



北京大学
PEKING UNIVERSITY



ICH E17 落地实施蓝皮书

北京大学

中国外商投资企业协会药品研制和开发工作委员会

2023年12月

目录

第一章 背景和目的	2
第一节 ICH E17	2
第二节 ICH E17 落地实施的总体挑战	3
第三节 主要挑战及应对策略	5
第二章 研究设计阶段考量	8
第一节 一般考量	8
第二节 特殊考量	12
第三章 研究执行阶段考量	14
第四章 研究结果解释评价	15
第一节 一致性结果解释考量	15
第二节 不一致性探索框架	18
第五章 特殊考量	22
参考文献	25
附录：蓝皮书涵盖的相关方法学技术细节	27

第一章 背景和目的

第一节 ICH E17

随着药物研发日益全球化，不同区域和国家监管机构能够接受多区域临床试验（Multi-Regional Clinical Trials, MRCT）数据作为支持药物（医药产品）批准上市的主要证据来源已经变得很重要。但在传统的药物研发模式中，为了符合各个监管要求，通常在不同监管区域进行各自的临床试验，这种做法不仅重复进行了临床试验，增加了额外的成本，而且导致一些国家在药物上市方面产生不必要的延误。

面对这一挑战，国际人用药品注册技术协调会（The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, ICH）于1998年就发布了ICH E5指导原则《接受国外临床试验数据的种族因素》，推荐一个用于评估种族因素对药物疗效的影响的框架，尽可能减少重复临床研究，为患者提供药物使其获益的同时，又对种族因素的影响进行了充分的评估，从而帮助药品在国际协调会议（ICH）地区注册。在该指南中种族因素被定义为人群中与遗传和生理因素（内因）、以及文化和环境（外因）特征有关的因素。虽然ICH E5指导原则的意图明确，但实际应用中，仍有许多临床试验在不同国家独立进行，导致新药研发成本的持续升高和不同国家上市的延迟。

为了应对这一挑战，ICH于2017年发布了E17指导原则《多区域临床试验计划与设计的一般原则》，旨在说明MRCT计划与设计的一般原则，目的是提高MRCT在全球监管递交中的可接受度。在与其他ICH指导原则（包括E5等）的使用中，实现希望能从地域思维走向全球思维，从回顾性策略走向前瞻性策略，从桥接试验走向全球同步研发。

ICH E17详细介绍了多区域临床试验（multi-regional clinical trial, MRCT）设计的7项基本原则，即①鼓励药物研发过程中使用MRCT，但是需要注意不同区域间的差异；②提前识别可能影响药物研发过程的内在和外在因素，并在探索性和确证性临床试验中对相关因素进行收集；③整体疗效估计和区域样本量分配方法；④基于已知的相似性，预设区域或亚群合并策略；⑤基于结构化探索，评价整体和亚群间的一致性；⑥基于ICH E6指导原则对MRCT进行高质量的设计和和实施；⑦鼓励申办者与监管机构进行充分且有效地交流。可基于ICH E17提供的培训材料和一些研究者成果，结合ICH E5，对ICH E17基本原则进一步理解^[1]。

随着中国新药研发和临床研究水平的显著提升，国内越来越多的临床试验机构开始参与国际多中心临床试验MRCT。这一发展趋势不仅有效地缩短了国内外药品上市的延迟时间，提高了国内患者对新药的可及性，而且有助于中国医疗机构和研究人员获得全球药物开发的最新知识和技术，对于更深入地探索和分析新药在中国人群中的疗效和安全性具有重要意义。

尽管中国在MRCT中的参与度提高并带来了诸多好处，但在全球范围内执行ICH E17指导原则时，仍面临一系列挑战。首先，不同国家监管机构对临床试验的要求存在差异，增加了全球临床试验设计和实施的复杂性。例如，各国可能对药物的疗效评估和安全性监测有不同的标准和要求，导致数据收集和分析上的差异。其次，全球范围内的质量控制也是一个重要挑战，因为不同国家和地区在医疗基础设施和资源方面的差异可能影响临床试验的质量和结果的可靠性。此外，尽管ICH E17指导原则提供了基本原则和框架，但在具体的操作和实施过程中仍存在许多未解决的问题，如全球范围内试验药物的试验设计、试验同步启动、有效性和安全性一致性评价、合并策略、区域样本量分配等。面对这些问题和挑战，急需相应的行业共识来对相应设计、执行、结果解释过程中的细节进行指导。

第二节 ICH E17 落地实施的总体挑战

目前，基于 MRCT 的全球同步研发已成为新药研发中最快速、最有效的方法之一。通过参与 MRCT，可以显著加速某些区域患者对创新药物的可及性，同时增强该区域研究者在全球医药领域的影响力。然而，MRCT 的实施同时伴随着一系列独特的挑战。这些挑战不仅影响了 MRCT 的设计和统计分析，还涉及到各区域之间在临床诊断和实践上的差异、临床试验运营管理及研究者能力、各区域监管机构间的协调和统一，以及伦理问题等方面。具体来说，这些挑战包括但不限于：

1.2.1 设计和统计方面

MRCT 的目的是在不同地区同时评估治疗效果，对其设计提出了很大挑战，其中涉及多方面的考量要点。首先，需要根据疾病特点、影响疾病或治疗结局的内在或外在因素，预先定义同质区域及合并策略，因为它们会影响疗效的整体评估和一致性评估过程中统计方法的选择，及对结果的解释。其次，研究的总体样本量和区域样本量分配也是试验设计时需要考量的关键点，样本量的确定和主要终点、数据分析方法及单个区域监管要求等密切相关。例如，对于时间 - 事件结局指标，样本量将取决于区域研究中心开放的时间和目标区域的事件数量；如果在分析模型中使用包含地区间变异性的随机效应模型，则需要比固定效应模型更大的样本量。此外，在 MRCT 设计过程中，应及时和监管部门沟通，如果不同地区的监管机构有不同的批准要求，还应预先规定一个满足不同批准要求的处理方法。最后，在 MRCT 设计过程中还需要考虑很多其他因素，建议根据试验目的、主要 / 关键终点、分析的主要人群、主要分析的方法、预定义（合并）区域、区域差异内因外因、随机化问题、总体样本量、区域样本量分配、获益风险评估、一致性评价、非劣效性 / 等效性试验目标等多个因素进行考虑。

1.2.2 临床诊疗方面

如前所述，MRCT 的目的是在不同地区同时评估治疗效果，应在 MRCT 的计划阶段识别对药物研发项目有重要影响的内在和外在因素。在 ICH E5 指导原则中提及了可能影响处理效应的内在或外在因素，在外在因素中包括一些临床诊疗方面的考量：不同区域对疾病定义、诊断方法、疾病严重程度、常用的合并用药、入院标准、对某些终点的理解，或监管政策及生物标志物分布等因素，若存在很大区域差异，则可能会影响 MRCT 的实施，对此可以定义清晰入排标准以及研究步骤来减少相关差异。此外，患者的人口学特征（年龄、性别、种族等）、基线疾病状态、体重和 BMI、治疗依从性、饮食、吸烟及饮酒、环境、文化背景、社会经济因素（例如，避孕药的使用，给药途径的偏好）、教育水平、可获得的健康服务、安慰剂效应、以及受试者对不同药物的反应受内在因素敏感性影响等导致研究人群异质性高的因素也均是 MRCT 执行过程中需要面临的挑战。

1.2.3 临床试验运营环节

在临床试验执行过程中，最重要的是保证从 MRCT 中获得的试验数据的质量，只有高质量的数据，才能在异质性较大的全人群中确定试验药物的疗效，并分析该整体疗效是否适用于单个区域。为了保证顺利开展 MRCT，获得较高的数据质量，在计划 MRCT 过程中，需要对以下环节仔细考量：中心和区域选择、对英文 SOP 的本土化翻译并培训、患者招募流程要点及注意事项、患者筛选、临床试验过程中药物分发至临床试验中心及中心供给患者的途径、MRCT 全人群在各区域的随机化方案、对开展 MRCT 中心的研究者进行培训、获得的临床试验数据进行核查、试验实施过程中对于方案的遵循程度、试验过程中记录文件的保存、受试者脱落及停止研究的原因、报告规范性及利益相关性等。

1.2.4 监管方面

在 ICH E17 指导原则中，强调了与监管机构沟通的重要性。在试验计划、设计、实施、分析过程中，均存在需要与监管机构达成一致的要点。因为不同监管机构可能对试验的要求有所不同，如主要终点的选择和定义、如何对主要终点进行测量、试验对照组治疗选择、是否所有地区监管机构均认可该治疗的使用等。此外，MRCT 的主要思想是，首先检验药物在多个区域的全人群中是否具有治疗效应，当试验结果显示在所有区域的总体受试者中有疗效时，再评估单个区域的疗效。但是每个监管机构最关注的仍然是在本土人群中的疗效，因此不同区域监管对 MRCT 中其他区域数据的认可度、及本区域数据的要求可能会有不同，这也是执行 MRCT 过程中的一大挑战。

1.2.5 伦理问题

无论是传统的临床试验，还是 MRCT，均是在人群中开展的干预性研究，因此，都需要关注伦理相关问题。正如赫尔辛基宣言中规定的，“以人体为受试者的各项生物医学研究，必须首先关心受试者的利益，其次才是科学和社会利益”，所以在 MRCT 中首要关注的是各区域患者受益情况，如果存在区域获益 - 风险差异，则在设计、执行及分析 MRCT 过程中严格警惕，避免对某区域患者造成较大风险。

此外，当涉及有知情同意能力受试者的医学研究，必须充分告知每位潜在受试者研究目的、方法、可能的利益冲突、预期的获益和潜在的风险等任何与研究有关的信息，及其有拒绝参加研究、或随时撤回知情同意，而不会因此受到不当影响的权利，也就是在受试者参加试验前会签署知情同意书，知晓关于临床试验的相关信息和自己的权益。但是对于 MRCT，知情同意书的制定会有一些挑战，如不同区域的文化背景及语言环境不同，在制定知情同意书时，是否能使多区域的知情同意书的不同语言版本均表达统一的内容，是否能够避免文化背景造成的理解差异等。

同时，在伦理委员会进行伦理审查时也带来了新的困难：选择设立统一的中心伦理委员会还是设立区域伦理委员会？中心伦理委员会更能保证整个 MRCT 伦理标准在各区域审核的统一性，但实践中常见的是多个区域伦理委员会对 MRCT 进行伦理审查。如何制定统一的标准，使各区域伦理委员会能够对相关伦理问题进行均质审查，降低区域间的异质性也是一大难点。

除此之外，如何使各区域研究均保证透明化、如何保障各区域研究质量都在一个相近且较好的水平、各区域对试验研究审核质量、数据采集流程、数据安全保障、对药物临床试验质量管理规范（GCP）的遵从性等也都是实施 MRCT 伦理审查中的挑战。

第三节 主要挑战及应对策略

上一节从宏观上讨论了 ICH E17 落地实施中可能面临哪些方面的挑战，本节将从执行层面列出 E17 的实际挑战并探讨相应应对策略。在执行 MRCT 的过程中，核心围绕一致性评价，对种族因素的识别、样本量分配、对照组选择、其他 ICH 指导原则、合并策略等实施以及在 MRCT 中的执行以及与监管机构的沟通等方面都是需要重点考虑的问题。本节中深入分析这些关键问题，探讨它们在 MRCT 中的影响和可能的处理策略。

首先，我们将基于对一致性评价的讨论基础上，首先聚焦种族因素在临床试验中的重要性以及可能利用的新思想；接下来，将关注样本量的合理分配及其在区域一致性评价上的影响；此外，本节还将分析一致性评价的复杂性和对照组选择中遇到的困难，特别是在不同地区标准治疗可能存在差异的情况下，及在 MRCT 中实施《ICH E9 (R1)：临床试验中的估计目标与敏感性分析》的挑战；随后，我们将讨论合并策略等的实施问题。最后，将探讨与监管机构沟通的重要性，以期提供全面的见解和解决方案，以支持有效且可靠地设计和实施 MRCT。实践过程中的主要挑战及应对策略如下：

1.3.1 一致性评价

评价临床结果一致性，其核心是基于完整证据下的总体获益 - 风险评估。当试验结果显示研究药物在所有区域的全部研究参与者中有疗效时，需进一步评价各区域处理效应的一致性，也就是处理效应在纳入的所有区域中均无临床相关差异。

一致性评价并不仅仅是基于整体和区域结果在数值上的相似性，还应考虑通过生物学可解释性、统计学评估、有效性和安全性结果、预期风险、早期试验结果，及其他内部或外部数据等构成的完整证据链分析，进一步确认一致性结论和上市后的医疗预期，评估整体获益风险和区域获益风险。

随着各国监管机构大力推进新药研发进程，越来越多的复杂性创新试验设计在临床中得到应用，如适应性设计，允许根据期中分析结果按照方案预先规定进行试验调整，加快药物上市的同时也增加了试验的复杂性。此时，若同时计划开展 MRCT，则使试验的复杂度进一步增大。因此，急需相关共识阐明适应性设计结合 MRCT 时，解决如何计划期中分析，进行样本量分配等难题，来实现区域一致性评价。此外，由于罕见病药物研发领域存在大量未被满足的临床需求，各国监管机构推出相应激励政策鼓励罕见病药物研发。MRCT 为罕见病快速招募提供有效方法，但相较于一般 MRCT，罕见病人数少，历史及同类信息不足，目前无有效治疗存在单臂试验等特点，使合并策略的应用，和样本量的分配，更具挑战性。如何能在这种情况下，进行一致性评价，也是目前急需解决的一大问题。

