

实习 4 队列研究

【目的】 掌握队列研究的基本概念和研究方法；熟悉队列研究资料的处理方法。

【学时】 3 学时

【内容】

一、队列研究的基本原理

队列研究是指在一个特定的人群中选择所需的研究对象，根据他们目前或者过去某个时期是否暴露于所研究的危险因素，或者暴露于研究因素的不同水平而分成不同的组，由此随访并比较不同组别研究对象的各种预期结果发生情况，从而检验研究因素与所研究的结局是否存在关联。

【课题一】 为了解目前我国广泛应用的国产低剂量复方口服避孕药(COC)与脑卒中发病危险性的关系，于 1997 年 7 月至 2000 年 6 月在江苏太仓市和如东县 25 个乡镇，随访并比较 44408 名使用甾体激素避孕药(HC)和 75230 名使用宫内节育器 IUD 妇女的脑卒中发病情况。表 4-1 是两组人群出血性卒中的发病情况：

表 4-1 不同避孕方式妇女出血性卒中的发病情况

避孕方法	观察人年	病例数
复方口服避孕药	129648.63	52
宫内节育器	216752.63	23
合计		

问题 1：上述研究属于何种类型的流行病学研究？

问题 2：用什么指标描述各组人群的发病危险？

【课题二】 为了证实非职业性环境接触青石棉与发生肺癌和间皮瘤危险的关系，对云南省大姚县青石棉污染区和作为对照的同省无石棉污染的禄丰县进行了既往 9 年（1987 年 1 月 1 日到 1995 年 12 月 31 日）的死亡率调查，两县在民族构成、生活习惯、文化教育、地理气候以及性别和年龄构成上均具有可比性。调查结果见表 4-2：

表 4-2 接触青石棉与各种肿瘤的发病情况

	调查人数	观察人年	死亡人数(死亡率, 1/10 万人年)				
			全肿瘤	肺癌	间皮瘤	胃癌	肠癌
暴露组	4543	39430.05	72(182.60)	21(53.26)	7(17.75)	6(15.22)	6(15.22)
非暴露组	5626	48236.48	60(124.39)	12(24.88)	0	9(18.66)	3(6.22)
	10169	87666.53	132	33	7	15	9

问题 1：上述研究属于何种类型的流行病学研究？和课题一比较，有何区别和联系。

二、队列研究资料的分析

不管是前瞻性队列研究，还是回顾性队列研究，队列研究资料分析的核心就是测量暴露组和非暴露组、或者不同暴露水平组的发病或死亡情况并比较它们之间的差异。然后在此基础上计算暴露和疾病发病（死亡）的关联强度指标；当暴露因素是按照暴露水平进行分级时，还需要分析是否存在剂量反应关系；如果收集了可能的混杂信息，还可以进行分层

分析。

问题 1：队列研究常用的率的指标有哪些？分别适用于什么情况？

问题 2：常用的计算暴露与发病的关联强度指标有哪些？

【课题三】 对【课题一】的资料进行分析，填写下表。

表 4-3 不同避孕方式妇女与出血性卒中发病的关系

避孕方法	观察人年	病例数	发病密度	u	RR	RR95%CI
复方口服避孕药	129648.63	52				
宫内节育器	216752.63	23				
合计						

提示：u 检验公式：

$$u = \frac{P_1 - P_0}{\sqrt{\bar{P}(1 - \bar{P}) \left(\frac{1}{N_1} + \frac{1}{N_0} \right)}}$$

能否根据上表的结果得出结论，认为与宫内节育器相比，服用口服避孕药增加了已婚妇女出血性卒中的危险？为什么？你可以参考表 4-4 的信息进行讨论。

表 4-4 进入队列时两组妇女的基本特征

	复方口服避孕药组 n=35930	宫内节育器组 n=75230
年龄		
≤24 岁	406	4998
25~	13142	34899
35~	13482	22279
45~	7342	9531
55~	1558	3524
文化程度		
文盲	13708	21518
小学	3724	6326
高中	154	39191
高中及以上	3046	8195
职业		
农民	29989	52346
非农民	5941	22884
吸烟	421	777
饮酒	1583	2982
高血压	154	90
脑卒中家族史	758	1159

