



罕见病临床试验设计的方法学进展

蔡思雨, 聂晓璐, 彭晓霞*

国家儿童医学中心, 首都医科大学附属北京儿童医院临床流行病与循证医学中心, 北京 100045

* 联系人, E-mail: pengxiaoxia@bch.com.cn

收稿日期: 2018-06-25; 接受日期: 2018-07-23; 网络版发表日期: 2018-09-07

“扬帆”计划重点医学专业(批准号: ZYLY201840)资助

摘要 罕见病的发病率低, 种类繁多, 且总受累人数众多. 然而, 由于罕见病相关研究可获得患者的样本量小、异质性大、难以招募等诸多因素, 难以按照常规研究方案开展, 导致许多罕见病缺乏有效的治疗手段, 而在其有限的治疗手段中, 也缺乏高质量的疗效与安全性评价证据的支持. 因此, 积极探讨适用于罕见病的临床试验设计对推动该领域的研究进展具有重要意义. 近年来, 方法学领域的不断突破, 新研究设计的不断诞生为推进罕见病临床试验研究提供了新的希望. 本文根据罕见病的特点, 有针对性地介绍4种研究设计方法, 为研究者选择适宜的研究设计提供参考.

关键词 罕见病, 临床试验, 方法学

目前, 罕见病在全球尚无统一且被广泛接受的定义. 欧盟将罕见病定义为患病率低于1/2000的疾病; 美国则定义为患者人数少于20万人的疾病^[1]. 罕见病发病率虽低, 却种类繁多, 因此罕见病患者的人数众多. 欧洲数据显示: 约6800余种疾病被归为罕见病, 受累人群约占总人口数的6%~8%^[2]. 罕见病常具有如下特征: (i) 常为遗传性疾病; (ii) 多于儿童时期发病, 疾病伴随终生; (iii) 病情通常比较严重, 致死致残率高, 生活质量差; (iv) 疾病诊断困难, 常见误诊、漏诊, 因此影响患者获得及时、准确的治疗, 严重影响患者生活质量, 给患者的家庭和社会都带来了巨大疾病负担.

然而, 大多数罕见病患者面临无药可医的困境. 研究报道仅有10%的罕见病具有有效的治疗手段^[3,4]. 迄今为止, 美国食品药品监督管理局共批准了360余种

罕见病治疗药物, 欧洲批准了60余种, 其他国家批准的药物很少^[5]. 截至目前, 中国尚无一种针对罕见病治疗的原研药上市, 而对于那些海外上市的罕见病治疗药物, 由于政策和价格等多种因素, 也极少能够进入中国投入临床使用^[5]. Tudur Smith研究组^[6]对Clinical Trials(www.clinicaltrials.gov)上登记的罕见病临床试验进行分析, 结果显示相对于非罕见病注册的临床试验, 更多罕见病临床试验是单臂、非随机、非盲法和开放标签的, 且相比非罕见病注册的临床试验, 罕见病临床试验更易出现提前终止. 罕见病临床试验受限的主要原因有: (i) 单中心可获得罕见病患者数量有限, 使得经典随机对照试验缺乏可行性; (ii) 大约2/3的罕见病累及儿童, 进一步限制了潜在临床试验参与者的可获得性^[7].

引用格式: 蔡思雨, 聂晓璐, 彭晓霞. 罕见病临床试验设计的方法学进展. 中国科学: 生命科学, 2018, 48: 937-942

Cai S Y, Nie X L, Peng X X. Methodological progress in the design of clinical trials for rare diseases (in Chinese). Sci Sin Vitae, 2018, 48: 937-942, doi: 10.1360/N052018-00089