1.3.2 种族差异的识别

MRCT 是在多区域进行的临床试验，涉及多个地区的受试者招募，种族是一个关键话题。由于 MRCT 假设处理效应对整个目标人群均适用，各区域样本量不足以进行描述性评价判断药效，而应采用描述性的一致性评价评估药物疗效，故需要提前识别区域间可能影响疗效的因素，如种族因素。在计划和设计 MRCT 过程中，种族因素是一个关键问题，ICH E17 指导原则也建议在计划阶段，仔细考虑区域差异、内在和外在因素（在 ICH E5 指导原则中描述为种族因素）影响^[2]，以确定 MRCT 在药物研发策略中可以发挥的作用。未来，随着 MRCT 在其应用中纳入的亚太地区试验中心逐渐增加，参与试验的中国、日本和韩国患者越来越多，在设计 MRCT 设计时更需要考虑内在和外在种族因素。目前，越来越多的研究者提出使用真实世界数据 (real-world data, RWD)^[3,4] 来评估区域间的种族差异，这虽然可以节省一部分探索种族差异的临床试验，但具体应用的统计方法和实施细节仍是一大挑战，需要进一步对其进行研究。

1.3.3 对照组的选择

原则上，MRCT 中设定的对照组在所有参与地区应该相同，但由于不同地区标准治疗可能不同，批准药物的剂型、用法用量、生产质量也可能有所差异，导致计划 MRCT 时对照组选择的复杂性和两组之间的可比性。此时可以考虑增加分层随机的策略，比如不同的化药，按对照药物分层随机等。此外，也有研究者提出 RWD^[3, 4] 可作为临床试验的外部对照，从而减少试验样本量。但 RWD 具有来源多样性、质量异质性高等特点，在 MRCT 中应用 RWD 作为外部数据会进一步增加整个试验设计、分析、及结果解释的复杂性，是否能应用仍需进一步探索。

1.3.4 样本量分配

MRCT 基于多区域整体人群主要疗效的效应值和变异度确定总体样本量，但是多区域人群整体疗效变异相对较大，可能会需要较大样本量。同时，根据各区域疾病特点、药物机制、内在或外在因素影响不同，整体人群疗效的估计可能存在较大挑战，可在相关人群中实施的试验药早期探索性试验，获得相关影响因素和参数数据，用以支持样本量确定。如前所述，一些监管机构仍要求使用本土临床试验数据用于注册，所以区域样本量分配时，可能会要求基于区域显著性或效应保留方法，也可能按照一定的比例分配本土样本量。如果 MRCT 中有多个区域均要求这类分配方法时，会增加 MRCT 整体样本量，失去了计划和设计 MRCT 的意义。前面提到的合并策略如果能够很好地落地实施，可能会降低监管机构对本土样本量的某些要求，增加 MRCT 样本量分配的灵活性和助力一致性评估。

1.3.5 估计目标与敏感性分析

目前，多个 ICH 成员国均已开始实施《ICH E9 (R1)：临床试验中的估计目标与敏感性分析》指导原则，但迄今为止对该指导原则的实践经验都较为有限，尚无适用 ICH E9 (R1) 的标准操作文件。因此，在 MRCT 实践过程中，考虑执行 ICH E9 (R1) 指导原则，将会更具有挑战性。

ICH E9 (R1) 指导原则主要涉及两大方面，估计目标的构建和缺失数据的填补。估计目标可在设计 MRCT 时从其五个属性来逐一考虑，尽可能使五个属性在各区域间保证一致，但由于地区间的种族差异等因素，整体人群和区域人群间的临床问题可能存在差别，需要在制定 MRCT 估计目标时，逐个进行考量。更多内容详见 E9 (R1) 蓝皮书第二版 6.2.3 国际多中心试验部分^[9]。缺失数据的填补方法原则上应尽可能保持各区域一致，但实际应用细节仍需要进一步探索，如数据缺失模式在某一区域和其他区域不一样时是否还能使用相同的填补方法，如果采用的填补方法不同，得到的证据强度是否会引起质疑；用全球的数据填补单区域数据是否恰当等。

1.3.6 合并策略的实施

合并策略有助于样本量分配和一致性评估^[1]。2020 年，欧洲制药行业协会联合会 (European Federation of Pharmaceutical Industry Associations, EFPIA) 曾对其成员公司调查，评估 E17 指导原则在一些 ICH 成员国和观察国实施的挑战。EFPIA 调查发现，在 ICH E17 指导原则定稿后，一些监管机构本土临床数据的数据需求有所降低，这与 ICH E17 指导原则设立最初目的相符合，呈现积极趋势^[5, 6]。

合并策略的前提是对潜在影响试验开展的內因和外因的合理分析，但很多区域本土的法律法规使监管机构仍要求本土临床试验数据用于注册，这也导致合并策略实践不足。种族因素可能涵括了多个內因和外因，为此应尽早参与早期研究或早期 MRCT，积累相关科学知识和数据，确定种族因素对疗效的影响。同时尝试从现有的经典案例研究中获得的经验，辅助制定合并策略。经过充分评估区域差异对药物有效性和安全性影响，在没有其它明确的内外因影响下，可以考虑 MRCT 中亚洲人群的合并，特别是将中国、日本、韩国的患者合并为东亚区域人群。

1.3.7 监管机构沟通

前文提到的 EFPIA 调查也发现，巴西、韩国、印度、俄罗斯、沙特阿拉伯等国家与监管机构沟通的 EFPIA 成员公司比例不足 1/3^[5, 6]，Asano 等人对日本批准的 167 项 MRCT 中涉及的 ICH E17 关键点执行情况进行统计，也发现日本对于科学咨询会议的考虑不足^[7, 8]。计划 MRCT 时与监管机构沟通不充分，会大大降低 MRCT 实施的可能性，和后期向多个区域监管机构同时递交新药研究数据的效率。此外，因为多区域监管可能提出截然不同、甚至相互冲突的要求，未能提前了解与沟通，不利于事先对研发项目形成统一的监管观点和认可不便事先对研发项目形成统一的监管观点和认可除了要提高在各个阶段与监管机构的沟通频率外，也需要保证沟通的质量^[1]。建议积极推进申办者和监管机构召开科学的咨询会，对影响试验方案制订的不同方面进行讨论。例如，在实际执行 MRCT 时，由于监管程序、临床研究中心的设置、药物供应等多种因素，某些国家可能无法与其他国家同步启动试验，导致入组时间窗口缩短。此外，各地区的竞争性入组机制可能会使部分区域难以完成计划的受试者入组。针对这种情况，一些申办者建议采取延展策略，即在同一 MRCT 方案中延长特定国家的入组时限。然而，延展策略是一种特殊情况下的应对策略，并非首选方法。在实践中，申办者应与监管机构保持积极沟通，并根据具体情况审慎评估策略的适用性。同时，也鼓励不同区域的监管机构之间进行科学讨论，以便在全球水平上统一研究要求。

第二章 研究设计阶段考量

第一节 一般考量

MRCT 通常可作为计划在多个区域监管递交的新药研究的首要选择，但在设计 MRCT 时，必须仔细考虑区域差异对试验设计和一致性评估的潜在影响。这意味着，在制定药物开发计划和 MRCT 设计时，应进行详细的流行病学和疾病情况调查，并总结相关信息，从而提前识别那些对药物研发具有重大影响的内在和外在因素。在 MRCT 的探索阶段，需要考察这些因素的潜在影响，并在确证性阶段继续收集相关信息。这对于后续进行的一致性评估至关重要。基于区域 MRCT 间相似性的已知信息，可以预先设定合并区域或亚群，这一决策应该反映在试验设计方案和统计分析计划中。这样的方式不仅有助于灵活分配样本量，还有助于评估区域间的一致性，并支持监管决策过程。在采用 MRCT 作为药物开发策略时，MRCT 的基本假设是治疗效应适用于整个目标人群。因此，在设计阶段，基于对产品和目标人群的早期认识，就应评估治疗效应适用于整个目标人群的假设，并确保这些考虑反映在 MRCT 的整体设计中。在此过程中，鼓励申办者在 MRCT 计划阶段与各监管机构进行有效沟通，以获得不同监管区域对全球研究设计方法的认可。这种沟通有助于确保试验设计符合各个区域的监管要求，从而加速全球新药的研发和审批过程。以下是更为详细的设计考量点：

2.1.1 疾病流行病学特征

在药物研发过程中，了解疾病的流行病学特征是需要考虑的重要因素，对 MRCT 的整体研发策略具有关键性的指导意义。流行病学的研究涵盖了疾病在特定人群中的发病率或患病率、病因和风险因素，以及组织学情况（如鳞状细胞癌、腺癌）和基因突变情况（如突变型或野生型），甚至预后情况。因为这些流行病学信息可以为包括患者入排标准制定、分层因素选择和样本量的分配等关键决策提供有价值的参考，因此对于 MRCT 的设计至关重要。例如，了解某种疾病在不同地区的发病率可以帮助申办者决定在哪些地区进行招募，以及预期招募所需的时间和资源；了解病因和风险因素则有助于确定哪些患者群体可能对治疗有更好的反应，或者更有可能发生特定的不良反应。

为了更好地理解和比较全球不同地区的流行病学情况，可以通过文字描述或图表形式对各区域的疾病情况进行总结和对比。这包括疾病的整体发病率或患病率，不同年龄、性别和种族等的发病率，以及常见病因和各种相关影响因素的分布情况。通过这种全面的数据收集和分析，研究者可以更准确地制定 MRCT 的设计方案，以确保试验结果既科学可靠又具有全球适用性。这种基于流行病学的方法不仅提高了 MRCT 的效率，而且有助于确保研究结果对各个参与区域都有意义和价值。

2.1.2 疾病诊断与治疗现状

深入了解特定疾病的诊断和治疗现状对于 MRCT 的设计至关重要。这不仅涉及对疾病的基本认识，还包括诊断方法、诊断标准、标准治疗流程（例如一线和二线治疗等）、临床实践以及治疗效果的全面理解，因为这些内容直接影响到研究的入选和排除标准的制定、分层因素的选择、随机化方法的设置、对照组的选择以及终点的确定等关键信息。

在临床研究中精确的诊断是临床试验成功的基础。不同疾病的诊断可能依赖于一系列测试，包括实验室检查、影像学检查和可能的基因检测。例如，在癌症临床试验中，精确的诊断出肿瘤分型和分期对于研究结果的成功和解释至关重要。同样，对于心血管疾病，了解病人的心脏功能评估方法如心电图和超声心动图等也是一样重要，尤其是在不同国家开展 MRCT 时，各种国际和地区的诊断标准可能会影响患者的入选，对研究结果的影响较大。

对于任何特定疾病，了解当前的标准治疗及其在临床上的应用是制定有效试验设计的关键。例如，在研究新的抗肿瘤药物时，了解现有的化疗、靶向疗法或免疫疗法对于设定适当的对照组和评估新疗法的效果至关重要。同样，对于像糖尿病这样的慢性疾病，了解全球现有的药物治疗和生活方式干预措施对于设计一个现实且具有可操作性的临床试验至关重要。

2.1.3 临床药理学研究和剂量

临床药理学研究在 MRCT 设计中扮演着关键角色，这种研究的作用不仅体现在药物开发的早期探索阶段，也贯穿于整个研究的后期确证阶段。早期这些研究不仅包括对药物的药代动力学 (PK) 和药效学 (PD) 的评估，还涉及剂量与药物暴露量和效应 (包括 PD、安全性、有效性) 之间的关系研究，以及剂量选择和确认等关键内容。如有必要，鼓励收集入选受试者的遗传数据 (例如，药物代谢酶的基因型)，以考察遗传因素对 PK 和 PD 的影响，其中特定基因型的受试者可能会被视为一个亚群。建议尽早在中国境内开展临床药理学和其他早期研究，以获得可评价数据。在早期研究中应尽量探索区域人群与整体人群的差异因素。早期探索性临床试验中获得的不同人群的药代动力学、药效学等数据可用于初步分析种族和区域间差异，为确证性 MRCT 研究的设计、实施和结果分析解读提供依据。有必要时，考虑增强 MRCT 临床风险管控措施，或进行剂量调整，甚至独立开展区域 (或亚洲) 人群的临床药理学或其它临床试验，以确保区域受试者的安全性和足够的治疗获益。

为确保区域受试者人群的风险可控，可通过临床前药代动力学、药效学、毒理学，以及疾病机理、临床操作、治疗手段等进行评估。在全球同步研发中，可考虑区域尽早加入早期 MRCT 研究，有助于高效的收集数据并提高对种族和区域间差异的评估。除了区域内的目标人群数据，已有的境外人群的早期临床数据，尤其是海外华人 (华裔)、亚洲人 (亚裔) 的数据，也是种族和区域间差异评估的有效证据。早期临床试验可在较宽泛的药物剂量范围内获得适应症人群中的剂量 - 暴露量 - 效应的关系，可尽早考察潜在的内在和外在因素，为在保障受试者安全的前提下开展确证性 MRCT 提供充分的支持，最终在注册时支持药物用法用量，指导上市后整体区域目标患者人群以及特殊人群的合理用药。