【课题四】 下表是 Framingham 研究中 1045 名 33-49 岁男性随访 10 年的冠心病的发生

情况，研究者首先检测了 1045 名 33~49 岁男子的血清胆固醇含量，然后按其水平高低分为 5 组，随访观察 10 年后各组冠心病发病人数如下表所示：

表 4-5 33~49 岁男子不同血清胆固醇水平组的冠心病发病情况

血清胆固醇 (mg/dl)	观察人数	病例数
114~	209	2
194~	209	11
214~	209	14
231~	209	26
256~	209	32
合计	1045	85

问题 1：以血清胆固醇水平 114~193mg/dl 组作为参照组，分别计算各暴露水平组的 RR、RR95%的可信区间、AR 以及 AR%。填入下表并解释各个指标的意义。

问题 2：根据以上结果，能否认为随着血清胆固醇水平的升高，研究人群冠心病发病的危险也增加？

表 4-6 33~49 岁男子血清胆固醇水平与冠心病发病的关系

血清胆固醇 (mg/dl)	暴露水平 (X _i)	病 例 数 (a _i)	非病例数 (b _i)	观察人数 (n _i)	CI(%)	χ^2	RR	95%CI	AR	AR%
114~	0	2	207	209	0.96					
194~	1	11	198	209	5.26					
214~	2	14	195	209	6.70					
231~	3	26	183	209	12.44					
256~	4	32	177	209	15.31					
合计	-	85	960	1045	8.13					

提示：多个样本率比较：

$$\chi^2 = n \left(\sum \frac{A^2}{n_r n_c} - 1 \right)$$

$$\gamma = (n_r - 1)(n_c - 1)$$

趋势性 χ^2 检验公式：

$$\chi^2_{m-EXT} = \frac{n^2(n-1) \left[\sum a_i \chi_i - \left(\sum a_i \right) \left(\sum n_i \chi_i \right) / n \right]^2}{\sum a_i \sum b_i \left[n \sum n_i \chi_i^2 - \left(\sum n_i \chi_i \right)^2 \right]}$$

【课题五】 以下数据来源于英国石棉工人死亡研究，该研究的目的是寻找石棉工人的肺癌死亡率为什么高于普通人的原因。该研究在 1955 年进行，研究对象是那些曾经或现在还在从事接触石棉的工作，并且接触时间至少 20 年以上的工人。

研究者从各个石棉加工工厂得到合格研究对象 113 名，他们开始进入队列的时间从 1923 年到 1952 年。那些在工厂工作超过 20 年并有 20 年的完整资料，同时暴露时间超过 20 年的石棉接触工人才被认为是合格的研究对象。一旦成为合格的研究对象，这些工人便被认为是“暴露人群”，不管他们是否继续在原来的岗位或工厂工作。

研究对象被确定后，研究者从各种渠道对他们进行了随访，确定研究对象是否死亡，如果已经死亡，查证死亡原因。结果 113 个研究对象中共有 39 人死亡。表 4-7 给出了 39 名死者的死因。同时给出的还有同期英格兰和威尔士相同年龄普通男性的死亡率。

表 4-7 39 名石棉工人死因构成

肺癌	其他呼吸道疾病和心血管疾病	其他系统疾病	合计
10	21	8	39

表 4-8 同期英格兰和威尔士同年龄段的普通男性死亡率 (/1000)

年代	肺癌	其他呼吸道疾病和心血管疾病	其他系统疾病	合计
1930	0.3	6.3	7.0	13.6
1937	0.4	6.5	6.6	13.5
1942	0.6	7.6	8.2	16.4
1947	0.7	7.0	6.7	14.4
1952	1.0	7.7	6.0	14.7

问题 1：根据以上资料能否提示肺癌可能是石棉接触工人的一种职业危害？为什么？石棉接触工人肺癌死亡比例高于普通人群还有哪些可能的原因？

表 4-9 记录了 113 名研究对象接触暴露满 20 年，被纳入队列的时间，以及死亡或失访时间。以普通人群为标准，按下列步骤估计石棉接触工人肺癌，其他呼吸道疾病和心血管疾病，其他系统疾病和总的疾病的期望死亡数，以完成表 4-10。

