

药物临床试验样本量估计指导原则

（征求意见稿）

2023年10月

目 录

一、 概述.....	4
二、 影响因素	4
(一) 试验设计	5
(二) 检验水准及检验效能.....	6
(三) 统计分析方法	7
1. 数据分布假设	8
2. 统计分析模型假设.....	9
(四) 预期治疗效应及方差.....	9
1. 参数设置依据	10
2. 伴发事件的影响	11
3. 缺失数据的影响	11
三、 样本量调整	12
(一) 样本量调整的原则.....	12
(二) 样本量调整的情形.....	14
1. 盲态样本量重新估计.....	15
2. 非盲态样本量重新估计.....	15
四、 试验实施过程中的考虑.....	16
五、 其他.....	17

(一)	基于贝叶斯方法的样本量估计	17
(二)	基于其他目的的样本量估计	18
(三)	样本量的敏感性计算	18
(四)	样本量重新估计与其他适应性设计的结合	18
(五)	与监管机构的沟通	19
六、	参考文献	20
附录 1	名词解释	28
附录 2	中英文词汇对照	30

1 药物临床试验样本量估计指导原则

2 (征求意见稿)

3 一、概述

4 样本量估计，又称样本量确定，是药物临床试验设计
5 的重要组成部分，也是确保研究具有合理性、准确性、可
6 靠性、完整性和科学性的重要手段。通常，临床试验的样
7 本量必须足够大，以可靠地回答研究假设所针对的临床问
8 题。

9 影响样本量估计的因素众多，当样本量估算相关参数
10 设置缺乏依据或依据不充分时，为样本量估计带来不确定
11 性，增大试验失败的风险。此外，若样本量调整的方法不
12 当或实施不当则会引入偏倚，导致总 I 类错误率 (FWER)
13 膨胀，破坏试验完整性。

14 为指导申办者进行科学合理的样本量估计，本指导原
15 则主要阐述基于统计假设进行样本量估计时的参数设置和
16 样本量调整等的基本考虑。本指导原则适用于以注册为目
17 的的确证性临床试验，也可供其他临床研究参考。

18 二、影响因素

19 对于临床试验提出的科学问题，统计学主要通过假设
20 检验回答针对估计目标提出的研究假设。样本量估计一般

21 依据主要估计目标，在 **FWER** 得到控制的情况下保证整体
22 检验具有一定的检验效能来确定。不同的假设检验方法，
23 有不同的样本量估计方法。正确的样本量估计应基于与估
24 计目标相一致的试验设计和恰当合理的统计分析方法。应
25 在方案中给出计算样本量的方法，以及在计算中使用的相关
26 参数及其依据。

27 影响样本量估计的因素众多，常见的影响因素有：①
28 试验设计，包括比较类型（如优效性试验、非劣效性试验、
29 等效性试验）、设计类型（如平行设计、交叉设计、析因
30 设计、群随机设计和适应性设计等）、变量类型（定量、
31 定性、时间-事件）、伴发事件的处理、组间分配比例及随
32 机化方法等；②检验水准和检验效能；③统计分析方法；
33 ④预期治疗效应及方差等。

34 （一） 试验设计

35 研究设计是样本量估计的重要影响因素。例如，在非
36 劣效性和等效性试验中，阳性对照药疗效以及相应的非劣
37 效和等效界值是样本量估计中的关键参数；对于单臂设
38 计，当采用目标值对照时，需设定合理的目标值；基于群
39 随机设计的样本量估计需要考虑群内相关性；适应性设计
40 在样本量估计时需考虑适应性决策可能引起的 **FWER** 膨胀
41 问题；采用分层随机化时可能需考虑各层比例与自然人群

42 保持一致以及分层因素导致的某个或某些层上研究者例数
43 稀疏等问题。

44 估计目标中伴发事件的处理也会影响样本量估计中对
45 治疗效应及方差参数的设置。疗法策略将伴发事件作为治
46 疗（处理）的一部分，其对样本量估计的影响主要体现在
47 治疗（处理）中伴发事件对治疗效应和方差的影响。在治
48 策略与复合变量策略，直接影响结局变量的定义，进而影
49 响治疗效应和方差。假想策略设想一种没有发生伴发事件
50 的情景，通常伴随着一定的假想假设和缺失机制假设，相
51 应的假设会直接影响治疗效应和方差。主层策略则直接影
52 响目标人群，该策略下主层人群的识别方法、主层人群所
53 占比例及相应人群预期的治疗效应和方差会影响最终的样
54 本量估计。

