暴露后时间+10 年；

(3) 如果 (2) 先于 (1), 个体对该组没有贡献, 否则, 所贡献的人年是 (1) 至 (2) 的时间。这个例子是一个三维的表格, 是按组对每个个体重复运算来计算人年的。

还可选择另一种算法 (任爱国, 私人通信), 它是按个体在每个日历年中对于各组所贡献的人年数来计算的, 该过程会产生一个相当冗长的数据库, 对于计算机的要求较高, 程序开始仅考虑了计算一个两维表 (年龄和日历年) 的观察人年, 以后我们发展了死亡数和死因的分配, 并增加了与暴露有关的时间协变量的多维表人年计算及死亡分配, 过程为:

- (1) 选择一个日历年 (在进入年-退出年之间);
- (2) 进入时间;
- (3) 退出时间;

(4) 如果 (2) \geq (1), 且 (3) \leq (1), 则贡献人年。该过程先对每一个体计算, 再按时间变量 (年龄, 日历年, 暴露时间等) 进行分组合计。

2. 近似法: 当只有关于时间的整数值时, 即缺乏月份和日期时, 可用一个认为相近的月份和日期来代替, 例如设为 7 月 1 日, 也可以把进入或退出的当年计为 0.5 人年, 如果在同一年进入退出则计为 0.25 年, 近似法在进行累计时, 由于个体间的互相抵消, 所以在应用中可达到足够的精确性^[4]。

四、举例 (三维表): 表 1 是一个数据库的结构, 为了简要说明, 仅考虑了 1 个观察对象。其中观察时间、年龄和日历年为动态的时间变量, 性别为固定变量。

表 1 数据库结构及一个观测例子

出生年	性别	死亡年月	随访开始年龄	进入队列时间	退出队列时间
1897	1	1981-03-15	76 岁	1972-01-01	1981-03-15

该研究定于 1972 年 1 月 1 日开始随访, 研究结束时间为 1984 年 12 月 31 日, 该对象在研究期间死亡, 应用前面讨论的程序进行人年计算和死亡的分配, 基于个体的计算结果见表 2。

程序的第二步按变量分层输出总结性结果。表 3 形成的数据库可用于年龄-时期专率的计算, 与标

准人群的年龄日期性别专率数据库相连接, 可计算 SMR, 与标准人群年龄构成数据相联结, 可计算直接标化率, 该数据库还可用于统计模型分析。

表 2 按个体计算人年和分配死亡

日历年	年龄	观察时间	性别	人年	死亡数
1972	77.0000	1.00000	1	1.00000	0
1973	78.0000	2.00000	1	1.00000	0
1974	79.0000	3.00000	1	1.00000	0
1975	80.0000	4.00000	1	1.00000	0
1976	81.0000	5.00000	1	1.00000	0
1977	82.0000	6.00000	1	1.00000	0
1978	83.0000	7.00000	1	1.00000	0
1979	84.0000	8.00000	1	1.00000	0
1980	85.0000	9.00000	1	1.00000	0
1981	85.2083	9.20833	1	0.20833	1

表 3 按亚组合计人年和分配死亡

日历年	年龄	性别	观测时间	人年数	死亡数
1970	75	1	0	3.00	0
1975	80	1	0	1.00	0
1975	80	1	5	4.00	0
1980	85	1	5	1.21	1

参 考 文 献

- 1 Enterline PE. Pitfalls in epidemiologic research; an examination of the asbestos literature. J Occup Med, 1976, 18: 150.
- 2 Wagoner JK, Infante PF, Saracci R. Vinyl chloride and mortality. Lancet, 1976, ii: 194.
- 3 Clayton DC. The analysis of prospective studies of disease aetiology. Commun Stat Theory Meth, 1982, 11: 2129.
- 4 Breslow NE, Day NE. Statistical methods in cancer research. Vol II. The design and analysis of cohort studies. Lyon: IARC, 1987.

(收稿: 1996-03-04 修回: 1996-05-15)