虽然开展罕见病临床试验面临重重困难, 但研究人员仍然应该致力于提供高质量的证据。虽然经典随机对照临床试验设计在评价罕见病药物疗效时可行性不高, 但是随机、盲法和对照这些设计要素, 仍可以通过其他巧妙的方法学设计得以实现^[8]。近年来, 研究者们相继提出了各种适用于罕见病药物疗效评价的研究设计, 为推进罕见病疗效评价研究提供了新的希望。本文根据罕见病的特点, 有针对性地介绍4种研究设计方法, 为研究者选择适宜的研究设计方案提供参考。

1 单病例随机对照试验

单病例随机对照试验(n-of-1 trial)是以单个病例自身作为对照, 对单个病例进行双盲、随机、多次交叉的试验设计, 从而比较某种药物与安慰剂或阳性对照药物疗效。研究目的是通过反复对同一病例进行多次交叉对照, 观察患者对某药物以及对照药物的反应, 从而为个体患者寻找最优治疗方案。研究设计为, 根据疾病特点设置三轮或以上的试验周期, 每一轮试验周期包括一个使用试验药物的观察期和一个使用对照药物的观察期, 每轮采用随机的方法来确定是先接受试验药物还是对照药物, 每期结束后设置洗脱期来消除研究干预措施的残余影响^[9]。当试验数据能充分表明试验药物对事先制定的目标结局是否有作用时, 则可终止试验。其设计模式见图1。

单病例随机对照试验的应用条件为: (i) 待评价干预措施起效快、半衰期短; (ii) 所研究疾病病情较为稳定; (iii) 试验药物与对照药物间无相互作用和延滞效应(carry over effect); (iv) 患者个体差异较大^[10]。单病例随机对照试验适用于罕见病临床试验的原因在于: 可应用最少的样本量, 花费较少人财物力, 短时间内从多种治疗中选出最有效者; 此外, 患者会先后接受

试验药物与对照药物, 因此失访率低。但是, 单病例随机对照试验具有一定的局限性, 试验中疾病病情的变化可使试验期和对照期的疾病相关特征基线不可比, 从而影响随机效果, 导致研究效应缺乏可比性; 而交叉性质的试验设计常常使研究者无法观测治疗的远期疗效与远期副作用; 较小样本量则可能导致研究结论外推性差, 出现II类错误的概率增大^[11~13]。

单病例随机对照试验的样本量概念与经典随机对照临床试验有所不同, 没有共识的计算公式。根据国内外发表文献, 单个患者的单病例随机对照试验的轮次数多为3次, 研究者在估算样本量时可通过参考文献或经验资料获得每轮次内需要测量疗效评价指标的轮次数来确定样本量; 针对系列病例开展单病例随机对照试验时, 样本量多为临床实际能够招募到的受试者数, 估算系列单病例随机对照试验的样本量时, 可根据预试验得出治疗措施间的平均差异, 先计算出RCT所需要的样本量大小, 继而根据模拟出的单病例随机对照中多个轮次的完成度计算样本量^[14]。

对于单病例随机对照试验结果的统计分析, 常采用统计图描述疗效差异或进行统计学假设检验。其中, 统计图往往能较为直观地显示数据分析的结果, 但无法给出统计学假设检验结果。单个病例的单病例随机对照试验中的数据, 是包含时间序列的试验数据集, 可使用动态回归模型进行统计分析。系列病例的单病例随机对照试验中, 研究人员需要分析整体的平均效果, 并根据其他参与者的结果对每个人的结果进行修正, 可使用分层贝叶斯模型。当数据出现自相关性和缺失问题时, 可使用混合效应模型^[15,16]。

案例^[17]: 重症肌无力是一种罕见的神经肌肉疾病, 最初通常使用乙酰胆碱酯酶抑制剂进行治疗。但某些病人使用乙酰胆碱酯酶抑制剂疗效不佳, 则使用皮质类固醇或其他免疫抑制药物, 但这些药物可能有严重

