

队列研究 (COHORT STUDY)

西安交大一附院临床研究中心

郑婕 副主任

2015-10-18

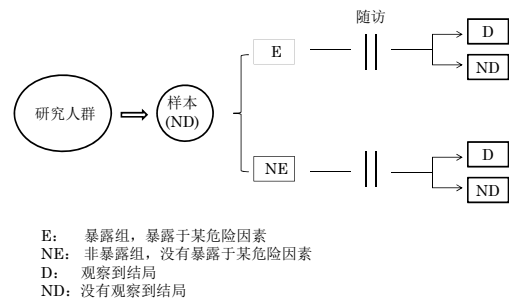
内容大纲

- ❖ 队列研究基本介绍
- ❖ 队列研究的过程
- ❖ 目前临床研究中需要注意的问题

队列研究

■ 队列研究 (cohort study) 属于观察性研究的一种，是临床研究的重要方法之一。将人群按是否暴露于某因素及其暴露程度分为不同的亚组，随访观察一定的时间，追踪其各自的结局，比较各组疾病结局的差异，从而判定暴露因子与结局之间有无因果关联及关联强度大小的一种观察性研究方法。

队列研究基本原理



队列研究特点

- 观察性研究
- 设立对照组
- 方向上是由因及果
- 能确诊暴露与结局的因果关系

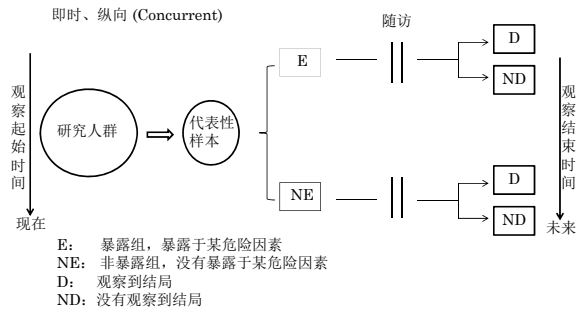
队列研究用途

- 病因学研究
- 预防、治疗和预后研究
- 描述疾病自然史

队列研究类型

- 前瞻性队列研究 (Prospective Cohort Study)
- 回顾性队列研究 (Retrospective Cohort Study)
- 双向性队列研究 (Ambispective Cohort Study)

前瞻性队列研究



前瞻性队列研究

■ 特点

- ✓ 研究队列的确定是现在 (concurrent)
- ✓ 根据研究对象现在的暴露分组
- ✓ 需要随访 (follow up)
- ✓ 结局在将来某时刻出现

前瞻性队列研究

■ 优点

- ✓ 时间顺序增强了病因推断的可信度
- ✓ 直接获得暴露与结局资料, 结果可信
- ✓ 前瞻性的, 设计相对完善, 事先不知道结局, 测量方法容易标准化, 减少信息偏倚
- ✓ 获得一种暴露与多种结局的关系

前瞻性队列研究

■ 缺点

- ✓ 不适于发病率很低的疾病的病因研究 (<5%)
- ✓ 易发生失访偏倚
- ✓ 耗时, 耗人力、物力、财力
- ✓ 设计要求严密, 资料的收集和分析难度较大
- ✓ 随访过程中, 已知变量的变化或未知变量的引入增加分析难度

队列研究 (1)

生命早期呼吸道合胞病毒 (RSV) 毛细支气管炎的研究,
RSV Branchiolitis in Early Life

课题类型: 前瞻性队列研究, 病因学研究

课题内容: 研究严重 RSV 毛细支气管炎后哮喘的影响因素, 建立能预测病后哮喘的模型。

研究人群: 1998 年到 2001 年间某医院一岁以内发生 RSV 严重感染的婴幼儿

队列研究 (1 - 续)

生命早期呼吸道合胞病毒 (RSV) 毛细支气管炎的研究
RSV Branchiolitis in Early Life

主要研究结局:

RSV 感染后是否发生哮喘, 测量指标: 医生诊断的哮喘

次要研究结局:

1. 活动性哮喘
2. 血清生物标记物的变化
3. 肺功能的变化

队列研究 (1- 续)

生命早期呼吸道合胞病毒 (RSV) 毛细支气管炎的研究
RSV Branchiolitis in Early Life

影响因素:

