

重点难点指导：

重点：基本原理、研究的实施、资料整理分析；

难点：偏倚及其控制。

1、队列研究的基本原理：在一个特定人群中选择所需的研究对象，根据目前或过去某个时期是否暴露于某个待研究的危险因素，或其不同的暴露水平而将研究对象分成不同的组，如暴露组和非暴露组，高剂量暴露组和低剂量暴露组等，随访观察一段时间，检查并登记各组人群待研究的预期结局的发生情况(如疾病、死亡、或其他健康状况)，比较各组结局的发生率，从而评价和检验危险因素与结局的关系。

在队列研究中，所选研究对象必须是在开始时没有出现研究结局，但有可能出现该结局(如疾病)的人群。暴露组与非暴露组必须有可比性，非暴露组应该是除了未暴露于某因素之外，其余各方面都尽可能与暴露组相同的一组人群。根据基本原理，可以分析出一些基本特点：

1) 属于观察法 2) 设立对照组 3) 由“因”及“果” 4) 能确证暴露与结局的因果联系。

2、队列研究的实施

(1)、确定研究因素：常称为暴露因素或暴露变量，通常是在描述性研究和病例对照研究的基础上确定的。

(2)、确定研究结局：结局变量(outcome variable)也叫结果变量，简称为结局，指随访观察中将出现的预期结果事件，即研究者希望追踪观察的事件，是观察的自然终点，不是观察期的终止。

(3)、确定研究现场：除要求有足够数量的符合条件的研究对象，还要求当地领导重视，群众理解和支持。

(4)、确定研究人群：包括暴露组和对照组：

1) 暴露人群的选择：有四种选择：(1)职业人群：研究某种可疑的职业暴露因素与疾病或健康的关系的选择。(2)特殊暴露人群：研究某些罕见的特殊暴露的唯一选择，如选择原子弹爆炸的受害者，研究射线与白血病的关系。(3)一般人群：即某行政区域或地理区域范围内的全体人群。(4)有组织的人群团体：一般人群的特殊形式，如医学会会员，工会会员等。

2)、对照人群的选择：(1)内对照：即先选择一组研究人群，将其中暴露于所研究因素的对象作为暴露组，其余即为非暴露组；(2)特设对照，也称外对照，选择职业人群或特殊暴露人群作为暴露人群时，在该人群之外去找对照组；(3)总人口对照：利用整个地区的现成的发病或死亡统计资料，对照中可能包含有暴露人群；(4)多重对照：或叫多种对照，即用上述两种或两种以上的形式选择的人群同时作对照。

(5)、确定样本大小：影响样本含量的几个因素 1) 一般人群(对照人群)中所研究疾病的发病率  $p_0$  越接近 0.5，样本量越大；2) 暴露组与对照组人群发病率之差 ( $d=p_1-p_0$ )  $d$  值越大，样本量越小；3) 要求的显著性水平，即检验假设时的第 I 类错误(假阳性错误) $\alpha$  值。 $\alpha$  值越小，样本量越大。4) 效力：效力(power)又称把握度( $1-\beta$ )， $\beta$  为检验假设时出现第 II 类错误的概率，而  $1-\beta$  为检验假设时能够避免假阴性的能力。若要求效力( $1-\beta$ )越大，即  $\beta$  值越小，则所需样本量越大。

(6)、基线资料的收集：收集每个研究对象在研究开始时的基本情况，包括待研究的暴露因素的暴露状况，疾病与健康状况，年龄、性别、职业、婚姻等个人状况，家庭环境、个人生活习惯及家族疾病史等。获取方式：①查阅医院、工厂、单位及个人健康保险的记录或档案；②访问研究对象或其它能够提供信息的人；③对研究对象进行体格检查和实验室检查；④环境调查与检测。

(7)、随访：1)、随访方法包括对研究对象的直接面对面访问、电话访问、自填问卷、定期体检，环境与疾病的监测等资料的收集等，应根据随访内容、随访对象、投入研究人力、物力等条件来考虑。在整个随访过程中，随访方法应保持不变。2)、观察终点 (end-point)就是指研究对象出现了预期的结果，达到了这个观察终点，就不再对该研究对象继续随访。3)、观察的终止时间是指整个研究工作截止的时间，也即预期可以得到结果的时间。4)、随访的