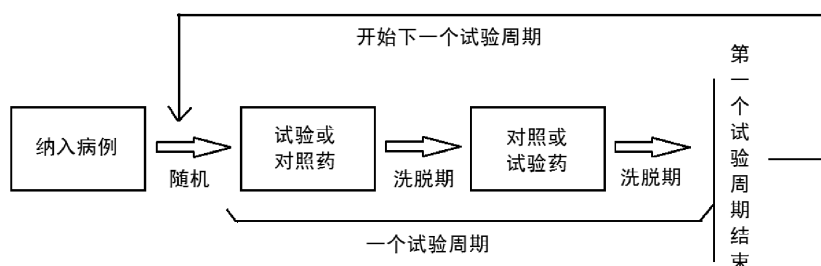


图1 单病例随机对照试验设计示意图

的副作用. 临床观察表明, 乙酰胆碱酯酶抑制剂或小剂量强的松联合使用麻黄碱, 可以推迟, 甚至无需使用免疫抑制治疗. 重症肌无力是一种罕见的病症, 且仅有部分患者需要使用麻黄碱, 所以大样本的随机对照试验难以开展. 因此, 该研究采用了单中心、安慰剂对照、双盲、随机的单病例随机对照试验. 研究共纳入 4 名患者. 1 次试验包括 3 个周期, 每个周期为期 2 周, 包括一个麻黄碱治疗期和一个安慰剂治疗期, 治疗顺序是随机分配决定的. 每周的周一至周五为服药期(服用麻黄碱或安慰剂), 周六至周日为洗脱期. 在 6 周的多重交叉阶段后, 是 4 周的评估阶段, 在此期间, 对多个交叉阶段的数据进行分析, 将结果反馈给患者和他们的治疗医生, 并对试验进行评估.

2 随机撤药设计

随机撤药设计(randomised withdrawal design)指所有符合纳入标准的患者全部接受治疗, 以鉴别能够从治疗中受益的患者亚群. 然后将该亚组中的患者随机分为试验组与对照组, 对照组撤药后, 对两组疗效进行比较. 其设计模式见图 2.

随机撤药设计的应用条件为: (i) 所研究疾病为非自限性疾病, 病情较为稳定. (ii) 撤药不会造成严重的伤害. 该设计的优势在于将治疗无反应的患者识别并排除出试验, 可以将无效治疗的时间最小化. 同时, 可以及时评估治疗是否需要继续或停止. 由于撤药产生的效应值较大, 因此, 可以最大程度节约样本量. 该设计缺陷为, 在分析治疗效果时纳入的人群为治疗有效的人群, 因此与对照组比较时治疗效果可能被高估.

随机撤药设计的样本量计算一般是在差异性假设检验的前提下进行. 以比较两组结局事件发生率为例.

n_c : 对照组样本量; π_t : 试验组结局事件发生率; π_c : 对照组结局事件发生率; k : 试验组例数与对照组例数的比例; $u_{1-\alpha}$ 和 $u_{1-\beta}$ 表示标准正态分布中对应 $1-\alpha$ 和 $1-\beta$ 的百分位数, 一般取常数为 1.96 与 0.84. 对应的样本量计算公式为:

$$n_c = \frac{[\pi_t(1-\pi_t)/k + \pi_c(1-\pi_c)](u_{1-\alpha} + u_{1-\beta})^2}{(\pi_t - \pi_c)^2}.$$

经以上公式计算获得的样本量, 除以预期估计的治疗有效率, 即为研究需要纳入研究对象的总数.

随机撤药设计的统计分析与经典随机对照临床试验一致, 以主要结局指标为二分类变量为例, 分别计算两组主要结局事件发生率, 对两组率的差异进行卡方检验, 并计算风险比值(risk ratio, RR)及其 95% 可信区间, 当 95% 可信区间不包含 1 时, 差异有统计学意义.

