

# 第二章 随机对照试验

彭继茂 (葛勤敏 译)

## 一、介绍

随机对照试验 (RCT) 是循证医学中最成熟的分支之一。高标准期刊采用标准化的清单表 (CONSORT 声明、Consolidated Standards of Reporting Trials 统一的试验报告标准规范), 以保证报告随机对照试验 (RCT) 的质量 (见 <http://www.consort-statement.org/>)。CONSORT 声明包括一个 25 项清单和一个流程图, 清单项目侧重于报告试验的设计、分析和解读。流程图显示了所有参与者在试验过程中的进度, 帮助我们区分受试者的计算按「符合方案数据分析」和「治疗意向分析法」。

## 二、RCT 的严格评价

当我们评价任何一篇临床文章时, 首先要看临床的指向性, 其次是有效性、结果和适用性。根据循证医学中心 (CEBM), “练习循证医学所需的基本技能之一, 是学习怎样建立一条完整可行的临床问题。为了方便患者和医生, 问题要与患者临床相关, 并能把问题用词组, 加以搜索后, 得到相关和精确的答案。

### (一) 临床指向性: PICO<sub>s</sub>/PEO<sub>s</sub> 模型

假设我们遇到了一个临床问题 (例如抗血小板药物如氯吡格雷, 对急性心肌梗死患者有额外的益处吗?) 并进行了文献检索。Lancet 中找到一篇文章“45852 例阿司匹林治疗的急性心肌梗死患者添加氯吡格雷: 随机安慰剂对照试验”。我们可以根据临床指向性 PICO<sub>s</sub> 模型 (见表 1) 制定一个可回答的临床问题, 及可被搜索的研究问题。P is patients/population 代表「患者/人群」。I is intervention 代表「干预」。C is comparison 代表「对照/比较」。O is outcome 代表「结果」。

表 1 临床指向性 PICO<sub>s</sub> 模型

	临床问题	研究问题
患者/人群 P	STEMI 患者	ACS 患者
干预 I	氯吡格雷+标准药物	氯吡格雷+标准药物
对照/比较 C	标准治疗	标准治疗

结果 0	死亡率和发病率	死亡、再梗死、出血、卒中
研究设计	RCT	RCT

PICOs 现在改为 PEOs, E 代表「暴露」。在上述案例中,「暴露」指氯吡格雷。

**患者:** 作者应明确定义。

纳入标准:

谁: 诊断标准、人口统计参数

地点: 数个研究中心或单一中心

时间: 时间框架

排除标准的设置是为了避免在数据分析中包含有问题的受试者。

弱势受试者或种族限制: 怀孕、囚犯

不适合获得知情同意的危重患者。(机构审查委员会/伦理委员会 可以免除知情同意)

低产率: 痴呆

“失访” 风险高: 游客

无法获得知情同意

**干预:** 预期在临床试验中得到有益结果的干预或治疗, 如在上述研究中增加氯吡格雷。

**对照:** 作为与治疗组比较的对照组。对照可以是安慰剂、标准治疗或甚至无治疗。

**结果:** 干预后的临床和功能状态: 用于评估干预的有效性 (治疗效果)。

可分为「主要结果」或「次要结果」。主要结果 (如死亡) 是研究最重要的部分, 也被用来计算样本量。次要结局是出现的 (预期或非预期) 结果, 以评价主要结局以外的治疗作用 (如并发症或住院时间)。

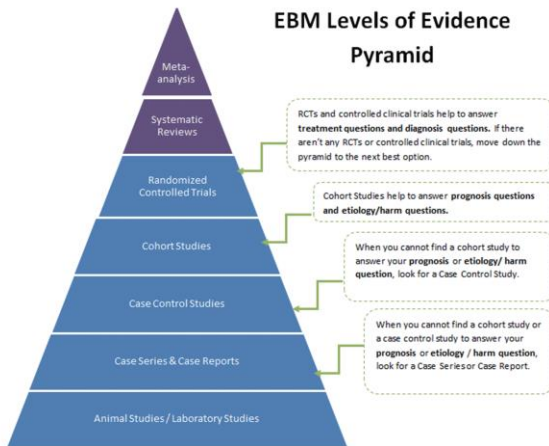
主要或次要结果可包括: 「综合结果」或「替代结果」。

**综合结果:** 将多个不同的独立结果合并或相加而成为一个单一结果, 如合并死亡、再梗塞或中风, 可算作一个综合疗效指标。一个受试者本身可能具备了 3 个独立的结果疗效指标。综合结果更容易具有统计学意义差异。但是, 若然独立地分析各自的结果, 可能就变成不具有统计学意义差异。