从表 4-8 可以看出，观察期内英格兰和威尔士普通男性的肺癌死亡率随年代不同而存在明显的差异，所以，计算观察期内总的期望死亡数，应该是不同年代的期望死亡数的总和。

第一步：按照研究对象进入队列的时间不同化为几个年代：1923-1934；1935-1939；1940-1944；1945-1949；1950-1954，分别计算各个年代的暴露人年。

第二步：应用表 4-8 中相应年代的率计算该年代的预期的死亡数

第三步：求各年代的期望死亡数的总和

表 4-9 113 名接触暴露满 20 年研究对象被纳入队列，死亡或失访时间

进 入 队 列时间	结 束（或死亡或 失访）时间	人 数	进 入 队 列时间	结 束（或死亡或 失访）时间	人 数
1923	1927	1	1941	1949	1
1923	1944	1	1941	1954	1
1924	1947	1	1941	1955*	4
1925	1946	1	1942	1942	1
1925	1952	1	1942	1950	1
1930	1935	1	1942	1951	1
1933	1942	2	1942	1955*	3
1933	1947	1	1943	1955*	2
1933	1950	2	1944	1948	2
1934	1954	1	1944	1955*	6
1935	1941	1	1945	1945	1
1935	1953	1	1945	1949	2
1936	1946	2	1945	1953	1
1936	1946	1	1945	1955*	11
1936	1955*	1	1946	1951	2
1937	1952	1	1946	1955*	6
1937	1953	1	1947	1954	1
1938	1955*	1	1948	1948	1
1939	1942	1	1948	1955*	7
1939	1954	1	1949	1952	1
1939	1955*	2	1949	1955*	6
1940	1944	1	1950	1951	1
1940	1951	2	1950	1955*	6
1940	1955*	2	1951	1955*	6
1940	1955*	2	1952	1955*	5
1941	1946	1	合计		39

*表示研究结束时仍然存活

表 4-10 不同时期期望死亡人数

时 期	暴 露 人 年	期 望 死 亡 数			
		肺 癌	其他呼吸道疾病和 心血管疾病	其他系统疾病	全 部 疾 病
1923-1934					
1935-1939					
1940-1944					
1945-1949					
1950-1954					
合计					

问题 2：与普通人群相比，石棉接触工人患肺癌、其他呼吸道疾病、心血管疾病、其他系

统疾病和总的疾病的危险怎样？这样的结果说明什么？能否根据以上分析认为长期接触石棉，导致死于肺癌的危险增加？为什么？

表 4-11 39 名石棉工人观察期内预期死亡人数和实际死亡人数

	肺癌	其他呼吸道疾病和心血管疾病	其他系统疾病	合计
实际死亡人数	10	21	8	39
预期死亡人数				
SMR				
SMR95%CI				

提示：SMR 是否为 1 的假设检验：

$$\chi^2 = \frac{(|A - E| - 0.5)^2}{E} \dots\dots$$

SMR95%CI 的估计：

$$A_L = (\sqrt{A} - 1.96 \times 0.5)^2$$

$$A_U = (\sqrt{A} + 1.96 \times 0.5)^2$$

$$SMR_L = A_L / E$$

$$SMR_U = A_U / E$$

【课题六】在 Framingham 研究中提供了该地区 35-44 岁男性人群中几种冠心病危险因素

素的相对危险度与人群暴露比例的资料，完成下表并回答问题：

表 4-12 35~49 岁男子中几种冠心病发危险因素的 RR 和 PAR%

危险因素	RR	Pe	PAR%
收缩压 $\geq 180\text{mmHg}$	2.8	0.02	
X 线心脏扩大	2.1	0.10	
吸烟	1.9	0.72	