间隔与次数将视研究结局的变化速度、研究的人力、物力等条件而定。一般慢性病的随访间隔期为1~2年。

#### (8)、质量控制

3、资料整理分析：队列研究的最大优点就在于它可以直接计算出研究对象的结局的发生率，能直接计算出暴露组与对照组之间的率比和率差，即相对危险度 RR(relative risk)与归因危险度 AR(attributable risk)，从而可直接准确地评价暴露的效应。

(1)、相对危险度(RR)也叫危险比(risk ratio)或率比(rate ratio)，是反映暴露与发病(死亡)关联强度的最有用的指标。RR 表明暴露组发病或死亡的危险是非暴露组的多少倍。RR 值越大，表明暴露的效应越大，暴露与结局关联的强度越大。RR 算出的只是一个点估计值，是一个样本值。若要估计数值的总体范围，应考虑到抽样误差的存在，需计算其可信区间，通常用95%可信区间。常用的有 Woolf 法和 Miettinen 法。

(2) 归因危险度(AR) 又叫特异危险度、率差(rate difference, RD)和超额危险度(excess risk)，是暴露组发病率与对照组发病率相差的绝对值，它表示危险特异地归因于暴露因素的程度。RR 说明暴露者与非暴露者比较增加相应疾病危险的倍数；AR 则一般是对人群而言，暴露人群与非暴露人群比较，所增加的疾病发生数量，如果暴露因素消除，就可减少这个数量的疾病发生。前者具有病因学的意义，后者更具有疾病预防和公共卫生学上的意义。

(3) 归因危险度百分比(AR%)：又称为病因分值 EF(etiological fraction)，是指暴露人群中的发病或死亡归因于暴露的部分占全部发病或死亡的百分比。

(4) 人群归因危险度(population attributable risk, PAR)与人群归因危险度百分比(PAR%) 人群归因危险度百分比也叫人群病因分值(population etiologic fraction, PEF)。PAR 是指总人群发病率中归因于暴露的部分，而 PAR%是指 PAR 占总人群全部发病(或死亡)的百分比。RR 和 AR 都说明暴露的生物学效应，即暴露的致病作用有多大；而 PAR 和 PAR%则说明暴露对一个具体人群的危害程度，以及消除这个因素后可能使发病率或死亡率减少的程度，它既与 RR 和 AR 有关，又与人群中暴露者的比例有关。

#### 4、偏倚及其控制

常见偏倚的种类：(1) 选择偏倚(selection bias)：发生于研究对象中有人拒绝参加；历史性队列研究中有些人的档案丢失或记录不全；研究对象由较健康志愿者组成；早期病人研究开始时未发现、抽样方法不正确、执行不严格。(2) 失访偏倚(lost to follow-up)在一个较长的追踪观察期内，总会有对象迁移、外出、死于非终点疾病或拒绝继续参加观察而退出队列。失访率最好不超过 10%。(3) 信息偏倚(information bias) 在获取暴露、结局或其他信息时所出现的系统误差或偏差，又称为错分偏倚(misclassification bias)，如判断有病为无病，判断有无暴露等。由于仪器不精确、询问技巧不佳、检验技术不熟练、医生诊断水平不高或标准不明确、记录错误甚至造假等造成。若发生于一组而不发生于另一组，或两组错分的程度不同，则结果可能比实际的相对危险度高或低。前者称为非特异性错分，将后者称为特异性错分。(4) 混杂偏倚(confounding bias) 混杂是指所研究因素与结果的联系被其它外部因素所混淆，这个外部因素就叫混杂变量。它是疾病的一个危险因子，又与所研究的因素有联系，它在暴露组与对照组的分布是不均衡的。性别、年龄是最常见的混杂因素。

常见偏倚的预防：(1) 选择偏倚的防止：首先要有一个正确的抽样方法，即严格遵守随机化的原则；严格按照规定的标准选择对象；对象一旦选定，坚持随访到底。(2) 失访偏倚的防止：提高研究对象的依从性。失访率达到 20%以上，则研究的真实性值得怀疑。(3) 信息偏倚的防止：选择精确稳定的测量方法、调准仪器、严格实验操作规程、同等地对待每个研究对象、提高临床诊断技术、明确各项标准、严格按照规定执行、做好调查员培训是重要措施。(4) 混杂偏倚的防止：对研究对象作某种限制以获得同质的样本；在对照选择中采用匹配的办法，在研究对象抽样严格遵守随机化的原则等措施。