同时，除了早期研究阶段，临床药理学的相关研究也会在后期研究中继续进行，比如说群体药代动力学、群体药效学、特殊人群药代动力学等等。确证性 MRCT 的给药方案，原则上对所有参与的各种族人群均应相同。如果预期 PK 和 / 或 PD 以及剂量 - 暴露量 - 效应关系中不存在重要的区域差异 (例如，药物反应可能对内在和外在外在因素不敏感)，则不必从所有计划纳入确证性 MRCT 的区域获得受试者的 PK、PD 或剂量 - 暴露量 - 效应关系数据，可直接采用相同的给药方案。如果确认区域人群剂量 - 暴露量 - 效应存在明显差异，可综合内在在外在因素和剂量效应关系等整体情况考虑是否对给药方案进行调整。建议慎重计划采用不同剂量或者给药方案的评价策略。

2.1.4 效应因子

效应因子（Effect Modifier）在统计学和流行病学中广义地定义为，可能影响药物治疗效果的内在或外在因素。效应因子包括但不限于年龄、性别、体重、疾病严重程度、种族、区域、医疗实践和环境因素等。在早期临床研发阶段，真正的效应因子往往很难确定，隐藏在种族和区域等因素之中。地理区域或监管区域是许多种族相关因素的复合体或替代因素。种族相关因素比地理区域更为具体，依然混杂了多种影响因素。ICH E5 指导原则详细列出了各种内在、外在影响因素，如图 1 所示。内在种族因素是指有助于判断和鉴别亚群的因素，并可能影响临床数据在区域间外推的可能性，例如，年龄、性别、人种、遗传因素；外在的种族因素，是与人群居住地环境和文化相关的因素，例如饮食、医疗措施、大气污染等。

内因		外因
遗传学	生理和病理状况	环境
性别	年龄 (儿童-老年人)	气候
身高	肝脏	日照
体重	肾脏	污染
ADME (吸收、分布、代谢、排泄)	心血管功能	文化
受体敏感性		社会经济因素
种族		教育状况
药物代谢的基因多态		语言
		医疗措施
		疾病定义/诊断
		治疗方法
		药物依从性
		吸烟
		饮酒
		饮食习惯
遗传性疾病	疾病	压力/应激
		管理规范/GCP
		方法学/终点

图 1 内在和外在种族因素的分类

(1) 内在因素

1) 性别

男性和女性生物体内某些激素水平受性别影响较大，同时性别也会影响由于种族、社会或宗教背景而产生的心理和文化行为。在 ICH E5 中，性别因素主要研究女性和男性之间的差异及其相关的生物学机制，以及男性和女性之间在人类生理、病理生理和临床结局上的共同点。自该指导原则发布以来，性别医学已成为医学实践和临床试验设计中的一个重要横向维度。在药物研发中应该考虑，同一种疾病的症状、临床演变和男女之间的治疗效果的差异。因此，了解性别差异，并在 MRCT 中纳入具有代表性的性别人群，将有助于更好地解读结果。

2) 种族

ICH E5 指导原则附录 C- 对药动学、药效学和量 - 效关系的考虑中指出, 评估 ICH 区域最相关的三个主要种族群体 (亚洲人、黑种人和白种人) 的药动学、药效学以及他们的可比性, 对于 ICH 区域的药品注册至关重要。但是应如何定义种族是一大挑战。Collins 等人提出, 种族本身并不能作为衡量有临床意义疗效的指标, 在药物开发中应谨慎使用^[10]。因为种族是一种自我报告的社会结构, 具有一定程度的主观性。因此, 在药物研发早期, 识别具有预后意义或具有与疾病病因和治疗相关的预测性生物标志物更为重要。这些预后或预测生物标志物可能与种族相关, 也可能与种族不相关。所以, 根据自我报告的人种或族群对人群进行分组并不能准确预测患者的基因型或药物应答, 应积极研究其他方法来定义种族以外的人群, 从而更加有效地评估和促进药物研发^[11]。

3) 遗传因素

遗传因素主要指能经由基因的传递, 使后代获得亲代特征的所有因素, 不包括受环境影响的表观遗传因素。为了解 PK 遗传多态性, 过去十年曾进行了多项研究和分析, 以评估东亚地区内 (日本、中国和韩国) 及区域间 (例如东亚 vs. 白种人) 的 PK 差异及相似性^[8]。在这些研究中, 东亚人群具有相似的遗传背景, 因此可以作为一个群体, 并与白种人进行比较。这些研究从方向性上提示, 这种差异可能仅限于由 CYP450 酶代谢的药物, CYP450 的遗传多态性已在不同人群和区域广泛研究^[12]。此外, 在 ICH E5 指导原则发布时, 新药多集中在小分子或新化学实体, 对与药物吸收、分布、代谢和排泄相关的群体内变异较为敏感, 可能导致不同人群 (或区域) 的药物代谢和配置差异^[13]。目前, 新疗法多集中在生物制剂, 通常不太容易受到此影响。

(2) 外在因素

随着对疾病的遗传学和生物学的研究, 制药行业和学术界已经考虑了 ICH E5 中提出的内在因素, 但是, 在考虑外部因素方面仍存在滞后性。通过将不同的患者人群纳入 MRCT, 可以解决与社会经济因素、教育状况、获得预防性医疗保健相关的外在因素。然而确定外在因素通常相当主观, 很难纳入研究设计, 增加了 MRCT 实施的困难。同时, 目前外在因素的识别和收集往往存在困难, 如食物效应或医疗保健的标准一直难以收集和标准化。

(3) 效应因子的确定

确定效应因子一般分为两个阶段: 1) 收集信息。除了 E5 列出的内在和外在因素, 还应包括生物标志物等更多因素, 既要考虑疾病的预后因素也要包括对疗效的预测因素。可以通过搜索医学和科学文献、指导原则和其他公开信息, 收集疾病信息和遗传信息, 也可以通过检索各种医学相关数据库, 如 WHO 疾病数据库和各项登记研究, 收集流行病学数据和相关历史数据, 还可以咨询本地专业医疗工作者, 了解其所在地区临床实践和特别的治疗方法。2) 检查评估。根据收集的有关药物和研究的信息, 检查评估内在和外在因素对药物的影响。通过进行临床试验或者运用模型和外推的方法进行评估, 比如早期临床试验、暴露效应分析、PK/PD 模型和研究者主导的试验等。

在关键性 MRCT 开始之前确定效应因子是一种理想状况。如果不能明确, 可设定一些潜在影响可能性较大的因素, 在临床研发进程中通过不同阶段的数据结果逐步验证。在某些情况下, 关键性临床试验完成后, 数据结果揭示潜在的影响因素, 需要更多分析甚至开展更多临床试验才能进一步确认效应因子。

第二节 特殊考量

2.2.1 合并策略

ICH E17 指导原则的第 4 项基本原则强调了合并策略在 MRCT 中的重要性。该原则指出，基于对研究参与者在某些内在或外在因素的已知相似性的认识，预设合并区域或亚群有助于灵活分配区域样本量，这不仅有利于评估治疗效果在不同区域间的一致性，还支持监管决策。简而言之，通过基于相似性合并区域或亚群，可以在保证跨区域治疗效果和安全性评估的一致性的同时，为样本量分配提供更大的灵活性，进而促进 MRCT 的成功实施。

E17 涵盖两种合并策略：合并区域与合并亚群。合并区域指，如果认为某些区域受试者的研究疾病和 / 或药物相关的内在因素和 / 或外在因素具有足够的相似性，则在计划阶段合并这些地理区域、国家或监管区域。合并亚群指，将一个区域的受试者子集与另外一个区域具有相似定义的受试者子集进行合并，这些受试者需要具有一个或一个以上对药物研发项目非常重要的相同内在或外在因素。合并的亚群被认为是与种族相关的亚组，这在 MRCT 中尤为重要。

对合并策略进行论证应基于已知影响治疗反应的内在和外在因素的分布，研究疾病的分布和各区域上述因素的相似性。区域因素往往是潜在的区分区域或人群差异的内在和外在因素的替代物。在 MRCT 中，合并区域也是一种常见的合并策略。例如，由于相似的医疗实践和相似的合并用药使用，合理地将加拿大和美国合并为北美；中国、韩国和日本合并为东亚。在某些领域，在已有信息尚不完备的情况下，可考虑采用常见的合并策略比如合并东亚、亚洲、北美人群并在全球研发过程中持续收集潜在关键因素的信息为后续一致性评估提供帮助。除区域外，也可以根据影响治疗效应的关键内在或外在因素确定亚群，再将这些新定义的亚群也纳入分层、分析和一致性评估中。种族通常可跨越区域边界，并且可成为疾病或与治疗反应相关的一个重要风险因素。如果适用，应在方案、统计分析计划或者其他文档中预先定义分层，样本量分配，和一致性评估的合并策略。

应根据已知影响药物治疗效果的内在或外在因素的分布、疾病的特征及其在各区域之间的相似性来找到效应因子并论证合并策略的合理性。确定合并策略的关键是找到真正的效应因子。如果不能明确判别，可以基于科学论据，并兼顾可操作性来指定灵活的合并策略。首先用两阶段法从相关的内在和外在影响因素中探索和甄别效应因子。如果能确认真正的效应因子，可根据因子不同因素水平（类别）定义不同的亚群。换句话说，依据效应因子的不同水平进行分层（如图 2），针对影响疗效的因子水平合并亚群。如确定体重低的人群从某试验药物明显获益，按体重水平（低、中、高）分为三个亚群，亚群 1 为体重偏低的人群，在各个区域人群中，区域甲人群体重偏低，亚群 1 能很好地代表区域甲总体人群，同时在较大的样本量下支持评估在区域甲人群中的药物疗效。另一个例子是，某一生物标志物人群对药物的相似反应，如果区域人群在该生物标志物的某个类别比例较高，那么来自区域之外的类似人群可以与之合并，帮助支持区域患者人群的疗效估计。

	分层 1 (低体重)	分层 2 (中体重)	分层 3 (高体重)	总数
区域甲				
区域乙				
区域丙				
区域丁				
总数	合并亚群 1	合并亚群 2	合并亚群 3	

图 2 按效应因子 (Effect Modifier) 进行分层和合并亚群

确认真正的效应因子极具挑战性，如果不能明确判定，可考虑采用合并区域策略进行药物评估，既能考虑潜在混杂因素的影响，又兼顾可操作性。区域定义为：地理区域，国家或监管区域。在很多情况下，区域针对某个或某些效应因子，在对疗效的潜在影响和因素数据分布两方面有共性。可以考虑中国所属的整个亚洲区域或者东亚区域。因为亚洲地域广大，西亚、中亚、南亚和东亚区域的各方面差异巨大，所以可优先考虑东亚区域。

计划采用东亚区域合并策略时，应尽可能从流行病学和早期临床研究中获取数据，特别是种族敏感性数据，如药代动力学和药效学、基因数据、生物标志物等，探索东亚国家和地区的人群间是否存在差异。如果没有强有力的证据表明各人群之间存在明显差异，可以合并东亚人群，同时关注中国人群。如果有证据表明，东亚人群如中国、日本和韩国人群之间有疗效差异，合并策略可能不适宜，而中国人群可能应成为主要的关注点。

2.2.2 区域样本量分配

样本量分配应考虑对合并亚群间治疗效应差异进行计划分析。在 MRCT 中，目前尚无公认的或最佳样本量分配方法。如图 3 所示，ICH E17 介绍了 5 种区域样本量分配方法。每种方法都有各自的优缺点。一般来说，建议在按比例分配（#1）和平均分配（#2）间保持平衡，确保招募工作切实可行，试验能够及时完成，实现真正的 MRCT，同时还可为区域一致性评估提供充足信息。上述区域样本量分配原则也可适用于合并亚群。样本量评估，需要注意确保不存在单个或部分区域样本量占主导地位，从而主导试验结果的情况。



图 3 区域样本量分配的五种方法

样本量分配面临的一大挑战是如何选择并符合一致性评价标准，这可能需要不同职能部门之间以及审批机构和申办者之间进行数轮讨论。根据 ICH E17 原则，每种一致性评价方法都有其各自的优点和局限性。因此，建议采用整体方法进行一致性评价。因此，相较于首先使用某一个固定的一致性评价标准然后确定样本量，建议根据方案设计选择几个候选标准或终点进行一致性评估。可采用文字或者表格提供在一定样本量范围内符合这些候选一致性评估标准的概率的操作特征总结。然后，可以使用在按比例分配（#1）和平均分配（#2）间保持平衡的方法基于整体性证据确定国家和 / 或合并区域的样本量。基于已知的相似性，预先设定合并区域或亚群可有助于区域样本量的灵活分配，便于评价处理效应在各区域间的一致性，并支持监管决策。