问题 1：PAR%和 AR%有何区别？他们的意义有何不同？

问题 2：相对危险度 RR、人群暴露比例和人群归因危险度 PAR%之间有什么关系？这种关系对于决策部门制订公共卫生政策有什么指导意义？

思考题：请根据上述课题的分析，总结不同队列研究类型的分析思路。

参考答案

一. 队列研究的基本原理

【课题一】

该研究是回顾性队列研究，与病例对照研究比较，主要区别是对暴露与其引起的后发病的观察时间顺序不同，回顾性队列研究的观察的方向为从因至果，而病例对照研究的观察的方向为从果究因。

【课题二】

该研究是前瞻性队列研究，与“课题一”比较，两者的相同点均是在靶人群中抽取一个样本，按目前或过去某个时期内是否暴露于所研究的疑似病因（暴露）或其不同的水平而将研究对象分成不同的组，观察比较各组某病的发病率或死亡率，用以研究暴露和与之相连的后发病或死亡危险的关联。

两者的基本区别在于研究的起点不同。前瞻性队列研究是对研究对象从研究开始时追踪到将来，回顾性队列研究是从过去追踪到现在。

二. 队列研究的资料分析

【课题三】

根据“课题一”的资料计算的各统计量值及各关联指标值

肿瘤	死亡率 (1/10 万人年)		U	RR	RR95%CI (Miettinen)	AR (1/10 万人年)	AR%	RR95%CI (Woolf 法)
	暴露组	对照组						
全肿瘤	182.60	124.39	2.2098*	1.47	1.0445~2.0687	56.21	31	1.04~2.08
肺癌	53.26	24.88	2.1547*	2.14	1.0695~4.2815	28.38	53	1.052~4.35
间皮瘤	17.75	2.07*	2.4176**	>8.57	1.5017~48.9067	>15.68	>88	1.05~69.68
胃癌	15.22	18.66	0.3875	0.82	0.3005~2.2374	-3.44	-23	0.29~2.31
肠癌	15.22	6.22	1.3080	2.45	0.6398~9.3825	9	59	0.61~9.80

* $P < 0.05$, ** $P < 0.01$; *假设对照组中有 1 例间皮瘤，死亡率则为 2.07; $RR95\%CI = RR^{(1 \pm 1.96u)}$;

(罗素琼, 1997)

u 检验结果可知，暴露组肺癌的发病率高于对照组， $RR=2.14$ ，结合其 95%CI 可认为暴露者发生肺癌的危险性是非暴露者的 2.14 倍。 $AR=28.38/10$ 万人年，说明单纯由于暴露于青石棉而使暴露组的发病率增加了 28.38/10 万人年。 $AR\%=53\%$ ，说明暴露人群中肺癌发病归因于暴露于青石棉的成份占全部病因的 53%。

从表中资料可以看出暴露组的全肿瘤、肺癌的死亡率高于对照组，有统计学意义；间皮瘤是与石棉接触密切相关且较为特异的肿瘤；本次研究未发现支持青石棉污染引起胃肠道肿瘤发病增高的证据。

【课题四】

根据“课题二”的资料计算的统计量值

血清胆固醇 (mg/dl)	暴露水平(x_i)	病例数 (a_i)	非病例数 (b_i)	观察人数 (n_i)	CI(%)	χ^2	RR	RR(95%CI)	AR(%)	AR%
114—	0	2	207	209	0.96	—	1	—	—	—
194—	1	11	198	209	5.26	6.43	5.5	1.69—17.79	4.3	82
214—	2	14	195	209	6.70	9.36	7.0	2.23—21.81	5.74	86

231—	3	26	183	209	12.44	22.05	13	4.42—37.99	11.48	92
256—	4	32	177	209	15.31	28.81	16	5.39—45.08	14.35	93
合计	—	85	960	1045	8.13	—	7.1	—	7.2	—

与“课题三”相比，“课题四”的暴露水平分为多个等级，因此可进行趋势检验，来判断暴露因素与发病之间是否存在剂量反应关系

(1) H_0 : 不同暴露水平组冠心病的发病率相等； H_1 : 不同暴露水平组冠心病的发病率不相等

$$\alpha=0.05$$

$$\chi^2=36.88$$

$$\nu = 4$$

本例 $\chi^2=36.88 > \chi_{0.05(1)}^2$, $P < 0.05$, 按 $\alpha=0.05$ 水准, 拒绝 H_0 , 接受 H_1 。说明不同暴露水平组冠心病的发病率不相等。