**替代结果:** 不具有直接实际意义, 但是反映结果测量。例如, 血压对患者不

直接重要，但在临床试验中经常用作替代临床结果，因为它是卒中和心脏病发作的风险因子。替代结果往往是生理或生化指标，检测相对快而简单，并被认为是重要临床结果的预测因子。

研究设计：研究类型，参考是根据文献的证据等级金字塔。



### 随机对照试验（RCT）：

将参与者随机分配（随机分组）到不同的治疗组后，比较两个或多个干预措施（可能包括对照组或不干预组）的实验。

证据级别最高，因为有效的随机化使已知和未知混杂因子在治疗组和对照组之间均匀分布。在 RCT 中也可采用盲法，减少观察者和反应者的偏倚。因此，随机化和盲法提高了研究的有效性。此外，RCT 完全控制了「暴露」- 这一独立因子(自变量)。

与前瞻性队列研究相比，队列研究缺乏随机分配。队列研究对独立因子「暴露」仅有有限操作(不能完全控制)。它可以控制已知的混杂因子，但不能控制未知的混杂因子。

然而，RCT 耗时而昂贵。它是以志愿者为基础的，存在不依从的问题。对研究微小差异的效果比较差，对罕见结果需要漫长的随访时间。此外，将受试者随机分组并参与「伤害有关的 RCT」是不道德的。因此，与伤害有关的试验通常采用病例对照研究（病例对照研究证据级别/有效性是低于 RCT）。

### （二）有效性（寻找偏倚）

有效性是探讨在研究或测量当中得出的结果, 其真值的程度、没有偏倚（系

统误差)的程度。

RCT有效性的要素：随机分配的过程、随访充分性、盲法和客观的结局测量。为方便记忆，可以概括为RAMBO（R是随机分配的方法和分配隐匿，A是确定随访充分性，MBO=测量必须盲法和客观）。Randomization/allocation concealment, Ascertainment, Measurement-Blinded and Objective.

### 1. 随机分配的方法与分配隐匿

总体人群是指研究问题中所有感兴趣的研究对象。样本是指总体人群的一小部分，选取代表整个总体人群，参加研究。选择/纳入过程称为抽样。

抽样之后，受试者进入随机分配。

随机分配是将受试者在对照试验内，被随机分配到的其中一个试验组。随机分配过程有两个组成部分：1 随机分配方法，和2 分配隐匿。

随机分配方法是指「随机序列」的产生及其应用，使得研究者不知道他所招募受试者的分配序列。随机分配可以有效地将已知和未知的混杂因子平均分配给两组。随机分配方法包括1. **简单**随机分配方法（例如，从帽子里抽出数字、掷硬币、使用随机表，或使用计算机生成的序列）或2. **限制性**随机分配方法（如区组随机分配方法或分层随机分配方法）。

分配隐匿：用于确保招募受试者的调查员，事前不知道受试者将被分配哪个组。这是为了减少「选择偏倚」。一些企图如何进行分配隐匿的尝试，比其他更容易被人为操纵。所以，怎样分配隐匿法被确立，成为被用来评价研究试验的质量。

2. 确认足够的追踪（寻找失访） 这是为了确定治疗组和对照组之间观察到的结局差异，是由两组接受的治疗差异所致。

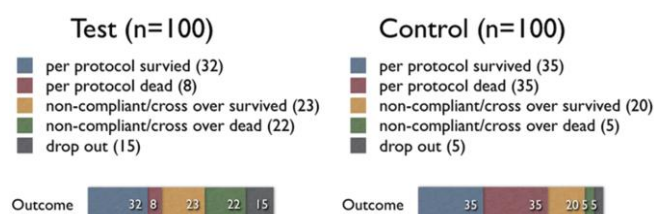
1) FU 充分性：受试者应被随访足够时间和完整。如果出现失访（LFU）或脱落，我们需要评估 LFU 的严重性。5%的脱落率很少引起偏倚，20%以上的脱落对效度构成严重影响。我们可以进行敏感度分析，在一定的概率范围内检验结论的稳定性。通常将 LFU 的百分比与效果大小进行比较； LFU%是否大于效果大小（治疗减去对照）。

2) 治疗意向分析（ITT）：随机分配后，即使受试者停止治疗（不随访）、从未接受治疗（不依从医嘱）或交叉治疗（对照组接受治疗），受试者应在原先被

随机分配的组中进行分析。这样在分析中，可以保持随机效应，从而偏倚小，有效性高。通常，所得 RR(相对结果)将小于通过「符合方案分析」计算得出的值。

3) 符合方案分析 (PP): 不依从的患者将不包括在分析中。将最终实际接受并完成治疗方案的患者，放入治疗组分析。因此计算 RR 的分母较小，得出的 RR 将大于 ITT。