案例^[18]: 此研究的目的是评估缓释盐酸胍法辛(guanfacine hydrochloride, GXR)治疗注意力缺陷/多动症儿童的远期疗效. 该研究为第 III 期, 双盲, 安慰剂对照, 随机撤药临床试验. 研究共纳入 528 名患者, 患者在第一阶段接受开放标签的 GXR 治疗. 13 周后, 治疗有效者进入第二阶段的随机撤药阶段. 共 316 人(59.8%)进入随机撤药阶段. 患者以 1:1 的比例被随机分配到 GXR 组或安慰剂组中. 研究的主要结局指标为治疗失败的百分比.

3 单组目标值

单组目标值法(single-arm objective performance criteria), 指根据前期研究或行业内广泛认可的效应指标值, 制定本次研究预期获得效应指标值的大小, 将其作为在不设立对照组情况下, 试验组至少应取得的目的.

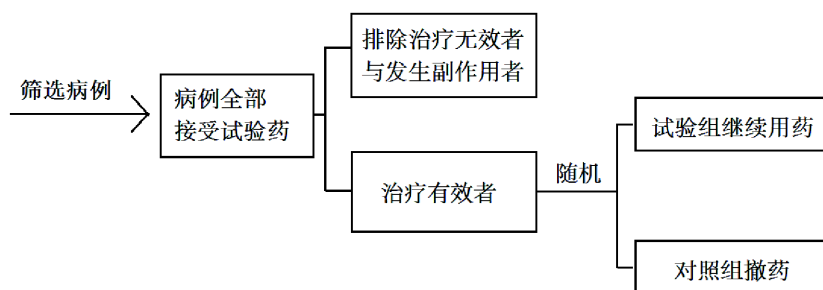


图 2 随机撤药设计示意图

标试验效应. 研究设计的步骤为: (i) 根据专业知识确定诊疗方案的有效性或安全性相关效应指标; (ii) 根据历史数据设定该效应指标的目标值; (iii) 估算样本量; (iv) 所有受试者均接受试验治疗, 观察其有效性和安全性; (v) 数据统计分析.

单组目标值法的应用条件为: 出于伦理学考虑, 设立同期对照组存在困难时. 该设计方法的优势在于可以避免一些伦理学风险, 同时减少对样本量的需求. 然而, 单组试验目标值法难以控制选择性偏倚, 而且由于缺乏同期平行对照, 其研究结果的因果论证强度远低于对照试验, 尤其是应用于自限性疾病的临床治疗时.

单组目标值法的样本量估计所需的参数包括: 主要评价指标的总体参数、目标值、I类错误率 α 、II类错误率 β ($1-\beta$ 为检验效能). 以均数样本量计算为例. μ_1 : 被试产品主要评价指标总体均数; μ_0 : 被试产品主要评价指标均数的目标值; σ : 主要评价指标的预期标准差. $Z_{1-\alpha}$ 和 $Z_{1-\beta}$ 为标准正态分位数. 对应的样本量计算公式为

$$n = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2 \sigma^2}{(\mu_1 - \mu_0)^2}.$$

单组目标值试验的统计学推断有两种方法. (i) 假设检验: 单组目标值法相应的假设检验为单侧检验. 当 $P \leq \alpha$ 时, 拒绝 H_0 , 认为试验产品达到设计要求. (ii) 置信区间法: 当评价指标为高优指标时, 评价指标 $(1-2\alpha)\%$ 双侧置信区间的下限高于目标值, 则认为被试产品达到设计要求. 相反, 评价指标为低优指标时, 评价指标 $(1-2\alpha)\%$ 双侧置信区间的上限低于目标值, 则认为被试产品达到设计要求^[19].

案例^[20]: 研究的目的是为评估Minerva子宫内膜消融术治疗绝经前妇女过度子宫出血的安全性和有效性. 该研究的设计为多中心单组目标值研究. 研究共纳入105名患者. 目标值基于美国食品和药物管理局报告的所有批准的子宫内膜消融术的成功率而制定. 研究所确定的目标值为: 治疗成功率为75.6% (95% CI: 65.6%~83.5%). 有效性分析旨在确定Minerva子宫内膜消融术的成功率是否等于或高于目标值的95%置信下限. 研究结果显示, 基于Minerva子宫内膜消融术的治疗成功率为96.2%, 优于目标值置信下限66% ($P < 0.0001$).