基线特征: 是否早产, 年龄, 性别, 种族, 出生体重, 身长

家庭情况: 母亲哮喘史, 过敏史, 吸烟史, 家庭收入

过敏源暴露史: 居住地尘埃过敏源分类定量测试, 收养宠物史

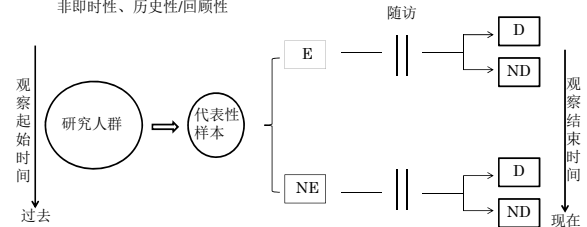
血液淋巴细胞百分比, IgE

三岁时皮肤过敏试验结果

观察时间长度: 12年

历史性队列研究

非即时性、历史性/回顾性



回顾性队列研究

■ 特点

- ✓ 研究队列的确定是过去的某一个时间点
- ✓ 根据研究对象过去某时刻的暴露情况分组
- ✓ 研究开始时, 暴露和结局均已发生

回顾性队列研究

■ 优点

研究所需的时间短, 耗费人力物力少

■ 缺点

暴露资料和疾病结局资料均为历史记录, 不受研究者控制, 真实性较差, 还常常因不能获得混杂因素的资料而影响资料的分析

队列研究 (2)

Intravenous Home Inotropic Use is Safe in Pediatric Patients Awaiting Transplantation

课题类型: **回顾性**队列研究, 预后研究

课题内容: 评价某种心脏移植前的治疗方法的安全性和疗效。

研究人群: 1999-2012年间某医院等待心脏移植的婴幼儿

队列研究 (2 - 续)

Intravenous Home Inotropic Use is Safe in Pediatric Patients Awaiting Transplantation

主要研究结局:

死亡率

次要研究结局:

1. 再次入院率
2. 严重并发症发生数、率

队列研究 (2- 续)

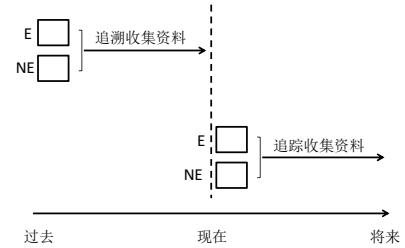
Intravenous Home Inotropic Use is Safe in Pediatric Patients Awaiting Transplantation

影响因素:

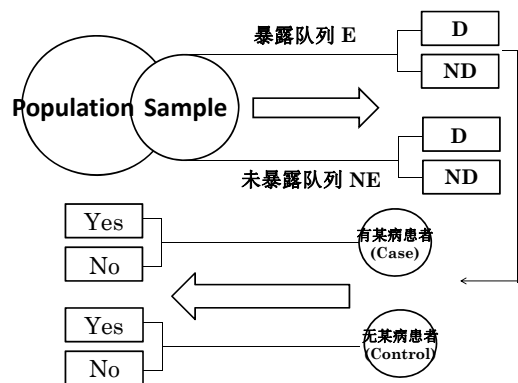
是否有先天性心脏病
是否有ECMO的支持
登记心脏移植时的心功能评级
心脏手术史
登记心脏移植年代

观察时间长度: 2年

双向性队列研究



巢式病例-对照研究 (NEST CASE-CONTROL STUDY)



巢式病例-对照研究

■ 优点

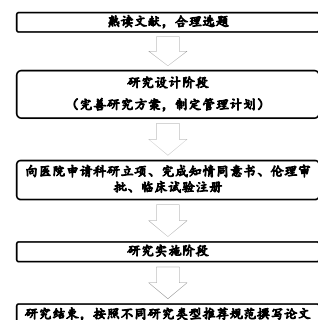
充分利用前瞻性研究收集的数据，随着生物样本的收集和储存方法的改进，这种设计越来越多的被用于大型研究

■ 缺点

研究感兴趣因素必须在已经收集好的数据里面

如何进行队列研究

临床研究的过程



研究方案 - 研究结局

■ 清楚定义所研究疾病 (Condition)

■ 确定研究结局 (Outcome)

明确研究结局的衡量指标

- 主要研究结局 (Primary Outcome, 1-2个)
- 次要研究结局 (Secondary Outcome, 多个)