4) 敏感度分析: 做病例最坏情景的计算，对于表现较好的组，假设所有失访者结局都较差；对于表现较差的组，假设所有失访者结局都较好。然后将此分析与原始分析 (ITT 或 PP) 进行比较，如果随访足够完整，且结果支持原始结论，我们可以得出一个强有力的结论，否则结论是不具说服力。



#### 死亡率作为主要结局

	试验 事件发生率	对照 事件发生率	ARR
PP 分析	8/40=20%	35/70=50%	50-20=30%
ITT 分析	30/85=35%	40/95=42%	42-35=7%
敏感度分析	45/100=45%	40/100=40%	40-45=-5%

#在敏感度分析（最坏情况）中，我们假设治疗组的所有脱落者都有不良的结局，而对照组的所有脱落者都有良好的结局。

从上表可以看出，在 PP 分析中 ARR 最大，ITT 分析中 ARR 较小，敏感性分析中 ARR 最小（甚至有负数，即有害影响）。因此，我们可以得出治疗组 15% 的脱落率是影向显著的，结论：该治疗结论是不具说服力。

假设治疗组脱落人数为 4 人（不是 15 人），我们将存活的人增加 6 人，死亡的人增加 5 人，然后再计算：

#### 死亡率作为主要结局

试验组事件发生率	对照组事件发生率	绝对风险下降率 ARR

	EER	CER	
PP 分析	13/51=25%	35/70=50%	50-25=25%
ITT 分析	35/96=36%	40/95=42%	42-36=6%
敏感度分析	39/100=39%	40/100=40%	40-39=1%

然后，EER 和 CER 之间仍有 1% 的差异，方向是有益的。因此，在治疗组中 4% 的脱落率无显著性影响。我们可以得出一个强有力的结论：该治疗具有良好的疗效。

通常，作者会在他们的研究中显示最大的差异值。他们可能没有做敏感度分析，以免导致结论说服力不强。

### 3. 结果的测量应该是盲法和客观

结果测量应该是客观的，即两个不同的数据收集人员对特定变量提供相同的检测。检测可分为：①生理测量（如血压、温度）；②自我报告（如疼痛评分、VAS 视觉类比量表、日记、问卷）：缺点是对感兴趣的变量可能检测不精确，通常取决于之前的「确认」；③观察（利用音频-视频）：只是记录下来、不提取，可用作重新检查。观察性研究中经常出现霍桑效应。

#### 1) 效应测量：

治疗组=治疗效果+随机误差+系统误差

安慰剂组=安慰剂效应+随机误差+系统误差

2) 随机误差：由于偶然性而产生的误差。置信区间和 p 值允许存在随机误差。

3) 系统误差：由于研究设计缺陷而产生的偏倚。只有通过更好的研究设计才能纠正。置信区间和 p 值不能纠正偏倚。

4) 安慰剂：指一种非活性步骤或给受试者服用非活性但外表相似物质/假药/伪药，通常是为了比较其与实际干预的效果。安慰剂在临床试验中用于对（医生和患者）设盲的治疗分配。

4. 盲法 盲法主要针对已经纳入受试者的治疗过程，不依赖随机化。盲法不同于分配隐藏法。盲法有助于减少观察者偏倚和应答偏倚，防止参与试验的人知道具体参与者所属的组别。当尽可能少的人知道谁在治疗组中、谁在对照组时，偏倚最小化。研究参与者、护理人员、结果评估者都可以被设盲。某些组无法做到一直被盲，如试验中的外科医生。应明确列出具体被盲法的人员，避免混淆。

当护理人员、参与者或结果评估者其中之一被盲时，称为单盲研究。当一组以上的人员被盲时，称为双盲研究。

### 三、结果

循证医学涉及定量试验。结果指标以数据形式呈现，数据可以大致分为两类：二分类数据和连续数据。只有原作者才有原始数据。原作者可以通过在二分类数据和连续数据之间交换来呈现原始数据集。例如，如果我们想比较温度的改变，我们可以在“绝对量的温度变化”（连续数据）和“大于或小于 38 摄氏度”（二分类数据）之间进行交换。