第三章 研究执行阶段考量

在 MRCT 试验中，良好的执行是保障高质量 MRCT 的重要前提。在实施 MRCT 各个环节期间，应谨记实施和操作性问题，操作应遵守 GCP，确保临床试验的质量和完整性，并随时跟踪并进行调整。在 MRCT 启动时，应尽量与全球同步启动，缩小与全球同步启动和入组的时间；在研究参与者入组时，应及时监控入组速度与人群特征。同时，要特别注意 MRCT 在不同区域操作的一致性，建议尽量使用全球一致的标准系统和流程，包括疾病诊断、患者入排标准、随机化方法、药物供应、数据收集、数据清理、医学监查、质量控制、统计分析、研究者培训等。应尽量确保试验的高质量和良好的一致性，减少不必要的因素造成的变异和难以解释的结果，更好地促进有效的药物研发，增加在不同区域的多个监管机构同步递交上市许可申请的可能性和成功率。

当临床试验设计包含合并策略时，应尽早就关键问题与监管机构进行沟通，沟通内容包括但不限于合并策略依据、统计分析方法和样本量确定等。在单个 MRCT 中采用合并策略，应在临床试验方案中明确说明，并在统计分析计划中详细阐述。如果针对多个 MRCT 采用合并策略，应记录在单独文件中（如综合统计分析计划）。在整体临床研发过程中，合并策略可能依据早期临床数据结果和临床研究的实际运行情况进行调整。调整时间可以在关键性临床试验的设计阶段，也可以在试验进行之中。申办方应提前与监管机构讨论，至少在试验数据库锁定之前明确并记录最终策略。如果申办方和监管方不能对合并策略的具体实施达成一致，可以探索其他方法，例如延展策略，保证区域人群可以参与关键性的全球临床试验。

第四章 研究结果解释评价

第一节 一致性结果解释考量

ICH E17 建议采用整体证据的方法（holistic approach）从多个维度评价区域一致性，而不是只依赖于某一种终点或某一种标准得出结果。评估时应考虑生物学上的合理性、内部一致性、外部一致性、证据强度、临床相关性、以及统计学不确定性系统地进行分析和解读。

4.1.1 结果解读

在进行结果解读时，建议采取三层法进行思考。即，第 1 层，考察整体的有效性和安全性结果。第 2 层，利用合并策略（如适用），考察合并区域或合并亚群结果。注意合并策略并不止有合并区域 / 国家。可以视研究情况，考虑选择合并区域或合并亚群。第 3 层，考察特定国家或区域的获益 / 风险。其中，在第 2 层利用合并策略（如适用），可促进评估各区域或亚群治疗效果的一致性，并支持监管决策。对合并策略进行论证应基于已知影响治疗反应的内在和外在因素的分布，研究疾病的分布和各区域上述因素的相似性。如果没有强有力的证据表明各人群之间存在明显差异，可以合并相应区域或人群。

合并策略是 ICH E17 强调的重要概念之一。此策略不仅在计划阶段非常重要，在结果分析和探索阶段也具有重大意义。一般推荐提前定义合并策略。不建议在不考察合并区域和 / 或合并亚群的一致性的情况下，直接进行第三层特定国家 / 区域层面的亚组分析和评估相应的获益 / 风险特征。

4.1.2 临床药理学

评估药物暴露和效应种族因素影响的基本技术考虑要点通常包括：①药代动力学特征：暴露（局部作用、无或极低系统暴露量）；剂量 - 暴露关系（线性）；代谢途径（单一或多元）；相关药物代谢酶或转运体的基因多态性（CYP2C9、CYP2D6、OATP1B1）；药物形式（抗体、RNA）。②药效学特征：靶点的表达或基因多态性（VKORC、EGFR、HLA），剂量范围暴露量 - 效应关系特征（平缓或陡峭）；疾病的病理机制亚型分布（HCV）；临床实践（联合用药、伴随用药的可及性）。③安全性：治疗窗（宽或窄）；风险管控措施或监测方法；合并用药（中草药）。④其他：罕见病、儿童用药。同时，在临床开发的不同阶段，充分利用模型引导的药物研发手段，采用建模、模拟等方法和技术，对境内外获得的临床药理学研究数据进行分析，区分亚群间变异性和个体间变异性，以支持 MRCT 试验中药物剂量、用法、疗效、安全性等的一致性评估和解读。

4.1.3 有效性 / 药效学

4.1.3.1 内部一致性

内部一致性是指，是否可在 MRCT 试验的多个设置（例如，终点、分析、亚组）中观察到一致的药物治疗效应。一般可通过检查主要分析和敏感性分析的一致性、主要终点和重要次要终点的一致性、亚组之间的一致性进行评价。在 MRCT 中，我们通常需要采用合并策略，检查不同区域和不同亚群的治疗效果的一致性。在考察区域、亚群、和亚组一致性的时候，需要考虑样本量对变异度的影响，结果解读需谨慎。在结果呈现的时候，可以考虑用图表的形式，例如，图 4 把重要的终点结果整合展现在一张表里，重要的区域、亚群和亚组结果可展现在一张或多张图表里。

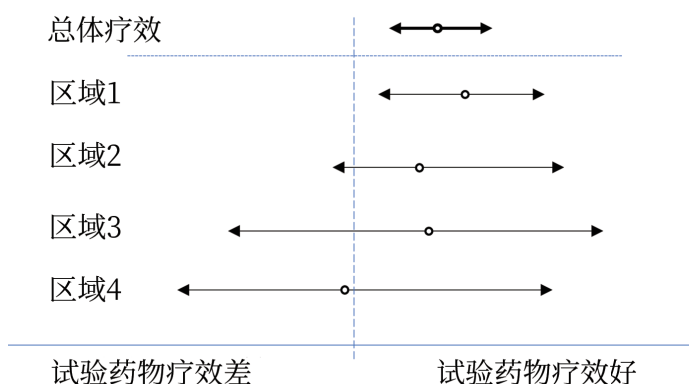


图 4 区域一致性的图形示例（来源：ICH E17 培训模块 6）

4.1.3.2 外部一致性

外部一致性是指，是否在多个数据来源中观察到一致的药物治疗效用或者一致的特定因素对治疗效应的影响。在某些情况下，可能需要递交多个试验来支持药物注册递交和监管决策。考察外部一致性的时候，一般考虑研发项目里的相似 III 期注册试验或 II 期探索性试验的一致性，也可以考虑来自研发项目之外但具有相似试验条件的试验。可以考察不同试验间的药物效用（组间差异、治疗组、或者对照组）是否一致。也可以考察特定因素、合并区域、和合并亚群对治疗效应的影响是否一致。除了展现每个试验的结果，如果有需要，也可以采用有效性综合分析（integrated efficacy analysis）的方法。有效性综合分析是对药物的拟申请注册同一适应症的全部临床有效性研究数据进行系统分析，比较不同研究数据的优势和不足，以描述总体有效性特征。在有效性综合分析的时候，也可以考虑采用合并策略。有效性综合分析的统计分析计划应与有效性综合分析报告一起提交给监管机构。建议与监管机构就有效性综合分析的统计分析计划进行充分沟通。

4.1.4 安全性

当进行区域 / 亚群和整体人群安全性比较的时候，对于区域 / 亚群的分析，无需重复整体人群上的所有分析，应根据产品安全性特性、目标人群、疾病等确定主要分析。可分析的数据可以包括：不良事件（AE）概况，重度（如 3 级以上）AE，严重不良事件（SAE），导致死亡的 AE，导致退出治疗的 AE，导致用药中断或剂量减低的 AE，根据药物特点选定的重要 AE（如特殊关注的不良事件），以及实验室检查等。除了比较上述参数的发生率、严重程度、和研究药物的相关性和不良事件类别外，必要时还可以进一步分析某些 AE 的发生时间，持续时间，转归及结局或者暴露量调整的 AE 分析。

分析策略包括评估研究药物组在区域人群和整体人群中的安全性是否一致，区域人群中研究药物组和对照组的组间差异是否和整体人群的组间差异相一致，报告的 AE 类别是否相似等。结果的呈现多采用描述性统计方法并用表和 / 或图形式展现。除非预设为研究终点，不建议在区域间安全性结果比较时，进行事后分析的假设检验。评估区域人群与整体人群的安全性是否一致，主要考虑是否存在有临床意义的差异，亚组和整体人群的不良事件发生率单纯在数值上的差别并不能确证存在真实的差异，因为可能受一些因素影响（例如整体和亚组的样本量，当区域亚组样本较小时，数值上的变异会比较大）。

4.1.5 获益风险

可基于前面的论证过程对整体证据进行综合评估得到对整体人群的获益风险结论并在此基础上对区域和整体一致性进行评价。内容包括：①治疗背景：如发病率 / 患病率、病因、危险因素、标准治疗、预后情况；②临床药理：如药代动力学、药效学、暴露 - 效应关系；③获益：如获益的终点，获益的幅度，获益时间；④风险：如不良事件的性质、发生率、严重程度等等；⑤获益风险的权衡和评估。

第二节 不一致性探索框架

即使经过慎重计划和良好执行，仍有可能观察到非预期的区域差异。如果观察到各区域处理效应存在可能的临床相关差异，则应对这些差异进行结构化探索，探索可能的差异来源，用整体证据衡量差异的重要程度及可接受性。同时要注意到考察特定区域 / 国家的数据应该采用描述性的框架，而不是假设检验。在探索中，应按照图 5 ICH E17 的原则系统得对内外因素展开探讨。

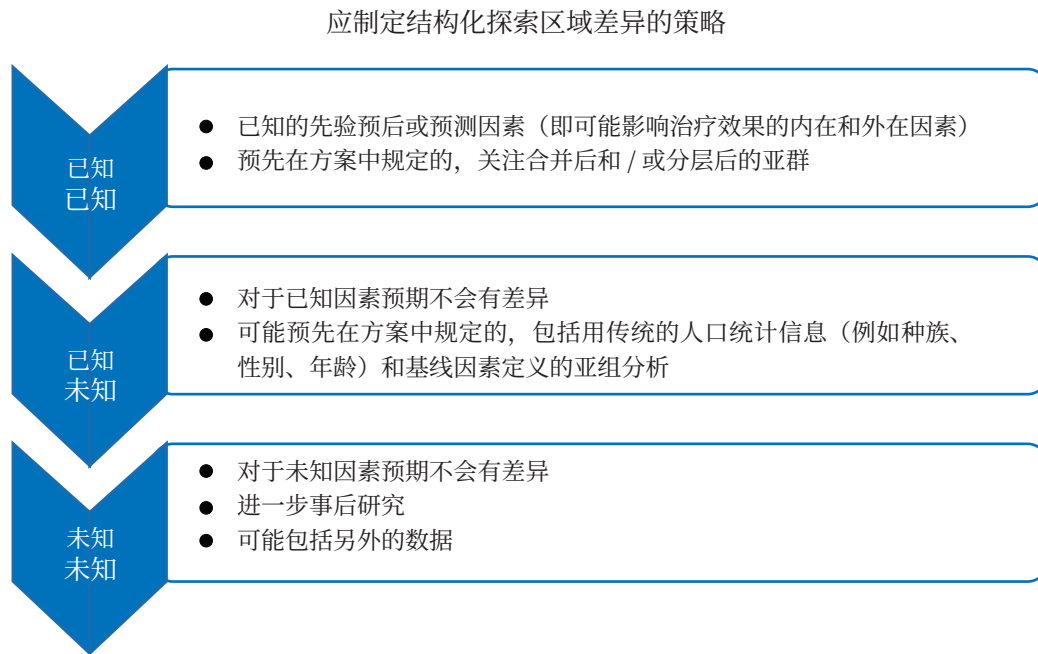


图 5 区域差异的结构化探索（来源：ICH E17 培训模块 6）

4.2.1 临床相关性

在 MRCT 背景下，各区域或国家间的差异可能不具有临床相关性。考量点包括目标人群、疾病负担、研究终点、和医疗实践。临床相关性还取决于治疗效应的特征，包括总结治疗效应的指标（例如，风险比或风险差异）。如果观察到区域间存在任何差异，首先需要评价该区域差异在大小和意义上是否具有临床相关性。例如，评估临床相关性时考虑区域背景，评价药物在疗效存在差异的区域是否均获益大于风险。对此，应在试验计划时考量，并进行相应的事后分析和深入分析。

4.2.2 疾病和治疗

除了设计和计划阶段对疾病流行病学、诊断、和治疗的总结，应使用最新的相关数据对相应的内容进行重新考察。在某些情况下，在试验期间某些区域 / 国家可能发生一些重要的变化比如试验结果更新、药物批准，或者指导原则更新从而导致对试验执行（退出试验 / 治疗）、后续治疗和治疗效果的影响。

4.2.3 临床药理

如在临床研究发现药物暴露量，剂量 - 暴露量 - 效应（PD、安全性、有效性）存在有临床意义的显著区域差异，应开展体外和 / 或体内机理研究，探索跟 PK、PD 有关因素的影响（如 4.1.2 章节所列举）。也可采用群体药代 / 药效、模拟等方法，分析各协变量可能带来的影响。