55 组间样本量的分配比例是样本量估计中考虑的重要参
56 数，药物临床试验常采用平衡设计，即各组样本量相同。
57 当出于伦理考虑或其他合理理由需要降低某组或某些组的
58 样本量的分配比例时，可采用非平衡设计。样本量在各组
59 的分配比例直接影响最终样本量估计，研究方案中须明确
60 说明。

61 （二） 检验水准及检验效能

62 检验水准和检验效能是样本量估计中考虑的基本参数，

63 须在方案中明确。

64 设置合适的检验水准 α 可以达到控制 FWER 的目的。

65 对于确证性试验，FWER 通常要求控制在单侧 0.025、双侧
66 0.05 以内。当涉及多重性问题时，名义检验水准的设置可参
67 考《药物临床试验多重性问题指导原则（试行）》。

68 对于检验效能，通常设定不低于 80%，当进行样本量
69 调整时，建议调整后的检验效能也不低于 80%，当涉及多
70 重性时，需考虑多重性对检验效能的影响。对于时间-事件
71 类型的变量，最终决定检验效能的主要是随访中观测到的
72 结局事件数，因此入组速度、随访时长等因素也会进一步
73 影响最终样本量。对于析因设计，如果以检验 A 和 B 的交
74 互作用为研究目的，若基于检验主效应计算样本量，则交
75 互作用的检验效能可能不足。对于多中心试验，样本量和
76 检验效能的计算通常基于各中心的组间治疗差异是相同的
77 无偏估计的假设，因此，制定共同研究方案并给予实施很
78 重要，同时试验的实施流程应该尽可能标准化。

79 (三) 统计分析方法

80 恰当合理的统计分析方法，是科学的样本量估计的基
81 础和前提，样本量估计前须确保所选择的统计分析方法与
82 研究设计相匹配。例如，统计分析方法应适合研究设计的
83 设计类型、比较类型、随机化方法、决策策略（如期中分

84 析) 以及变量类型等。方案中需明确给出样本量估计所基
85 于的统计分析方法, 且原则上应与主分析方法相一致, 否
86 则需有合理的理由认为所依据的方法能满足主要估计目标
87 所需的样本量, 即不会导致样本量低估。方案中应明确样
88 本量估计的具体计算方法、工具及出处, 当采用统计模拟
89 估计样本量时, 模拟的参数设置、模拟方法以及模拟代码
90 等应在相关文件中详细描述并提供。

91 需要注意的是, 每种统计分析方法都有其相对应的假
92 设, 在选择统计分析方法时需充分评估所选择方法违背其
93 假设时的风险, 以及违背假设对样本量估计可能产生的影
94 响。统计分析方法中常见的假设有①关于数据分布的假设;
95 ②关于统计分析模型的假设等。

96 1. 数据分布假设

97 统计分析常需要对数据的分布进行假设, 例如, 定量
98 变量的数据服从正态分布假设, 时间-事件变量的数据服从
99 指数分布假设等。需评估偏离数据分布假设的可能性, 一
100 方面当偏离风险较大时应采用对分布假设更稳健的分析方
101 法或不依赖于分布假设的分析方法, 另一方面在参数设置
102 时, 也应考虑偏离分布假设带来的治疗效应高估或方差低
103 估的可能性。当采用非参数或半参数方法分析时, 样本量
104 估计基于方便计算的考虑可能会依赖一定的参数假设, 例

105 如生存分析 log-rank 检验的样本量估计，有些方法会假设生
106 存数据服从指数分布，有些会假设满足等比例风险，在应
107 用时需考虑偏离假设所带来的风险。

108 当数据存在相关性并影响治疗效应或方差估计时，若
109 忽略相关性，则会导致样本量低估并对分析结果带来偏倚。
110 例如基于群随机的设计，群内个体相关性越高方差越大，
111 样本量估计时若忽略群内相关性则会低估方差造成样本量
112 的低估，统计分析时若忽略群内相关性则会带来 FWER 膨
113 胀风险，因此在选择统计分析方法和样本量估计时，需评
114 估数据间相关性，并在方案中明确描述其大小。