4 成组序贯设计

传统序贯设计(group sequential design)指将每一对研究参与者随机分配至试验组和对照组, 在这一对完成试验后即进行一次统计分析, 判定试验药物是否可以得到有效/无效的结论, 从而确定是否终止研究.

传统序贯设计可以在确定试验组与对照组相比确实存在疗效差异的情况下, 及时结束临床试验, 从而节约样本量^[21]. 但是, 传统序贯设计需要在每一对受试者完成试验进行一次统计分析, 会因重复假设检验导致I类错误发生概率风险增加. 成组序贯设计则在传统序贯设计上进行了一定改进, 在完成一定比例的样本量之后对累积试验数据进行期中分析, 以判定试验是否可以提前得到有效/无效结论而提前结束试验^[22]. 其设计模式见图3.

成组序贯设计的应用条件为必须预先制订成组序贯设计方案, 并严格按方案执行研究. 成组序贯设计方法的优势主要体现在: 当试验药物明显优于或劣于对照药物时, 成组序贯设计可以在期中分析时提前得出结论, 节约样本量并缩短试验周期. 这样做使得有效药物可以尽早上市, 而无效药物可以在试验中尽早淘汰, 避免更多的受试者仍然暴露于无效治疗之中. 但成组序贯设计可能会因多次期中评估而揭盲数据, 需要特殊统计分析方法来进行调整和控制.

成组序贯设计的样本量估计所需的参数包括: I类错误率 α , II类错误率 β ($1-\beta$ 为检验效能)、组间均数或率之间的差值 δ 以及标准差 σ . 每次期中分析的样本量为

$$2n = \theta \times (\sigma / \delta)^2.$$

在成组序贯设计中, 还要考虑预期进行几次期中分析. 整个试验进行 K 次分析, 则试验所需要的总样本量为 $(2n)K$. 通常情况下, 试验停止的越早, 节约的样本量越多.

成组序贯设计期中分析常见的两种方法为Pocock法和O'Brien-Fleming法. 期中分析的关键在于, 首先必

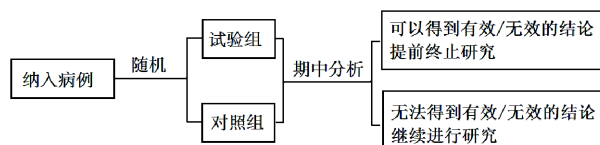


图3 成组序贯设计示意图

表 1 各项研究方法优缺点与适用范围

研究设计	主要特征	随机	应用条件	优点	缺点
单病例随机对照试验 (n-of-1 trial)	研究仅纳入一个患者, 为患者随机分配设计旨在对一个个体评估几种治疗的效果	治疗和安慰剂的顺序	①预措施起效快、半衰期短; ②所研究疾病为非自限性疾病, 病情较为稳定; ③所研究治疗间无相互作用和延滞效应; ④疾病个体差异较大	①所需样本量小; ②研究短时间内从多种治疗中选出最有效者, 患者从试验中可直接获益; ③失访率低	①试验中疾病的病情变化可使试验期和对照期基线不一致; ②无法观测到治疗的远期疗效与副作用; ③外推性差
随机撤药设计(randomised withdrawal design)	用于评估治疗有效患者的治疗持续效应. 所有患者最初都接受治疗, 随后治疗有效者被随机分配到试验组或安慰剂组	治疗有效者被随机分配到试验组或安慰剂组	①所研究疾病为非自限性疾病, 病情较为稳定; ②撤药不会造成严重的伤害	①无效治疗的时间最小化; ②及时评估治疗是否需要继续或停止; ③节约样本量	分析治疗效果时纳入的人群为治疗有效的人群, 因此与对照组比较时治疗效果可能被高估
单组目标值(objective performance criteria)	无随机和对照, 疗效指标与设定的目标值进行比较	无	设立同期对照组存在挑战时	①可以避免一些伦理学风险; ②节约了样本量	①控制选择性偏倚; ②研究结果的因果论证强度远低于对照试验
成组序贯设计(group sequential design)	期中分析后判定试验是否可以提前得到有效/无效的结论	为每一组患者随机分配治疗和安慰剂	必须预先制订成组序贯设计方案, 并严格按方案执行研究	①节约样本量; ②缩短试验周期	会因多次期中评估而揭盲数据, 需要特殊统计分析方法来进行调整和控制