(2) 将队列研究中累积发病率资料按暴露水平分组, 组成如下表所示的 2×5 有序列联表。

队列研究中累积发病率资料 (暴露因素分为 5 个水平)

暴露水平	0	1	2	3	4	合计
	y_1	y_2	y_{c-1}	y_c	
发病人数	2	11	14	26	32	85
未发病人数	207	198	195	183	177	960
合计	209	209	209	209	209	1045

$$H_0: RR_0=RR_1=RR_2=RR_3=RR_4$$

$$H_1: RR_0 < RR_1 < RR_2 < RR_3 < RR_4$$

$$\alpha=0.05$$

$$\sigma^2 = 2.00$$

$$\chi^2 = 35.98$$

$$\nu = 1$$

对于本例 $\chi^2 (=35.98) > \chi_{0.05(1)}^2$, $P < 0.05$ 。按 $\alpha=0.05$ 水准, 拒绝 H_0 , 接受 H_1 。认为血清胆固醇水平与冠心病发病之间存在剂量反应关系, 随着血清胆固醇水平的升高, 冠心病发病率 (CI)、RR、AR 也增高, 从而为病因推断提供了更为有力的证据。[注 $\chi_{0.05(1)}^2=3.84$]

【课题五】

弗明汉 35—44 岁男子中几种冠心病危险因素的 RR 与 PAR%

危险因素	RR	Pe	PAR%
收缩压 $\geq 180\text{mmHg}$	2.8	0.02	3.47
X 线上心脏扩大	2.1	0.10	9.91
吸烟	1.9	0.72	39.32

(Kahn HA, 1983)

问题 1:

PAR%：人群归因危险度百分比，是一种测量某危险因素对靶人群潜在影响的指标，用来说明靶人群中某种疾病的发生归因于某种暴露引起的发病占人群中发病的百分比。

AR%：归因危险百分比，是一种测量联系强度的指标，用来说明靶人群中暴露者中某种疾病的发生归因于某种暴露引起的发病占暴露者中发病的百分比。

问题 2：

当 RR 一定时， P_e 越大，PAR% 越大；当 P_e 一定时，RR 越大，PAR% 也越大。只有当 RR 较大，而且 P_e 也较大时，才会有较大的 PAR%。所以当人群中暴露于某因素的比例很小时，该因素对靶人群的潜在影响不会很大。本例中，虽然收缩压 $\geq 180\text{mmHg}$ 者的 RR 值最大，但由于其 P_e 值很小，所以其 PAR% 值，即公共卫生意义也不大。

【课题六】

Mount-Sinai 研究中石棉工人标准化死亡比

死因	观察数	期望数	SMR	SMR95%CI
肺癌	475	89.4	5.31	4.84—5.79
胃肠道肿瘤	140	67.9	2.06	1.72—2.40

(Selikoff, 1978)

问题 1：

(1)对职业人群肺癌的 SMR 与 1 的差别作假设检验

$H_0: \text{SMR}=1$

$H_1: \text{SMR}>1$

$\alpha=0.05$ (单侧)

$$u = \frac{D - E}{\sqrt{E}} = 40.7820$$

$u = 40.7820 > u_{0.05(2)}$, $P < 0.05$ 。按 $\alpha=0.05$ 水准，拒绝 H_0 ，接受 H_1 。可以认为职业人群肺癌 $\text{SMR} > 1$ ，说明该职业人群发生肺癌的危险超过一般人群。

(2)计算总体 SMR 可信区间

$$\left(\frac{D - 1.96\sqrt{D}}{E(D)}, \frac{D + 1.96\sqrt{D}}{E(D)} \right) = (4.84, 5.79)$$