115 2. 统计分析模型假设

116 统计分析模型通常会基于一系列假设，需关注模型的
117 适用性，对模型假设是否成立应进行预先判断并进行事后
118 验证。关于协变量调整可参考《药物临床试验协变量校正
119 指导原则（试行）》。是否调整协变量会影响治疗效应及方
120 差的估计，样本量估计中应予以考虑。另外，当临床试验
121 中采用了分层随机时，若分层不合理，校正分层因素或采
122 用分层分析可能存在降低检验效能的风险，在设计及样本
123 量估计时应进行考虑。

124 （四） 预期治疗效应及方差

125 在基于给定的统计分析方法进行样本量估计时，需在

126 原假设和备择假设成立的情况下对各参数进行合理设置，
127 主要包括对影响预期治疗效应及其方差的相关参数设置。

128 1. 参数设置依据

129 确证性试验在进行样本量估计时各参数的设置须有充
130 分依据，一般应基于历史数据（前期研究结果或已发表的
131 数据等），并注意其临床意义与合理性。预期治疗效应可基
132 于最小临床意义差别或基于试验药的预期疗效，不应低于
133 最小临床意义差别。不建议出于减少样本量的目的，设置
134 激进的参数。当参数设置无依据（无历史数据）或参数设
135 置依据不充分（历史数据过少）时，建议开展探索试验获
136 得所需参数。

137 对于所参考的历史数据，应充分评估当前拟开展研究
138 的估计目标与历史数据相关研究估计目标之间的差异，尤
139 其是在人群、治疗（处理）、变量（终点）等方面的差异。
140 由于伴发事件的不同处理策略会直接导致人群、治疗（处
141 理）或变量（终点）定义的变化，因此还应考虑伴发事件
142 的处理策略对各参数估计的影响。当历史数据的相关研究
143 与当前研究的估计目标相同或相近时，所提供的参数较为
144 可靠；当差异较大时，建议进一步开展探索性试验。

145 对于优效性试验，应根据临床获益确定达到优效的判
146 断标准。对于非劣效性和等效试验，应考虑制定非劣效界

147 值和等效界值时所采用研究的估计目标与当前研究的差异，
148 非劣效界值的确定方法建议参考《药物临床试验非劣效设
149 计指导原则》。需要注意的是不应基于样本量的考虑进行界
150 值的确定，尤其是“需要的样本量大”不应成为选择较宽
151 松界值的理由。

152 当单臂设计采用目标值对照时，目标值的设定应基于
153 专业领域具有共识或认可程度较高的标准，当采用历史对
154 照或基于历史数据确定目标值时，应选择可靠的历史数据。
155 对于历史数据的要求可参考《单臂临床试验用于支持抗肿
156 瘤药上市申请的适用性技术指导原则》。

157 2. 伴发事件的影响

158 伴发事件的处理策略会影响到预期的治疗效应及方差。
159 参数设置时，若假定的参数或参考的历史数据未能反映伴
160 发事件的影响，则建议评估伴发事件的比例和可能带来的
161 疗效损失，尽可能降低样本量低估的风险。

162 3. 缺失数据的影响

163 当所有受试者均能为分析提供信息，且在预期治疗效
164 应和方差中已考虑了相应策略来反映伴发事件的影响时，
165 则在计算样本量时通常不需要额外增加样本量。但在临床
166 试验中，通常会存在不能为分析提供信息的受试者，因此，
167 样本量估计还应考虑缺失数据的影响，一般采用对计算所

168 得样本量增加一定比例的方法进行处理。

169 三、样本量调整

170 样本量调整是指临床试验期间对初始设计的样本量所
171 做的调整。应在方案中说明样本量调整的必要性和合理性，
172 并描述为保持盲态所采取的措施及其对 I 类错误和置信区间
173 宽度的影响（如果有）。还应在方案中明确样本量调整的时
174 间节点、样本量调整的决策规则、样本量调整的计算方法
175 以及 FWER 控制方法，即什么条件下对样本量进行调整，
176 如果要调整按照什么方法进行调整，调整后的统计分析方
177 法以及如何控制 FWER。对于涉及反推疗效的细节建议在
178 独立的期中分析计划中详述并设置访问权限，避免非必要
179 人员知晓相关内容，例如限定只有撰写样本量重新估计计划
180 和审批的统计师、DMC 及其独立统计团队有访问权限。应
181 在调整前与监管机构沟通，沟通时应提供调整依据及方法，
182 并递交相应的方案。