结局的结局的测量: 客观、明确、严格

采用国际或国内通用的标准

研究方案 - 研究结局

■ 研究结局

- 发病、死亡、健康状况和生命质量
- 终极结果（发病）或中间结局（血清变化）
- 正面（生命延长）和负面（发病）
- 定性或定量
- 可多个或多方面

研究方案 - 影响因素

■ 确定影响因素

- **主要暴露因素:** 在描述性研究和病例对照研究的基础上确定。
- 人口学特征
- 混杂因素

研究方案 - 入排标准

■ 确定研究人群

- 从目标人群中抽出的具有代表性的人群
- 未患所研究疾病
- 分为暴露人群和非暴露人群

■ 对照组的选择

与暴露人群可比（除未暴露于所研究的因素外，其它各种因素或人群特征应尽可能与暴露人群相同）。

■ 研究类型的选择: 结合实际情况

研究方案 - 随访的设置

■ 随访的方法

1. 确定随访时间节点
2. 明确每一个时间节点上的任务
3. 随访起始时间: 理论上应在疾病的最短诱导期和潜伏期之后进行
4. 随访时间长短: 取决于暴露因素与疾病的联系强度
5. 随访的间隔: 根据人群中暴露因素的变化情况和疾病的最短潜隐期和最长潜隐期间的长短决定
6. 随访时间窗的确定

研究方案 - 随访方法

■ 随访的方法

- 1) 查阅记录和档案
- 2) 面谈
- 3) 信访
- 4) 电话访问
- 5) 环境检测

研究方案 - 估计样本量

■ 影响样本含量大小的因素

- 显著性水平 α (0.05 或 0.01)
- 把握度 Power = $1 - \beta$
- Effect size
例：两组不同人群均值的比较, 需要用两组测量指标的均数、标准差来计算
- 暴露组与对照组的比例
- 失访率

研究方案 - 估计样本量

■ 样本含量计算

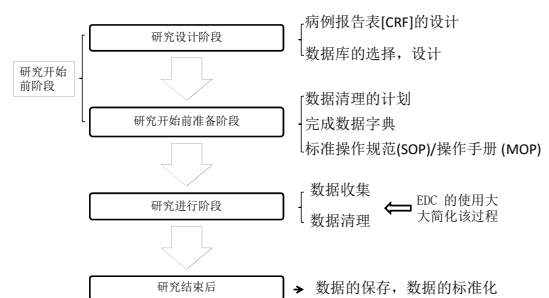
- 探索性质的研究不需要计算样本量
- 很多软件可以计算样本量：
G*power - 免费
PASS, SAS - 收费软件
- 查表
- 公式计算（条件：暴露组和对照组样本含量相等）

研究方案 - 数据管理

■ 数据收集管理方面的设计与计划

- 病例报告表的设计
- 数据收集方式的选择
 - web-based
 - PC-based (单机数据输入系统, 局域网数据输入系统)
- 数据管理的计划
 - 数据质量控制的计划 (data quality control plan)
 - 数据的更新与传送
- 数据库输入系统的设计
- 数据字典

数据管理的流程和工作内容



病例报告表的设计 - 设计原则

- 遵循研究方案
- 方便使用
- 避免可能的数据输入错误
- 信息化的角度
- 有利于统计分析

病例报告表的设计 - 设计细节举例

- 填表者使用方便
- 关键变量一定要定义明确
- 所有分类变量用代码代替具体分类名称
- 混淆因素变量的设置
- 尽量输入的是直接测量的数据 (c.g 身高, BMI)
- 每一页上都有研究对象相对应的唯一的研究代号
- 每一个模块需要有专门选项记录该表格的完成状态, 完成时间
- 每一个模块需要有专门选项记录谁填写的表格

研究方案 - 数据的清洗

■ 数据清洗的计划

- 确定执行数据清理的内容及频率
- 数据质量控制措施
 - ✓ 随机抽取被研究者的已输入数据和原始数据核对
 - ✓ 双数据输入(double data-entry)
 - ✓ 遗漏数据的核查
 - ✓ 数据的逻辑核查
 - ✓ 特殊数据的核查(随访时间窗, 实验数据稳定性检查)