须明确试验分析的结局指标; 其次必须建立停止试验的规则^[23].

案例^[24]: 此研究为一项多中心, 双盲, 安慰剂对照的成组序贯设计研究, 评估哌醋甲酯治疗躁狂症的疗效和安全性. 研究的主要结局指标为青年躁狂评分量表总分的变化. 当研究纳入42名患者后, 期中分析结果显示两组患者青年躁狂评分量表的变化没有显著差异($P=0.64$), 于是提前终止该研究.

各项研究方法优缺点与适用范围见表1.

罕见病研究不仅为患者提供更多获得治疗机会, 也可以评估治疗的效果, 使有效的治疗更有可能用于罕见疾病患者. 但是, 关于罕见疾病的流行病学研究和临床研究证据有限, 导致临床实践中的诊疗决策常

在缺乏证据的前提下制定. 介绍以上适用于罕见病临床试验开展的方法学设计, 可以帮助研究人员为特定的罕见疾病选择适宜的研究方案. 罕见病临床试验设计更偏重于节约样本量和患者利益最大化, 但是临床试验的目标仍是尽可能遵循临床试验设计的一般标准, 即设立同期对照、随机分配、盲法的实施与评价, 从而保证足够的统计学功效以及偏倚最小化, 产生低偏倚风险的临床研究证据. 罕见病临床试验方法学的发展会极大地推动罕见病疗效评价相关研究, 及时、有效、低风险地评估治疗, 并优化医疗资源的在未来治疗中的利用和发展. 因此, 研究者可以了解上述适用于罕见病的临床试验设计方法, 并将其应用到未来的研究中.

参考文献

1 Bashaw E D, Huang S M, Coté T R, et al. Clinical pharmacology as a cornerstone of orphan drug development. *Nat Rev Drug Discov*, 2011, 10: 795–796

2 EURORDIS Rare Diseases Europe. “Rare diseases: understanding this public health priority www.eurordis.org”

3 Tambuyzer E. Rare diseases, orphan drugs and their regulation: questions and misconceptions. *Nat Rev Drug Discov*, 2010, 9: 921–929

4 Heemstra HE. From research on rare diseases to new orphan drug development. Doctor Dissertation. Utrecht: Utrecht University, 2010

5 马端, 李定国, 张学, 等. 中国罕见病防治的机遇与挑战. *中国循证儿科杂志*, 2011, 6: 81–82

6 Bell S A, Tudur Smith C. A comparison of interventional clinical trials in rare versus non-rare diseases: an analysis of ClinicalTrials.gov. *Orphanet J Rare Dis*, 2014, 9: 170