(3)研究结果说明石棉粉尘可以引起肺癌和胃肠道肿瘤，但发生肺癌的危险更大，说明接触粉尘与发生肺癌有一定的特异性。

【课题七】

石棉含量指数与肺癌死亡的关系

含量指数 (mppcf,years)	肺癌			胃肠道肿瘤		
	O	E	SMR	O	E	SMR
<125	18	11	1.7	27	22	1.2
125~249	11	5	2.2	5	9	0.6
250~499	16	5	3.0	19	10	1.9
500~749	9	2	5.0	6	3	1.8
>750	5	1	5.6	2	2	1.2

问题 1：先将资料整理成如下的 SMR 线性趋势 χ^2 检验用表。

SMR 线性趋势检验用表						
暴露等级 X_i	1	2	3	4	5	合计
观察死亡数 A_i	18	11	16	9	5	59(A)
期望死亡数 E_i	11	5	5	2	1	24(E)
$X_i(A_i - E_i)$	-9.04	-2.58	11.13	16.33	12.71	28.55
$X_i E_i$	27.04	24.58	36.88	19.67	12.29	120.46
$X_i^2 E_i$	27.04	49.16	110.64	78.68	61.45	326.97

$$E_i = E \times \frac{A_i}{A}$$

$$A = \sum A_i, E = \sum E_i$$

H_0 : 石棉含量指数等级与肺癌死亡率之间不存在剂量反应关系

H_1 : 石棉含量指数等级与肺癌死亡率之间存在剂量反应关系, 随着石棉含量指数的增高, 肺癌的死亡率增加

$\alpha=0.05$ (单侧)

$$\chi^2 = 10.06$$

$$\nu=1$$

本例 $\chi^2 (=10.06) > \chi^2_{0.05(1)}$, $p < 0.05$ 。故拒绝 H_0 , 接受 H_1 。认为石棉含量指数等级与肺癌死亡率之间存在剂量反应关系, 随着石棉含量指数的增高, 肺癌的死亡率增加。题中已知胃肠道肿瘤的 SMR 与粉尘含量指数的线性趋势检验结果 $\chi^2 = 1.65$, 说明石棉含量指数等级与胃肠道肿瘤发病之间不存在剂量反应关系。[注 $\chi^2_{0.05(1)} = 3.84$]。

【课题八】

表 6-10 随访人群观察人年计算表

年份	HBsAg 阳性人群				HBsAg 阴性人群			
	年初人数	不变人数	退出人数	人年数	年初人数	不变人数	退出人数	人年数
1977	1195	1193	2	1194.0	5077	5070	7	5073.5
1978	1193	1188	5	1190.5	5070	5059	11	5064.5
1979	1188	1187	1	1187.5	5059	5055	4	5057.0
1980	1187	1176	11	1181.5	5055	5037	18	5046.0
1981	1176	1166	10	1171.0	5037	5033	4	5035.0
1982	1166	1153	13	1159.5	5033	5012	21	5022.5
1983	1153	1135	18	1144.0	5012	4982	30	4997.0
1984	1135	1122	13	1128.5	4982	4949	33	4965.5
1985	1122	1113	9	1117.5	4949	4924	25	4936.5
1986	1113	1109	4	1111.5	4924	4902	22	4902.5
合计	——	——	86	11029.5	——	——	175	47654.0

HBsAg 阳性人群肝癌 ID=41/11029.5=37.17/万人年

HBsAg 阴性人群肝癌 ID=16/47654.0=3.36/万人年

RR=11.06

AR=33.81/万人年

AR%=91.0%

以所研究的人群代表全人群估计 PAR%

$I_t=(41+16)/(11029.5+47654)=9.71/\text{万人年}$

$PAR\%=(9.71-3.36)/9.71 \times 100\%=65.4\%$

【课题九】

该工厂人群全死亡 SMR=0.77(95%CI 为 0.61~0.95),说明该厂人群总死亡率低于标准人群。这在职业流行病学研究中是常见现象,往往是由于健康工人效应造成的。

该工厂人群肺癌死亡 SMR=2.14(95%CI 为 1.27~3.25),说明该厂人群接触肺癌危险因素的机会高于一般人群。