4.2.4 生物学合理性

此处生物学合理性是指基于与内在和外在因素相关的临床、药理学和机制考虑，可以预测或预期特定效应（在这种情况下，跨区域或亚组的治疗差异效应）的因果性和相关程度。合理性主要是临床和药理学判断，除非在计划阶段已经考虑，否则通常不是直接可量化或可测量的概念。一般来说对于局部用药和针对特定基因突变的靶向用药，不预期有大的区域差异。有些药物效果可能与基线体重/BMI、基线风险（高风险，低风险）、组织学（鳞癌，腺癌）、生物标志物表达水平（低表达，高表达）相关，而非与人种或者区域直接相关。此种生物学上的合理性可以用于考察有效性，也可以用于考察安全性；可以用于考察组间差异，也可以用于考察治疗组和对照组内差异。

4.2.5 入组情况及样本量

在一致性探索中，应考察详细的入组情况，包括全球及区域的入组开始时间、入组结束时间 / 状态、全球区域样本量分配及目标区域 / 国家 / 地区占比情况、和每个治疗组的样本量情况。入组情况对暴露时间和随访时间有一定的影响。样本量的占比和组间样本量的均衡会影响统计不确定性。

4.2.6 基线指标均衡性

评估区域人群是否与整体人群一致，治疗组与对照组之间是否均衡。除了预先定义的预测和预后因素，特别考察观察到基线差异和治疗差异的变量。可以考虑多变量模型矫正的方法（在模型中调整重要变量）进行调整来评估差异是否是由于基线差异引起。一般区域样本量较小，在变量较多，多变量模型不适用的时候，也可以考虑采用人群重抽样调整的方法（resampling）来评估是否特征类似的人群效果一致，而与人种无关。

4.2.7 暴露随访及分布情况

对于暴露情况、随访情况、分布情况（Disposition）包括退出试验的原因和退出试验治疗的原因，进行详细总结。对于联合用药得暴露情况，可以考虑不仅对联合治疗也对联合治疗里的每个药物的暴露情况进行总结。对区域与整体人群相似情况进行评估。如有不同，方向如何，是否对处理效应有影响。是否应该考虑采用暴露或者随访调整的分析进行评估。

4.2.8 内部一致性

内部一致性是指是否可在 MRCT 试验的多个情景（例如，终点、分析、亚组）中观察到一致的药物效应。如果不同分析方法显示结果一致、不同研究终点（例如，主要、次要和其他支持性终点）对应疗效一致、亚组结果基本一致和稳定、随时间变化的趋势一致，则加强内部一致性的力度。以下列出一些与内部一致性评估相关的考量点：

(1) 观察到特定区域终点的结果与整体人群结果存在差异而且此特定区域人群与整体人群在某个重要因素上存在差异。但在这个因素的每个亚群里区域人群与整体人群的结果一致（如图 6 PLATO）。

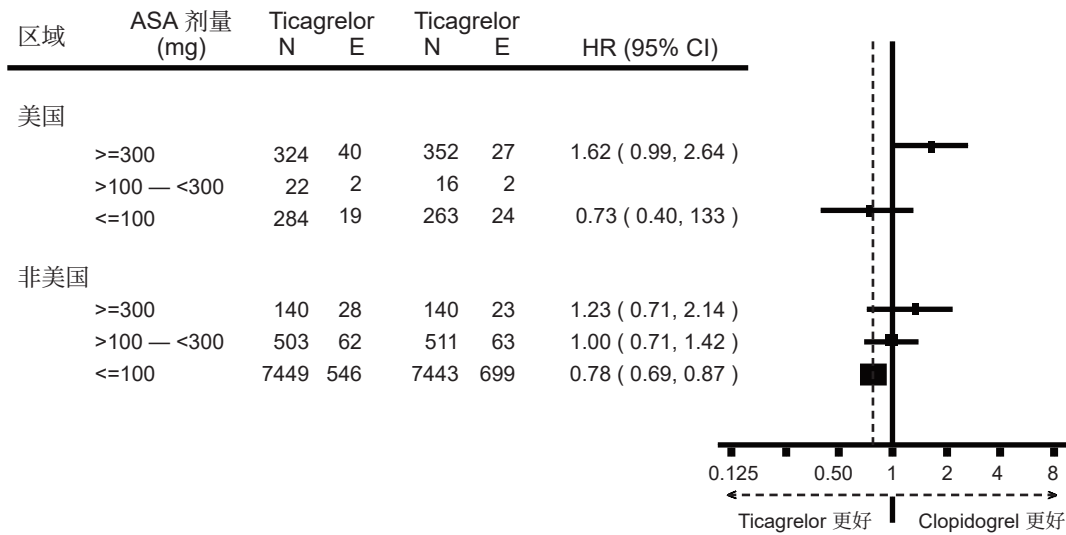


图 6 PLATO 试验中按地区和阿司匹林剂量水平列出的治疗效果估计值

(来源：ICH E17 培训模块 6)

(2) 特定区域终点主要分析的结果与整体人群使用同样的分析模型但是结果有差异。一些复杂的模型适用于大样本量数据。在小样本量上使用相同的复杂统计模型可能会造成模型不收敛或者不稳健。在使用优化的模型（AIC 或者 BIC 或者其他标准选择）后，结果变为一致。这种观察到的差异可能更多是由于复杂模型的不稳健的表现引起而非真实的临床差异。

(3) 特定国家主要终点的分析结果与整体人群结果有差异但是对于考察的重要因素，合并区域，合并亚群包括人种的结果均与整体结果基本一致表现出很好的稳健性。在这种情况下，观察到的区域差异可能更多是由于数据变异度引起而非真实的临床差异。

(4) 特定区域主要终点的分析结果与整体人群结果存在差异，但重要次要终点的分析结果与整体结果基本一致且具有临床意义的改善（12935 肺动脉高压研究）。这种观察到的差异更可能是由于某些指标在区域小样本量下的不稳定性（主观指标）和 / 或数据的变异度造成而非真实的临床差异。

(5) 特定区域终点的点估计结果与整体人群的点估计结果存在差异，但是考察按时间变化的趋势发现变化模式具有时间异质性（比如说延迟效应）。区域人群随时间变化的趋势与总部人群随时间变化的趋势一致（比如说分段 HR）。这种观察到的点估计总结的差异更多是由于随访时间不同造成的而非真实的临床差异。在随访时间类似的情况下或者随着随访时间的变化，可发现结果趋于一致。

(6) 特定区域终点的结果与整体人群的结果在某一个观察时间点存在差异，但是考察按时间变化的趋势，发现在之前的时间点和之后的时间点上区域人群结果与整体人群结果基本保持一致。这种观察到的差异更可能是由于数据的变异度引起的而非真实的临床差异。

(7) 特定区域终点的点估计结果与整体人群的点估计在期中分析结果存在差异。但是考察发现样本量随着分析时间的变长呈现特定区域样本量占比增加，变异度减少，一致性程度提高的趋势。这种观察到的差异可能是由于期中分析时样本量有限造成的变异度引起。

4.2.9 外部一致性

外部一致性是指是否在多个数据来源中观察到特定因素对治疗效应的影响：特别是，一个确证性临床试验中的不一致性是否也可见于外部临床试验数据，比如另一项 III 期试验、II 期探索性试验、或来自研发项目之外但具有相似试验条件的试验。外部一致性可以通过检查相似研究之间的一致性、与历史 / 外部数据的一致性、有效性综合分析或 meta 分析（如适用）来评估。在检查不同实验条件之间的外部一致性之前，重要的是要检查实验条件在治疗方案、研究人群、终点测量及其汇总指标以及伴发事件和多区域背景方面是否相似。以下列出一些与外部一致性评估相关的考量点：

(1) 在特定研究中观察到特定国家 / 区域终点的结果与整体人群结果有差异，但在其他类似研究中未观察到特定区域差异。有效性综合分析之后得到更加稳健的结果，发现特定区域结果与整体结果一致。

(2) 在特定研究中观察到特定国家 / 区域终点的结果与整体人群结果有差异，但治疗组效果类似，差异主要来自于对照组。但在历史研究中和同时期外部数据中未观察到特定区域疗效差异，这种情况下观察到的差异可能并不是真正的临床差异。

4.2.10 统计学不确定性

统计学不确定性是产生于多区域试验下偶然或随机的波动。当 MRCT 中包括多个区域、国家或亚组时，基于统计学不确定性的偶然或随机因素可能导致观测到的区域治疗估计值有数值上的不一致，特别是当区域样本量较小或者区域较多时。在分析上可以考虑合并策略的方法或者一些数据借用的方法（贝叶斯方法）来增加稳健性。同时也可以采用一些图形的方法比如说图 7 漏斗图的方法展现在不同的样本量下预计的变异区间（比如 95%CI）来更好的了解区域变异性从而进行一步考察。

4.2.11 安全性分析

当区域亚组和整体人群在安全性结果观察到明显差异时，比如 AE 的发生率、严重程度或 AE 类别出现明显差异的时候，则需要分析导致差异的原因。比如 PK 暴露和安全性的关系，内在和外在的因素，如患者的体重，基线，区域的医疗实践和 AE 的管理，合并用药（比如中药的使用）等等。有必要时还可以结合研究中与 AE 相关的辅助检查结果，和 / 或本研究药物在区域患者中的既往历史数据进行分析。如果存在亚组和整体人群暴露时间的差异，可以通过暴露校正后评估安全性是否一致，尤其是对于和暴露时间相关的某些事件，结合暴露时间校正分析更有意义。

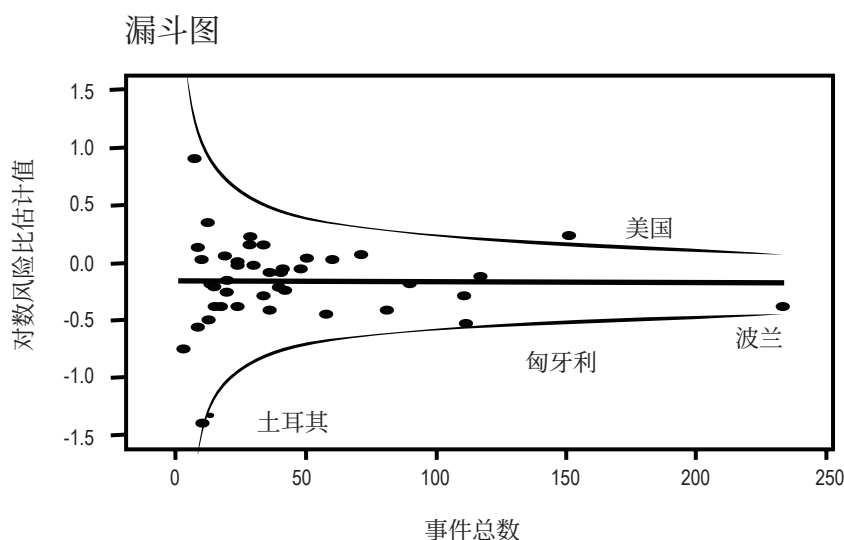


图 7 按样本量列出的治疗效果漏斗图（来源：ICH E17 培训模块 6）

第五章 特殊考量

5.1 研究设计类型

当确证某个药物疗效时，优效性检验一般会被首先考虑。当优效性检验不适用时，可考虑采用非劣效检验。非劣效检验需以阳性药物为对照，目的是确证试验药物的疗效，即使低于阳性对照药，其差异也是在临床可接受范围之内。如表 1 所示，非劣效试验一般采用固定界值法并且需在方案中提前定义非劣效界值 M （对于绝对度量指标 $M>0$ ，对于相对度量指标 $M>1$ ）。对于非劣效性研究，可将“治疗效应”替换为“非劣效性界值校正的效应”。之后，在非劣效性研究中可采用优效性研究中相同的方法进行样本量评估和一致性评价。

表 1 非劣效性界值校正的效应

绝对 / 相对指标类型	高优 / 低优指标类型	非劣效性界值校正的效应
绝对度量	$H_0: T - C \leq -M$ vs $H_1: T - C > -M$ ($M>0$)	$T-C+M$
绝对度量	$H_0: T - C \geq M$ vs $H_1: T - C < M$ ($M>0$)	$T-C-M$
相对度量	$H_0: T / C \leq 1/M$ vs $H_1: T / C > 1/M$ ($M>1$)	$T/C * M$
相对度量	$H_0: T / C \geq M$ vs $H_1: T / C < M$ ($M>1$)	$T/C / M$

绝对度量指标包括均值差和率差等；相对度量指标包括率比、风险比、比值比等
高优指标是其值越大表明疗效越好的指标；低优指标是其值越小表明疗效越好的指标
 T 代表试验组效应， C 代表阳性对照组效应， M 代表非劣效界值。

5.2 多个主要终点

在 MRCT 中可能会出现多个主要终点的情况。通常有两种情况，一种是要求所有主要终点均达到显著时才认为研究药物有效（此种情况通常称为共同主要终点）。一种是多个终点要求至少一个终点显著，即至少有一个主要终点显著时就认为研究药物有效（此种情况通常称为多终点）。在进行样本量评估的时候，也应该进行相应的调整。在共同主要终点时，建议提供多个终点均达到一致性的概率。在多终点时，建议提供每个终点达到一致性的概率。一致性评估采用描述性框架，不做假设检验。在进行一致性评估的时候，无需对多个终点进行多重性调整。