- 7 Braun J, Day P K, Toennies J P, et al. Design and analysis of cross-over trials. Chapman & Hall/CRC, 2003, 27: 464–490
- 8 Day S. Evidence-based medicine and rare diseases. *Adv Exp Med Biol*, 2010, 686: 41–53
- 9 Guyatt G, Sackett D, Adachi J, et al. A clinician's guide for conducting randomized trials in individual patients. *CMAJ*, 1988, 139: 497–503
- 10 Wulff H R. Single case studies. An introduction. *Scand J Gastroenterol Suppl*, 1988, 147: 7–10
- 11 李江, 田金徽, 杨克虎. 单病例随机对照试验研究现状. *中国循证儿科杂志*, 2011, 6: 453–455
- 12 刘建平. 单个病例随机对照试验的设计与应用. *中国中西医结合杂志*, 2005, 25: 252–254
- 13 Korn E L, McShane L M, Freidlin B. Statistical challenges in the evaluation of treatments for small patient populations. *Sci Transl Med*, 2013, 5: 178sr3
- 14 Nikles J, Mitchell G, McKinlay L, et al. A series of n-of-1 trials of stimulants in brain injured children. *Neurorehabilitation*, 2017, 40: 11–21
- 15 翟静波, 商洪才, 李江, 等. 单病例随机对照试验的统计分析方法. *中国循证医学杂志*, 2017, (4): 494–496
- 16 井含光, 孟庆刚. 单病例随机对照试验在中医药临床疗效评价中的应用与要点分析. *中华中医药学刊*, 2018, (4): 890–893
- 17 Vrinten C, Lipka A F, van Zwet E W, et al. Ephedrine as add-on therapy for patients with myasthenia gravis: protocol for a series of randomised, placebo-controlled n-of-1 trials. *BMJ Open*, 2015, 5: e007863
- 18 Newcorn J H, Harpin V, Huss M, et al. Extended-release guanfacine hydrochloride in 6-17-year olds with ADHD: a randomised-withdrawal maintenance of efficacy study. *J Child Psychol Psychiatr*, 2016, 57: 717–728
- 19 中国临床试验生物统计学组. 单组目标值临床试验的统计学考虑. *中国卫生统计*, 2017, 34: 505–508
- 20 Laberge P, Garza-Leal J, Fortin C, et al. One-year follow-up results of a multicenter, single-arm, objective performance criteria-controlled international clinical study of the safety and efficacy of the minerva endometrial ablation system. *J Minim Invasive Gynecol*, 2015, 22: 1169–1177
- 21 Demets D L. Sequential designs in clinical trials. *Cardiac Electrophysiol Rev*, 1998, 2: 57–60
- 22 何清波, 苏炳华. 成组序贯试验在临床试验中的应用. *中国临床药理学与治疗学*, 1998, 3: 130–133
- 23 李先涛, 梁伟雄, 张伯礼. 简单介绍成组序贯设计在临床试验中的应用. *中国临床药理学与治疗学*, 2006, 11: 1433–1436
- 24 Hegerl U, Mergl R, Sander C, et al. A multi-centre, randomised, double-blind, placebo-controlled clinical trial of methylphenidate in the initial treatment of acute mania (MEMAP study). *Eur Neuropsychopharmacol*, 2018, 28: 185–194

Methodological progress in the design of clinical trials for rare diseases

CAI SiYu, NIE XiaoLu & PENG XiaoXia

Center for Clinical Epidemiology and Evidence-based Medicine, Beijing Children's Hospital, Beijing 100045, China

Rare diseases have a low morbidity, include a wide variety of diseases, and involve a large number of people. However, because of the small sample size, high heterogeneity, and difficulty in recruiting, the research of rare diseases is difficult to carry out. This dilemma results in the lack of effective treatments for many rare diseases and the lack of high-quality efficacy and safety evidence of limited treatments. Therefore, it is of great significance to explore the design methods of clinical trials of rare disease. Recently, the breakthrough in methodology and the generation of new design methods provide a new clue for the clinical trials of rare diseases. The purpose of this paper is to introduce the innovative methods suitable for the clinical trials of rare diseases. We introduce four design methods according to their characteristics for researchers' reference.

rare disease, clinical trials, methodology

doi: [10.1360/N052018-00089](https://doi.org/10.1360/N052018-00089)