数据也可以分为绝对结果和相对结果。例如，“绝对风险下降率”和“需治疗人数”是绝对结果。“相对危险度减少率”、“相对危险度”，“比值比”是相对结果。

#### 1. 影响程度：点估计

二分法结果，可以用一个 2x2 表格/二分数据表来表示结果。

	事件（结果+）	非事件（结果-）
实验组	实验 事件（a）	实验 非事件（b）
对照组	对照 事件（c）	对照 非事件（d）

$$\text{实验事件率 (EER)} = a / (a+b)$$

$$\text{对照事件率 (CER)} = c / (c+d)$$

$$\text{相对危险度 (RR)} = \text{EER} / \text{CER}$$

$$\text{比值比 (OR)} = ad/bc$$

$$\text{相对危险度减少率 (RRR)} = (\text{CER} - \text{EER}) / \text{CER}$$

$$\text{绝对风险下降率 (ARR)} = \text{CER} - \text{EER}$$

$$\text{需治疗人数 (NNT)} = 1 / \text{ARR}$$

$$\text{得到有利结局的治疗成本} = \text{NNT} \times \text{每日成本} \times \text{治疗时间}$$

RRR 只显示相对值，因此看上去比 ARR 好得多。即使是一个非常小的 EER 和 CER，也能给出一个显著的 RRR。

RRR 和 ARR 仅表示临床意义，不代表统计学意义。

95% CI 或 P 值表示差异的统计意义。二分类数据采用卡方检验。

#### 2. 连续性数据 常用 student t 检验（未配对或配对）或单因素方差分析。

效应估计的精确度：区间估计。

3. 95%置信区间 如果在同一人群反复抽取的样本中多次重复同一研究，这些研究的 95%置信区间将包含未知量的真实值。区间大表示精确度较低（样本量较小），间隔小表示精确度较高（样本量较大）。95% CI=点估计 $\pm$ 1.96 SEM

Standard error of the mean (SEM) 平均值标准误差=Standard deviation/ $\sqrt{n}$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

4. P 值是某研究观察到的结果偶然发生的可能性，如果事实上「零假设」成立。即如果计算的  $p < 0.05$ ，也就是说，在研究中观察到的两组之间的差异是由偶然性引起的概率小于 5%，那么这个差异就被认为具有统计学意义。

P 值只说明具有统计学差异的一个临界点，而 95% CI 也能提示关于点估计的精确度和样本大小的充分性。因此，95% CI 应用更为广泛。

#### 5. 适用性

1) 5 Killer B 我们能把文章的结果应用到我们的临床情况或病人身上吗？有 5 个生物或社会经济障碍 (5 Killer B's)：①生物学：SCRAP，代表性别、共患病、种族、年龄和病理；②权衡：在临床利益和伤害/副作用之间进行权衡；③屏障：您所在地的可用性/易获得性；④信念：患者自己的信念/考虑；⑤负担：患病率和涉及的经济负担

2) 当地 NNT 当我们把文章的结果应用到自己病人身上时，需要评估我们能得到多少益处。假设相同的治疗效果和相同的相对危险度。当地 NNT=文章 NNT x (文章患病率/当地患病率)。例如文章中疾病的患病率为 0.1% (100/100 000)，NNT 为 30。我们对当地受试者应用文章中相同的相对危险度/治疗效果，如果当地患病率只有一半：0.05% (50/100 000)，那么当地 NNT 将为 60。这表明对于不同地区的罕见疾病，当地 NNT 更大，换句话说我们需要治疗更多的患者来得到一个结果。

#### 6. 附录

1) RCT 的样本量计算 样本量是为获得足够的统计检验能力（统计显著性）而招募的受试者的最小数量，它基于主要结果，应该在研究设计阶段进行计算（先验样本量计算），但在评价文章时可以用于交叉检查。



2) 决定样本大小的因素:

①研究类型 (如 RCT、队列、病例对照)

②显著概率阈值 (I 类错误,  $\alpha$ ): 数据中发现的关系因样本不凑巧而发生错误拒绝  $H_0$  的概率, 也称为 I 类错误 ( $\alpha$ ) 或假阳性。它在治疗性研究中更为严重 (如预期最低死亡率), 因此避免发生至关重要。

③研究的功率 (II 类错误,  $\beta$ ; 检验能力= $1-\beta$ ): II 类错误 ( $\beta$ ) 或假阴性错误, 是不拒绝实际不存在的零假设, 错误接受  $H_0$ 。统计检验功率 ( $1-\beta$ ) 是指试验适当拒绝  $H_0$  并证明治疗组和对照组之间差异的概率。对于任何给定的数据集,  $\alpha$ 、 $\beta$  都是反相关的: 一个发生的风险越小, 另一个发生的风险就越大。通常设  $\alpha=0.05$ ;  $\beta=0.2$  (检验功率=0.8)。