5.3 期中分析

在 MRCT 中可能会出现一次或者多次有效性期中分析。所有有效性期中分析应提前定义。如果试验在期中分析时有一个或者多个终点达到统计学显著性，申办方可选择采用期中分析数据进行注册递交上市。应按照方案的期中分析计划执行期中分析和一致性评估。在一致性评估时候，应该考虑全球试验终点的假设检验计划和结果。在期中分析时，应考虑不同区域入组时间的差别和期中分析样本量的差别而导致的对一致性评估的影响。建议提供入组情况和样本量帮助评估。如有需要并可能，可提供后续有效性和安全性随访数据。如果后续有盲态长期随访的要求（未检验的主要终点和 / 或者监管要求），应做好盲态控制。由专门的团队负责申请注册上市，负责继续开展试验进行长期随访的团队仍保持盲态以尽量减少对试验的影响。

5.4 延迟效应

在一些 MRCT 中，由于疾病或者药物的特性，可能存在延迟效应的情况（比如，前 6 个月 HR 为 1，后面 HR 为 0.7）。在这种情况下，如果区域有延迟入组，比如说比全球入组晚 6 个月，则可能造成直接进行一致性评估的困难和偏倚。此情况下，可考虑提供随时间变化的分段有效性和一致性评估数据，更长的随访数据，试验模拟或数据填补等支持性和敏感性分析。

5.5 适应性设计

目前 MRCT 应用适应性试验设计的情况已经变得越来越普遍。一般定义适应性设计为按照预先设定的计划，在期中分析时根据试验期间累积的数据对试验做出相应修改的临床试验设计。常见的适应性设计包括样本量重新估计、成组序贯设计、剔除治疗组、和无缝设计。可进行的适应性调整类型包括因无效或有效而停止试验以及对样本量、患者人群、治疗组或患者分配比例进行调整。相比于传统的固定设计，适应性设计可通过使用试验中的累积数据进行期中分析，根据预先规定的规则修改试验过程，使临床试验更灵活，更有效，更伦理。

近年来，适应性设计在 MRCT 中越来越常见。适应性设计和 MRCT 在试验设计、计划和执行方面都很复杂。一起使用在实施适应性 MRCT 的时候，应特别注意确保试验设计和执行能够达到目的。在实际情况中，一些国家的研究中心启动时间通常比其他国家晚几个月，原因可能包括监管和机构审查委员会批准要求的差异。对于适应性 MRCT，不仅要考虑整个试验的区域样本量分配情况，也要考虑期中分析时的区域样本量分配情况。因为进行适应性修改的计划需要在试验开始之前在试验方案以及统计分析计划中进行描述，所以如果 MRCT 被认为是可以支持到全球同步研发，那么对于最终结果一致性的考量，既要保证在潜在的方案的适应性改变下依然适用，也要保证其不会因方案改变而变化。如果有时间异质性，也需将时间异质性纳入考量。对于适应性 MRCT，可以采用计算机模拟以确定试验的操作特征来帮助试验设计和样本量分配。

5.6 单臂研究

单臂试验（single arm trial, SAT）是指设计为开放，不设立平行对照组的一种临床试验。单臂研究常用于新药早期探索性阶段。在某些特殊情况下，也可考虑采用 SAT 的研发策略支持药物的加速批准，显著缩短新药的上市时间，比如无药可治的晚期恶性肿瘤患者，治疗效果非常突出的新药，或者患者人数非常少的罕见疾病领域。鼓励申请人与监管机构就此策略进行沟通交流，共同讨论此策略的合理性。SAT 通常采用替代终点为主要终点，比如肿瘤单臂试验中常用客观缓解率作为主要终点。

按照 ICH E17 原则，可采用多个备选的一致性评估标准从不同的角度用整体证据的办法评估一致性。单臂研究一般全球整体样本量小，入组窗有限，留给区域的入组时间不多。由于单臂试验的设计本身存在一定的局限性，难以控制混杂因素，需要更严格的按照方案的预设执行。同时建议综合考虑早期研究和 / 或试验中的临床药理学评价整体评估一致性。

5.7 罕见疾病

罕见疾病是指发病率 / 患病率极低的一组疾病的统称。由于罕见疾病发病率 / 患病率极低、病情复杂、目前对其认识相对有限，使得罕见疾病药物研发所面临的困难远远超过常见多发疾病。其药物研发具有以下特点：通常流行病学和疾病自然史等数据有限，医疗信息不充分，可能缺乏公认的疗效评价方法和研究终点；患者参与临床研究的机会有限，药物研发经验较少；疾病可能存在多种亚型，患者异质性强；一般采用小样本临床研究，入组窗有限。

由于罕见疾病人数少，MRCT 适合被用作招募罕见疾病和实现全球监管递交的快速方法。全球各监管机构针对“罕见疾病”的药物研发可能设定了不同的审评标准。并且对于特定疾病的未满足需求以及特定开发项目需满足的一致性标准，在进行评估时实际操作也可能会有所不同。提前与监管机构进行讨论很有必要。

一般来说，对与罕见疾病，考虑到实际上稀缺的患者人群和高度未满足的医疗需求，区域样本量分配通常需要更大的灵活性。样本量可能更多的考虑实际可行性情况而不是统计学考量。在全球样本量分配上一般实施全球竞争入组而不是固定区域样本量分配。罕见疾病一般来说历史数据有限，研发经验较少。在历史数据和临床经验不充分的情况下，可考虑借鉴其他相关领域常见的重要因素和合并策略。罕见疾病很难通过传统途径研发这些疾病的治疗药物。全球研究经常采用创新复杂设计的 MRCT 单一研究，使得一致性评价更具挑战。在这种情况下，相比与传统的一致性评价，判断是否有其他证据可体现合理的种族差异可能更切实可行；鼓励采用更加灵活得方式，结合内外部数据对一致性进行整体评价。

参考文献

- [1] ICH, I.E., General principles for planning and design of multi-regional clinical trials [EB/OL] (2017-11-16) [2023-09-26]. https://database.ich.org/sites/default/files/E17EWG_Step4_2017_1116.pdf.
- [2] ICH, I.E., Ethnic factors in the acceptability of foreign clinical data [EB/OL] (1998-02-05) [2023-09-26]. https://database.ich.org/sites/default/files/E5_R1__Guideline.pdf.
- [3] Bunning, B.J., Hedlin, H., Chen, J.H., et al., The evolving role of data & safety monitoring boards for real-world clinical trials. *J Clin Transl Sci*, 2023. 7(1): p. e179.
- [4] Wasser, J.S., Greenblatt, D.J., Applying real-world data from expanded-access ("compassionate use") patients to drug development. *J Clin Transl Sci*, 2023. 7(1): p. e181.
- [5] Associations, E.F.o.P.I., ICH E17 Guideline on Multi-Regional Clinical Trials Catalyst for Simultaneous Global Development and Registration (2021 DIA Europe) [2023-09-26]. <https://globalforum.diaglobal.org/issue/june-2021/ich-e17-guideline-on-multi-regional-clinical-trials/#:~:text=A%20European%20Federation%20of%20Pharmaceutical%20Industry%20Associations%20%28EFPIA%29,and%20the%20existence%20of%20local%20legislation%20and%20For%20regulations>.
- [6] Associations, E.F.o.P.I., Industry Survey on Implementation of ICH E17 Guideline General Principles on Planning and Designing Multi-Regional Clinical Trials (2021 DIA Europe) [2023-09-26]. <https://globalforum.diaglobal.org/issue/june-2021/industry-survey-on-implementation-of-ich-e17-guideline/>.
- [7] Asano, K., Aoi, Y., Kamada, S., et al., Points to Consider for Implementation of the ICH E17 Guideline: Learning from Past Multiregional Clinical Trials in Japan. *Clin Pharmacol Ther*, 2021. 109(6): p. 1555-1563.
- [8] Asano, K., Uyama, Y., Tohkin, M., Factors Affecting Drug-Development Strategies in Asian Global Clinical Trials for Drug Approval in Japan. *Cts-Clinical and Translational Science*, 2018. 11(2): p. 182-188.
- [9] DIA 中国统计社区, ICH E9(R1) 估计目标及敏感性分析蓝皮书 (2021-11-01) [2023-09-26]. <http://www.diachampion.com/index.php?c=show&id=12>.
- [10] Collins, F.S., What we do and don't know about 'race', 'ethnicity', genetics and health at the dawn of the genome era. *Nature Genetics*, 2004. 36(11): p. S13-S15.
- [11] Bonham, V.L., Callier, S.L., Royal, C.D., Will Precision Medicine Move Us beyond Race? *New England Journal of Medicine*, 2016. 374(21): p. 2003-2005.

- [12] Zanger, U.M., Schwab, M., Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & Therapeutics*, 2013. 138(1): p. 103-141.
- [13] Wilson, J.F., Weale, M.E., Smith, A.C., et al., Population genetic structure of variable drug response. *Nature Genetics*, 2001. 29(3): p. 265-269.
- [14] Pharmaceuticals and Medical Devices Agency, P.E.N.N., Basic Principles on Global Clinical Trials [EB/OL] (2007-09-28) [2022-10-10]. <https://www.pmda.go.jp/files/000157727.pdf>.
- [15] Quan, H., Zhao, P.-L., Zhang, J., et al., Sample size considerations for Japanese patients in a multi-regional trial based on MHLW guidance. *Pharmaceutical Statistics*, 2010. 9(2): p. 100-112.
- [16] Hung, H.M.J., Wang, S.J., O'Neill, R.T., Consideration of regional difference in design and analysis of multi-regional trials. *Pharmaceutical Statistics*, 2010. 9(3): p. 173-178.
- [17] Lan, K.K.G., Pinheiro, J., Combined Estimation of Treatment Effects Under a Discrete Random Effects Model. *Statistics in Biosciences*, 2012. 4(2): p. 235-244.

附录：蓝皮书涵盖的相关方法学技术细节

1. 统计分析模型

可考虑采用固定效应模型（Fixed-effect Model, FEM），连续型随机效应模型（Continuous Random-effect Model, CREM）和离散型随机效应模型（Discrete Random-effect Model, DREM）对区域治疗效应和整体治疗效应以及区域之间的变异度进行深层次评估^[16, 17]。使用这些模型时要注意不同的模型假设，FEM 假设区域治疗效应和区域的样本量都是固定的参数，CREM 假设区域治疗效应是服从某种分布的随机变量而区域样本量是固定的参数，DREM 假设区域治疗效应是固定的参数而区域样本量是服从多项式分布的随机变量。

2. 延展策略相关技术细节

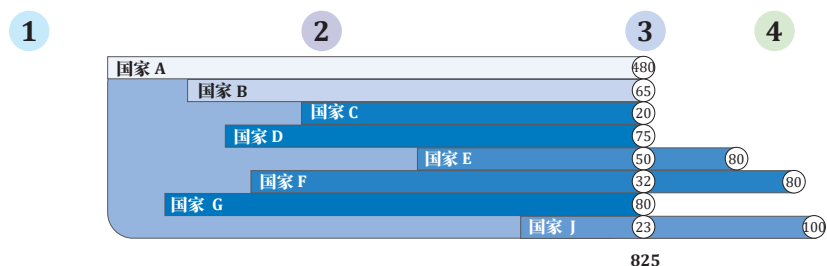
在 MRCT 中可能会出现多个主要终点的情况。通常有两种情况，一种是要求所有主要终点均达到显著时才认为研究药物有效（此种情况通常称为共同主要终点）。一种是多个终点要求至少一个终点显著，即至少有一个主要终点显著时就认为研究药物有效（此种情况通常称为多终点）。在进行样本量评估的时候，也应该进行相应的调整。在共同主要终点时，建议提供多个终点均达到一致性的概率。在多终点时，建议提供每个终点达到一致性的概率。一致性评估采用描述性框架，不做假设检验。在进行一致性评估的时候，无需对多个终点进行多重性调整。

（1）延展策略基本思想

如 1.3.8 中所述，在实际 MRCT 操作执行中，由于监管程序、临床研究中心设置、药物供应等因素，一些国家可能无法与其他国家同步启动试验，导致入组时间窗缩短。此外，各区域竞争性入组机制可能使一些地区难以入组计划数量的受试者。在其它已知的方法（例如：合并策略）都不适用时，可积极探索了各相关方都能接受的、理论上科学合理的、操作上可行的补救措施。一些申办者建议延长同一 MRCT 方案内特定国家的入组时限，即延展入组策略（extended enrollment strategy, EES）。

如果由于传统无延展入组的 MRCT 不可行（例如：无法同时满足各国监管机构的注册要求）而采用 EES，则在全球主要队列入组结束后，在扩展队列中以无缝衔接的方式在特定国家中继续入组受试者。扩展队列与主要队列的入组标准、研究程序和重要研究终点相同。扩展队列受试者的数据将由同一研究运营团队按照相同的 GCP 质量控制程序，执行相同的数据收集、数据清洗和医学监查程序。从扩展队列中收集的数据的质量标准与主要队列相同，并依据同一方案进行评估，以确保疗效和安全性的评估在各区域间的一致性。其主队列的分析以及特定国家的分析需要谨慎的设计，避免引入偏倚。