■ 数据更新及传送

研究方案 - 统计分析

❖ 统计分析方面的计划

- 主要终点的分析方法
- 次要终点的分析方法
- 出现预计的影响统计分析的问题后, 备用的统计分析方法

研究方案 - 统计分析

❖ 按计划进行统计分析

❖ 常用统计方法

- 基线数据连续变量分析(*t-tests, ANOVA, linear regression, non-parametric Wilcoxon*)
- 基线数据分类变量分析(*Chi-square, logistic regression*)
- 纵向及重复检测数据分析(平均值, 斜率, *repeated ANOVA, MANOVA, mixed model*)
- 生存分析(*Kaplan-Meier plot, log rank test, Cox proportional hazard model, other parametrical analysis methods*)

常见偏倚

- 选择偏倚
- 失访偏倚
- 信息偏倚
- 混杂偏倚

■ 选择偏倚 (SELECTION BIAS)

研究人群在一些重要因素方面与一般人群或待研究的总体人群存在差异, 而导致研究结果的偏倚。

■ 产生原因

- 选择对象的方法不当
- 最初选定参加研究的对象中有人拒绝参加
- 历史性队列研究中部分档案丢失或记录不全
- 志愿者队列
- 研究开始时未能发现早期病人等

■ 控制

预防为主, 抽样方法正确, 严格按照规定标准选择对象, 提高研究对象的应答率和依从性

■ 失访偏倚 (LOST TO FOLLOW-UP)

研究对象因迁移、外出、死于非终点疾病或拒绝继续参加观察而退出队列所引起的偏倚。

本质上属于选择性偏倚

失访偏倚控制

■ 设计阶段

- ✓ 选择便于随访的人群
- ✓ 在计算的研究样本的基础上扩大10%

■ 实施阶段

- ✓ 加强对随访员的管理
- ✓ 制定随访计划和监测措施

■ 数据管理、分析方面

- ✓ 对失访对象进行补查
- ✓ 比较失访者和急需观察者的基本特征差异，以评估失访可能造成的影响

■ 信息偏倚 (INFORMATION BIAS)

在收集和整理有关暴露和疾病资料时所出现的系统误差。

e. g. 测量不精确，受试者回答问题不准确。

混杂偏倚 (CONFOUNDING)

- 暴露因素与疾病发生的相关程度受到其它因素的歪曲或干扰
- 是疾病结局的风险因素，但同时又与所研究的暴露因素有关。在暴露组与对照组中的分布不均衡，造成研究结局在两组中不可比。

临床研究 (3)

肾移植后免疫抑制药物治疗的选择：通过大型数据库分析治疗效果和药物相关并发症的发病率做决策

Choosing Immune Suppression in Renal Transplantation by Efficacy and Morbidity

课题特点：

1. 三个大型数据库联合，OPTN+PCD+Medicare，超大样本，>5万病例，专门数据团队研究建立了整合三个大数据库的策略。
2. 通过大型数据库了解全美国目前免疫抑制药物的使用情况
3. 特殊统计学方法分析数据
4. 药效+并发症情况+药物经济花费

临床研究 (3 - 续)

肾移植后免疫抑制药物治疗的选择：通过大型数据库分析治疗效果和药物相关并发症的发病率做决策

Choosing Immune Suppression in Renal Transplantation by Efficacy and Morbidity

主要研究结局：

死亡率

次要研究结局：

1. 移植后感染
2. 移植后新发糖尿病
3. 排斥反应
4. 癌症

临床研究 (3 – 续)

肾移植后免疫抑制药物治疗的选择：通过大型数据库分析治疗效果和药物相关并发症的发病率做决策

Choosing Immune Suppression in Renal Transplantation by Efficacy and Morbidity

药物治疗的混杂因素：

年龄、性别、种族

组织配型匹配情况

合并症史：糖尿病

疾病严重程度

器官捐献者情况等14种因素。

现实临床研究中要考虑的问题

❖ 研究设计

- 没有明确研究的首要研究临床终点，次要临床研究终点及相应的衡量指标
- 自己的研究内容适合什么样的研究形式
- 对照组或研究组选择合理不合理
- 我的研究到底需要多少病人？
- 混杂因素的排除
- 前期数据的积累

现实临床研究中要考虑的问题

❖ 数据管理

- CRF表格设计合理不合理
- 避免队列研究中病人失访
- 数据清理（数据丢失、数据错误）
- 研究结束后数据保存差，导致研究没有连续性

谢谢聆听！

西安交大一附院

临床研究中心

029-85323614

xjtuyfy_crc@163.com