④单尾或双尾研究: 如果我们假设治疗组总是比对照组好/差, 我们可以使用单尾测试。如果我们不知道哪个组更好/更差, 我们可以使用双尾测试。双尾检验的样本量大于单尾检验。

⑤效应大小: 这是最小的临床上重要区别。效应值越大, 您越可能拒绝  $H_0$ 。如果效应值增加, 则样本量  $n$  将减小。

⑥平均值标准误差: 较大的样本量和较小的变异性 (标准差) 将产生较小的标准误。SEM= $\sigma/\sqrt{n}$  ( $\sigma$  是标准差)

3) 零假设  $H_0$  统计学假设是两组变量之间没有关联, 即两个群体之间没有差异。 $H_0: \Delta=0$ ; ( $\Delta$ : 试验组与对照组的区别)

4) 交替假设  $H_A$   $H_A: \Delta \neq 0$ 。如果 95% 置信区间不超过无差异值或  $P<0.05$ , 则拒绝接受  $H_0$ 。当接受  $H_A$  拒绝  $H_0$  时, 意味着一个变量与另一个变量之间存在关联 (即治疗组与对照组有差异)。

### CONSORT 2010 清单 随机对照试验报告标准规范

部分/主题	项目编号	清单项目	报告页码
标题和摘要	1a	在标题中确定为随机试验•	_____
	1b	试验设计、方法、结果和结论的结构化摘要 (有关具体指导, 请参阅摘要)。	_____

介绍	2a	科学背景与合理解释	_____
背景和目的	2b	具体目标或假设	_____
方法			
试验设计	3a	描述试验设计(如平行设计/析因设计), 包括分配入组比例	_____
	3b	试验开始后对试验方法所作的重要改变/原因	_____
受试者	4a	受试者合格标准	_____
	4b	资料收集的场所和地点	_____
干预措施	5	详细描述各组细节(2W)	_____
结果指标	6a	完整而确切地说明预先设定的主要和次要结局指标(何时、如何测评的)	_____
	6b	试验开始后对结局指标是否有任何更改/原因	_____
样本量	7a	如何确定样本量	_____
	7b	必要时, 解释中期分析和试验终止原则	_____
随机方法			
序列产生	8a	产生随机分配序列的方法	_____
	8b	随机方法的类型, 任何限定的细节(怎样分区组、各区组样本)	_____
分组隐匿	9	用于执行随机分配序列的机制(如序列编码封藏法), 描述任何干预实施前用于隐蔽随机序列方法	_____
实施	10	谁产生随机分配序列, 谁招募受试者, 谁给受试者分配干预措施	_____
盲法	11a	如使用了盲法, 分配干预措施后对谁设盲, 盲法是如何实施的	_____
	11b	如若相关, 描述每组干预的相似性	_____
统计学方法	12a	用于比较各组主要和次要结局指标的统计学方法	_____
	12b	附加分析的方法, 诸如亚组分析和校正分析	_____

<b>结果</b>			
受试者流 程（极力 推荐使 用）	13a	随机分配到各组的受试者例数，接受治疗例数，纳入主 要结局分析的例数	_____
	13b	随机分组后，各组脱落和被剔除的例数，并说明原因	_____
招募受试 者	14a	招募期和随访时间长短(具体)	_____
	14b	为什么试验中断或停止	_____
基线资料	15	用表格列出每组受试者的基线数据(人口学/临床特征)	_____
纳入分析 的例数	16	各组纳入每一种分析的受试者数目(分母)，是否按最初 分组	_____
结果和估 计值	17a	各组纳入每项主要和次要结局指标的结果，效应估计值 及精确性(如 95%可信区间)	_____
	17b	对于二分类结局，同时提供相对效应值和绝对效应值	_____
辅助分析	18	所作的其他分析结果，包括组分析和校正分析，指出哪 些是预先设定的分析，哪些是新尝试的分析	_____
危害	19	各组出现的所有严重危害或意外效果(报告方式见 CONSORT for harms)	_____
<b>讨论</b>			
局限性	20	试验局限性，阐述潜在偏倚、不精确性以及分析的多样 性（如果相关的话）	_____
	21	试验结果的一般性（外部有效性、适用性）	_____
	22	与结果一致的解释，权衡利弊，并考虑其他相关证据	_____
<b>其他信息</b>			
注册	23	注册号和试验注册人名	_____
方案	24	可以的话，如何获得完整的试验方案	_____
资金	25	资金来源和其他支助（如药品供应）、资助者的作用	_____