EES 的总体框架描述如下：1) 全球队列招募入组，在第一个参与国加入后立即开始 MRCT 招募，确保完全的竞争性入组，最大限度地减少“主要队列”的试验期；2) 样本量不足国家延展入组，其所需的入组患者数量可能在相应国家的“扩展队列”中得到满足。以上内容可借助下图中的假设图予以说明。



- 1 前瞻性地确定参与 MRCT 的国家，并分配相应的样本量
- 2 作为完全竞争性入组运行
- 3 当达到主要队列的全部样本量时进行数据库锁定与递交
- 4 “延展”未达到计划样本量国家的入组，直到达到计划数量为止

图 8 EES 的说明性假设图

在设计阶段根据主要目标确定 MRCT 主要队列的样本量。在示例（图 8）中，计算的样本量是 825 例。在示例中，国家 A、B、C 和 D 等的监管机构可能对按照竞争性入组方法分配样本量，没有特别的要求，而 E、F、G 和 J 等国家基于疗效或安全性角度对样本量可能有各自的监管要求。MRCT 以完全竞争入组方式，从所有参与 MRCT 的国家招募患者。根据研究方案规定，当达到科学评估主要研究目标所需的样本量时，停止“主要队列”招募。对于具有计划样本量的国家，如果其在主要队列中的样本量未达到计划要求，将继续在对应“扩展队列”中招募，直至达到计划的样本量。

与 ICH E5 指导的单独桥接研究相比，采取 EES 的 MRCT 更符合 ICH E17 原则，即一项研究由一个运营团队根据一个方案开展，以便保持入组标准、研究程序、数据收集、数据清洗、医学监查、质量控制流程、统计分析、试验计划和研究者培训的一致性。这样可以促进在各区域的高效操作，以及区域与总体、区域间的一致性评价。与 MRCT 相比，独立研究可能需要更大的样本量，入组时间和研究持续时间更长，存在更多时间上的差异和延迟。同时，独立研究可在不同的研究框架下进行，使用单独的方案、团队、运营过程和培训，存在更多的可变性，导致一致性降低，为一致性评估带来挑战。

(2) EES 中不同队列的作用

应用 EES 时要注意一些分析考虑因素，表 2 总结了将主要队列和扩展队列用于全球申报的关键考虑因素。其中，主要队列（C1）一般情况下用于支持对全球监管机构的注册，扩展队列（C2）仅用于在某区域监管机构的注册。C1 具有充足的把握度，以达到方案所述的主要目标，基于该队列的风险 / 获益分析用于支持全球监管注册；C2 的设计是为了满足当地监管要求样本量，且仅纳入该区域患者。C1 是主要数据库，整体疗效和安全性结论均基于该队列的数据得出。对于针对全球监管的注册，不会前瞻性地对整个试验的 C1 和 C2 合并。区域水平的分析只作为亚群分析，即汇总两个队列（C1+C2）的区域患者数据，评价与 C1 在疗效、安全性和获益 / 风险特征方面的一致性，并在相应文件中予以总结。由于在全球递交时并不会汇总 C1 和 C2 的结果，因此 C2 对 C1 的结论没有直接影响。

表 2 MRCT 中主要队列和扩展队列的申报目的

主要队列 (C1) (用于全球注册)	扩展队列 (C2) (用于区域注册)
<ul style="list-style-type: none"> · 在研究方案中预设以针对主要目标; · 代表全球试验人群; · 具有充分的把握度, 支持进行主要分析和得出风险/获益结论; · 安全性/疗效分析在 CSR 中报告; · 包括主要队列中的区域受试者; 	<ul style="list-style-type: none"> · 在研究方案(或方案附录)中预设; · 不是全球试验人群的一部分; · 患者疗效与安全性数据在描述区域数据的文档中报告; · 在全球主要队列最后一例患者入组试验后(LPET)开始(仅在区域开展; C1+C2 的区域患者数据将用于评估与 C1 全球数据的一致性); · 如果 C1 因获益或无效而提前结束(例如, 基于期中分析), 将重新评估 C2 计划。

注: CSR 为 clinical study report, 以下简称 CSR

(3) 方案描述

对于使用 EES 的 MRCT, 其主要队列和扩展队列都应采用相同的研究方案, 因此两个队列的试验程序应相同。研究方案要描述 EES, 可在方案的正文部分, 或通过方案附录的形式提供。由于试验执行阶段可能存在区域入组速度的调整, 因此 EES 的方案可适当灵活用词来保证可选择性。此外, 相较于两项独立研究, 遵循同一方案可最大程度地减少执行中的偏倚, 相较于启动另一项(桥接)研究, 可切实节省所需的时间和资源。

(4) EES 对 MRCT 全球递交的影响

一般情况下, 在区域使用 EES 不会对其他监管部门(例如: FDA/EMA)递交资料的审评时限产生任何影响。无论是否在区域使用 EES, 基于主要队列的全球数据包的递交将遵循正常的时间表。在区域扩展队列完成时可根据监管补充递交需要再向其他监管部门(例如: FDA/EMA)提交全球数据包, 由其审查和批准。下图展示了主要队列与扩展队列各自关键事件点的关系。

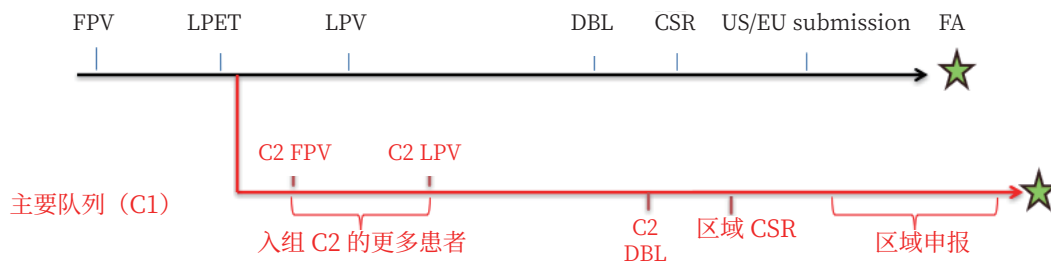


图 9 主要队列与扩展队列关键时间点的关系图

图注: FPV: first patient visit; LPET: last patient entering trial; LPV: last patient visit;
DBL: database lock; CSR: clinical study report; FA: final approval

(4) 含 EES 的 MRCT 样本量计算

延展策略中的样本量涉及主要队列的样本（C1）和扩展队列的样本（C2）。需注意的是，C1 中可能包括需要 EES 满足区域药品监管方要求的部分区域受试者。C1 的样本量基于 MRCT 的主要目标确定，并且没有涉及 C2 信息。C2 的样本量通常根据 C1 中的区域受试者和 C2 区域受试者数据合并后与 C1 全球数据的一致性统计标准确定。目前已有多种样本量计算方法用于没有任何扩展队列的理想 MRCT 的情况（即只有 C1 和没有 C2），开发了诸多方法用于评价一致性。其中，比较典型的是，日本厚生劳动省（MHLW）在“全球临床试验基本原则”指导原则^[14]中提出的方法 1。该方法规定，日本患者（或任何其他国家 / 区域）的样本量应采用以下方法确定：

$$P(\hat{D}_J / \hat{D}_A > \pi) \geq 1 - \beta' \quad (1)$$

其中，在日本患者和所有 MRCT 受试者中观察到的疗效分别用 \hat{D}_J 和 \hat{D}_A 表示， π 是一致性标准的阈值，而 $1 - \beta'$ 是预期表现出一致性的概率水平，指导原则建议此概率至少应为 80%。根据 Quan^[15] 的文章，如果 (1) 中的 $\pi=0.5$ 且整体 MRCT 把握度为 90%，则日本患者需至少占整体 MRCT 样本量的 22.4%。图 10 展示了在监管机构有样本量要求的关注区域 R1（如中国）中确定一致性的两种情况。

在此基础上稍作调整，上述一致性标准可应用于在区域使用 EES 的 MRCT。在设计阶段，假设主要队列将入组一定比例的区域患者，则可以通过模拟确定扩展队列所需的区域患者数量（以及需要的区域患者的总数），以满足一致性标准。更多统计学细节参见 (7)。

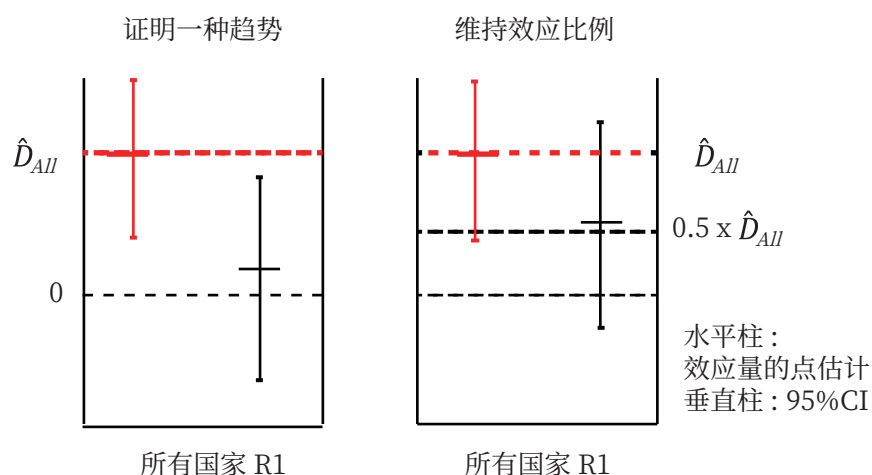


图 10 在 R1 中确定疗效的两种情况图示

(5) EES 实践中的考量

如果 MRCT 采用了 EES，可能需要提供方案附录（EES 附录），说明需要在某些区域入组更多患者。不过，提交附录并不意味着一定会执行 EES，附录的内容可以根据主要队列的入组情况和区域需求灵活调整。主要方案已定稿的情况下，无需为加入 EES 附录而修改方案，但如果主要方案和 EES 附录同时编写，简单提及附录会有帮助。EES 附录可在试验开始后撰写和执行，但需与监管机构达成一致，并在区域患者入组之前批准。附录中不必明确列出各队列的具体患者数量。统计内容应简明扼要，更多细节可在统计分析计划（SAP）中说明。为了透明，建议将 EES 的可能性告知所有监管机构，尤其是 FDA 和 EMA。

此外，还应为 EES 制定设盲和揭盲计划。在主要队列数据库锁定（DBL）之前，所有数据应保持盲态。主要队列揭盲后，扩展队列仍保持盲态，直到扩展队列 DBL 时才揭盲。EES 的区域队列数据（C2）与主要队列数据（C1）分开保存，并标记日期和国家。在 C1 锁定时，C2 数据会从主数据库中移除，并在 C2 锁定时提取出来。保持盲态有助于防止偏倚，并准确评估区域患者的安全性和疗效。

偏倚的程度取决于 EES 的计划和实施细节。如果扩展队列的入组和治疗在主要队列数据公开前完成，偏倚的可能性会降低。然而，如果扩展队列仍在进行，且主要队列显示明显疗效，可能会影响研究的完整性和数据解释性。新加入的研究中心也可能引入偏倚，因此需保持操作的一致性。

在使用 EES 时，扩展队列与主要队列的标准和程序应完全一致。扩展队列的数据由同一团队按照相同的质量控制标准收集和评估，确保疗效和安全性的一致性。这也符合 ICH E17 质量原则。与使用第二项研究（如桥接研究）设计相比，若采用 MHLW 方法 1 中的一致性标准，基于 EES 设计的 MRCT 在主要队列中的区域患者数量增加时，拟入组的扩展队列中相应区域患者的数量将减少，从而减少区域患者的总数（如表 3）。然而，实际操作中需要更全面地考虑这一问题。如果扩展队列仅入组少量区域患者（称为小型“尾队列”），且主要队列在竞争性入组期间已纳入了较多区域患者，是否有必要继续纳入“尾队列”值得探讨，特别是如果存在全球同步申报的机会。在这种情况下，申办者应结合所有现有证据（例如早期全球项目中的区域患者数据），评估是否确实需要小型“尾队列”。此外，申办者可以考虑以下两种方案：

① 切断“尾队列”：申办者可以在与监管机构协商后，利用主要队列中的现有区域数据，评估是否能够与其他国家同步递交 NDA。这可能涉及切断小型“尾队列”，或采用其他可行的策略，如合适的合并策略。

② 纳入“尾队列”：申办者也可以考虑将小型“尾队列”纳入 MRCT 的主要队列中，作为同一主要分析的一部分。

此时，为了确保主要队列能包含“尾队列”，全球 LPET（最后患者入组时间）可能需要推迟。相反，如果入组扩展队列的区域患者比例较高（称为大型“尾队列”），在计划 EES 时应谨慎。首先，根据定义，要求至少有 1 例患者入组或随机分配至主要队列才符合 EES 的条件。因此，无患者入组或随机分配至主要队列、所有区域患者均入组或随机分配至扩展队列不属于 EES，应视为另外一项单独桥接研究，相关要求的指导详见 ICH E5。其次，大型“尾队列”可能需要大量时间入组区域患者，之前讨论的因素可能会影响扩展队列结果的完整性和解读，如主要队列结果已知的情况下，扩展队列的盲态和潜在偏倚。建议当预期尾队列样本量合理时考虑 EES，且在公布主要队列结果之前能够完成扩展队列患者的入组和治疗。

表 3 使用与不使用 EES 的 MRCT 所需的区域患者数量 (示例)

	使用 EES 的 MRCT (1)	使用 EES 的 MRCT (2)	开展第二项研究的 MRCT
主要队列	800 例 (主要队列入组 80 例区域患者)	800 例 (主要队列入组 120 例区域患者)	800 例 (主要队列入组 0 例区域患者)
扩展队列 (或第二项研究)	132 例	66 例	242 例
区域患者总数	132+80=212	120+66=186	0+242=242

此外，在区域使用 EES 有助于满足区域特定的入组需求，且 EES 不影响主要队列。主要队列是全球申报的基础，所有风险 / 获益结论均基于主要队列中患者的结果。区域延展入组对其他区域的注册申报没有程序性影响，不管是否使用 EES，主要队列都将遵循正常的时间表。但应基于监管机构之前对该概念的了解与监管机构讨论 EES。通常在 II 期结束 (EOP2) 会议时告知 FDA 和 EMA，并可在简述文件中包含 EES。主要队列的数据构成通用技术文件 (CTD)，并用于所有全球申报。如需要的话，可向其他监管机构提供扩展队列中的区域患者数据。在某些情况下，即使扩展队列仍在进行，也可使用全球申报数据包向监管机构申报。但应考虑已入组主要队列的区域患者比例、支持注册申请所需的其他资源、提前申报的获益和监管风险等因素。总之，预计使用 EES 不会对其他区域或国家的注册申报产生重大影响。但是，需要考虑每个区域或国家的具体情况，并根据需要与监管机构进行沟通。

(6) EES 统计学技术细节

对于使用 EES 的 MRCT，样本量的计划涉及两个方面：一方面是主要队列，另一方面是具有监管要求的区域。为了方便起见，将监管机构有样本量要求的关注区域称为 R1，其余区域统称为 R2。R1+R2 构成参与 MRCT 的所有区域。

本节将重点关注区域样本量，因为无论是否使用 EES，“主要队列”的样本量均没有变化。如果 R1 要求固定数量的患者暴露于研究药物，通常是从安全性角度出发，问题相对简单。但是，如果 R1 要求提供来自该区域患者的有效性和安全性证据，可能比较困难。最简单方法是通过仅在 R1 中进行单独试验（具有充分把握度）来获得该类证据。但是，如果没有内在或外在因素证据表明在 R1 人群中可能存在不利获益 - 风险比，该方法可能会因试验重复而存在伦理争议。在 MRCT 框架内，也有一些使 R1 亚组人群检验具有统计学把握度的做法。尽管满足 R1 的监管需求并不违背伦理，但它扩大了 MRCT 的样本量并有违开展 MRCT 的目的，特别是当涉及两个以上的区域时。

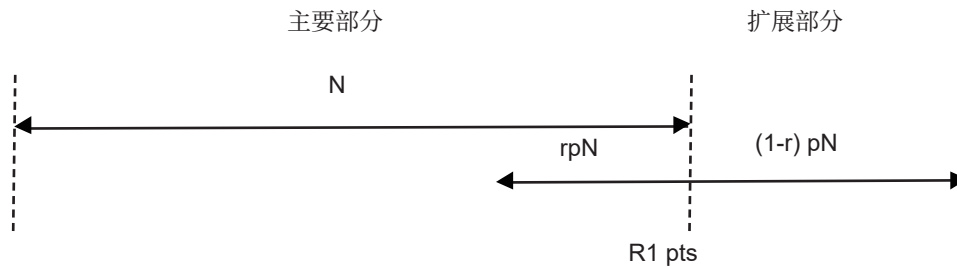
在 R1 的药物研发中，将一致性作为疗效证据较为常见。研究人员在过去的十年里已开发出诸多方法用于评价一致性。以之前在 (1) 中提到方法 1 为例，(1) 中的标准仅关注在 R1 中显示的治疗，不要求考察全球所有区域的一致性。

目前已开发了一些该方法的变体。在不失一般性的情况下，重点关注基于 (1) 中的标准进行 R1 样本量确定的讨论。可以通过强调对真实疗效 D_J 和 D_A 的依赖重新制定 (1) 中的标准：

$$P(\hat{D}_J / \hat{D}_A > \pi | D_J, D_A) \geq 1 - \beta'$$

如果本节中的观点不受此假设的限制，在接下来的讨论中，假设 $\hat{D}_J = \hat{D}_A$ 。在实践中， $\pi=0$ 和 $0<\pi<1$ 可能指向两种情况，前者描述一种“趋势”，后者维持 R1 的特定效应比例。“趋势”可能适用于满足高度未满足的医疗需求且风险特征合理的药物，而后者可能更适用于一般情况。

标准 (1) 仅在整体试验成功时才有意义。可用 $A=\{D_J / D_A > \pi\}$ 和 $B=\{\text{MRCT 成功}\}$ 表示两种事件。监管机构通常会在观察到 $A \cap B$ 事件的情况下，批准该药物在 R1 使用。对于试验申办者，条件事件 $A|B$ 为主要关注点。在后续讨论中， $P(A|B)$ 表示显示一致性的条件概率， $P(A)$ 则表示无条件概率。



使用 EES 的 MRCT 的样本量计划如图 11 所示。主要队列样本量的确定与是否使用 EES 无关。以下符号较为直观：

N = 主要队列的样本量；

p = $R1/N$ ，即 R1 的总样本量相对于 N 的比例；

r = 主要队列中分配给 R1 的样本量占 R1 总样本量的比例；以及

$pN=R1$ 的样本量，其中 rpN 是主要队列中 R1 的样本量， $(1-r)pN$ 是扩展队列中 R1 的样本量；

$[1+(1-r)p]N$ = 主要队列和扩展队列合并的总样本量，假设 R1 是唯一涉及扩展队列的区域（根据本节开头的讨论）。

在采用竞争性方式入组时，通常假设 r 是随机数字。如果 r 为一个固定数，可以将其视为特殊情况。

(1) 证明同方向趋势

证明一种趋势可指 $\pi=0$ 。下列结果较为直观。

对于固定的 p 值， $P(A)$ 与 r 无关。

采用正态逼近法，使 $P(A)$ 至少为 $1 - \beta'$ ，

$$p = \left(\frac{Z_{1-\beta'}}{Z_{1-\frac{\alpha}{2}} + Z_{\beta}} \right)^2 \quad (2)$$

无论真实效应量和随机化比例如何，均适用。

对于固定的 p 值, $P(A|B)$ 通常大于 $P(A)$ 。差异随着 r 的增加而增大, 并在 $r=1$ 时达到最大。最大差异取决于主要队列的把握度。

根据 EES 的性质, r 是随机数字, 标准 (1) 可在 $\pi=0$ 的情况下得到满足, 即

$$\int_0^1 P(\hat{D}_J / \hat{D}_A > \pi | p, r) f(r) dr \geq 1 - \beta' \quad (3)$$

$f(r)$ 是 r 在 $[0,1]$ 数值范围内的分布, 由于试验期间动态招募, 因此其未知。要计算 p , 可以仅使用 (2) (具有保守性), 并且真实概率将始终不低于预期的 $1 - \beta'$ 。

为了便于说明, 假设主要队列的把握度为 90%。根据 (2), 通过设定 $1 - \beta' = 0.8$ 使 $p=6.74\%$ 。采用双臂平行设计, 以 0.3 的效应量模拟 10000 项试验, 得到的 $N=470$ (按照 1:1 的随机比, 每个组的样本量为 235 例)。 $P(A|B)$ 总结见图 12。 $P(A | B)$ 与 $P(A)$ 之间的差异不超过 0.02。

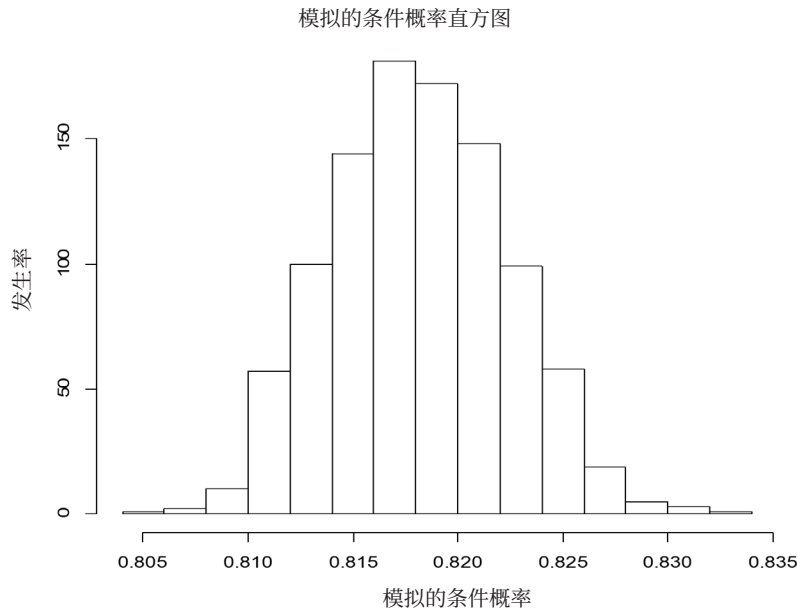


图 12 模拟的条件概率情况

(2) 维持效应比例

当 $\pi > 0$ 时, 指的是与 MHLW 方法 1 一致的一般情况, 作为特殊情况。以下结果可能较为显著。

对于固定的 p 值, $P(A|B)$ 和 $P(A)$ 均随着 r 的增加而增加。对于给定 $P(A|B)$ 或 $P(A)$, p 随着 r 的增加而下降;

当 r 值处在 $[0,1]$ 数值范围内时, $P(A|B)$ 不再一直 $\geq P(A)$ 。为了便于理解, 想象当 r 接近 1 时, 事件 B 表示在 R1 中具有良好的治疗效果, 从而通过分别向分子和分母增加一些值来影响事件 A 的分数, 使分数值增大, 随后 $P(A)$ 也会增大。相同的类比表明, 当 r 接近 0 时, 另一个方向可能保持不变。

当 $\pi > 0$ 时满足 (1) 的 p 应满足 (3)。 $P(\hat{D}_J / \hat{D}_A > \pi | p, r)$ 的表达式可见参考资料 Luo (4)。与之前相似, 严谨的做法是通过设定 $P(\hat{D}_J / \hat{D}_A > \pi | p, r^*) = 1 - \beta'$ 保守估计 p , 其中 r^* 是 r 最保守的估计值。

当 r 为 1 时, 满足 (3) 的 p 的显式解可参见参考资料 Quan^[15]。

模拟示例的结果请见图 13，其中 x 轴表示 rp ，y 轴表示 p 。当主要队列的把握度为 90%、 $1 - \beta' = 0.8$ 时，可根据参考资料 Quan^[15] 分析计算 $p = 22.4\%$ （图中的水平线）。对于 $P(A)$ （图中的蓝线），当 $rp=p$ 时，该线与水平线相交，证实了参考文献 Quan[15] 的结果。随着 r 的减小， p 增大，可在 $P(A|B)$ 看到类似的模式。当 r 接近 1 时， p 对应的 $P(A)$ 小于 $P(A|B)$ 。参考资料 Quan^[15] 中已经解释了 $r=1$ 的这种情况（并且其中还计算了无条件概率，注明了闭合解和保守性）。当 r 接近 0 时，预计方向相反。

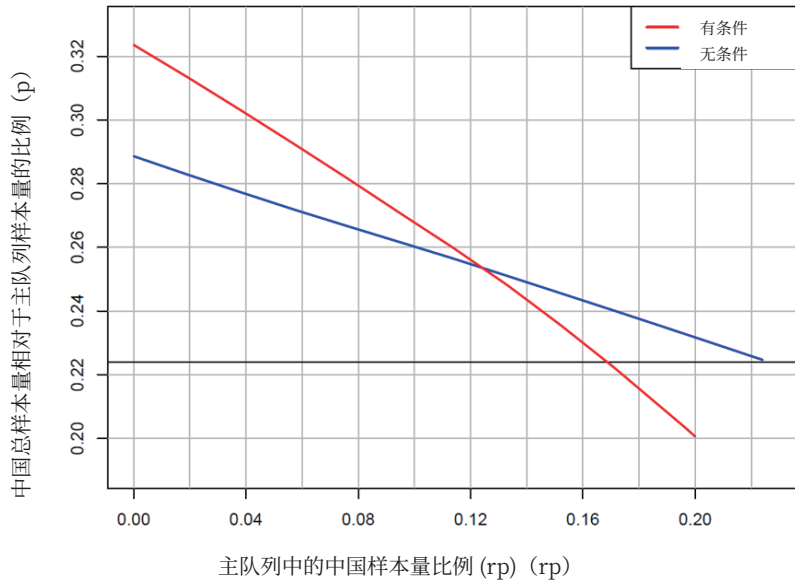


图 13 R1 样本量和主队列中 R1 样本量的关系