183 （一）样本量调整的原则

184 只有合理的样本量调整才能达到提高试验效率的目的，
185 应充分评估进行样本量调整的必要性、合理性、可行性，
186 并确保试验的完整性，样本量调整一般不建议减小样本量。

187 （1）必要性

188 不应无根据地随意调整样本量，需充分评估样本量调

189 整的必要性。当历史数据过少导致参数设置的依据不足时，
190 可通过开展探索试验获得所需的数据；如果历史数据较充
191 分可靠，相比样本量重新估计设计，固定样本量设计可在
192 保证检验效能的情况下，具有节省期中分析成本，避免期
193 中分析可能带来试验完整性破坏的风险，效率更高等优点。
194 对于非劣效性和等效性试验，阳性对照药疗效较明确缺乏
195 调整的必要性且样本量调整可能存在 I 类错误膨胀的问题，
196 建议谨慎考虑样本量调整。

197 (2) 合理性

198 样本量调整应以控制 **FWER** 和保证试验完整性为前提，
199 保证调整的合理性并根据调整方法制定正确的统计分析方
200 法。调整后的样本量应避免有统计学意义但没有临床意义
201 的情形。如果根据试验本身累积的数据进行样本量重新估
202 计，建议选择合理的调整时间节点，过早进行调整，可能
203 会面临由于数据较少导致调整所基于的参数可靠性不足；
204 过晚进行调整，可能面临有统计学意义没有临床意义以及
205 试验实施等方面的问题。

206 (3) 可行性

207 样本量调整需要考虑可行性。多次调整增加了引入实
208 施偏倚的风险，同时也增加了实施的复杂性，一般不建议
209 进行多次样本量调整；入组结束后的样本量调整会给临床

210 试验实施带来挑战，因此样本量调整需要评估试验入组进
211 度，以及数据收集和数据清理等速度。另外，如果试验终
212 点的观测需要随访的时间较长，还需要评估样本量调整对
213 试验整体时间的影响。

214 (4) 试验完整性

215 样本量调整应避免引入试验实施偏倚并保持试验的完
216 整性。当根据试验累积数据进行样本量重新估计时应严格
217 按照方案、统计分析计划和/或样本量调整计划中预先设定
218 的规则进行调整，对于非盲态的样本量重新估计应由独立
219 的数据监查委员会（DMC）及其独立统计团队完成，并确
220 保期中分析结果对研究者、受试者和申办者项目组成员的
221 盲态保持。

222 (二) 样本量调整的情形

223 常见样本量调整的情形包括：①基于外部试验数据修
224 订初始设计的样本量参数，包括治疗效应、方差、事件发
225 生率等，进而调整相应的样本量。此类调整既可以是预先
226 计划的也可以是非计划的，非计划的调整有较大风险破坏
227 试验完整性，建议谨慎考虑。②样本量重新估计，是指依
228 据方案预先设定的期中分析计划，利用累积的试验数据重
229 新计算样本量，以保证最终的统计检验能达到预先设定的
230 目标或修改后的目标，并同时能够控制 FWER。为了降低
231 引入操作偏倚和破坏试验完整性的风险，不建议在试验期

232 间增加非事先计划的样本量重新估计。样本量重新估计分
233 为盲态样本量重新估计和非盲态样本量重新估计两种。相
234 比盲态样本量重新估计，非盲态样本量重新估计可能引入
235 的操作偏倚、FWER 膨胀、及破坏试验完整性的风险更高，
236 在设计和实施时更需谨慎。

237 1. 盲态样本量重新估计

238 对于盲态样本量重新估计，针对样本量重新估计的期
239 中分析不对数据揭盲，不使用试验分组信息，不做任何涉
240 及组间疗效比较的分析，不能基于试验收集到的数据对治
241 疗效应大小进行重新估计。通常是在假定初始设定的组间
242 差异不变的情况下对其他参数进行重新估计来调整样本量，
243 主要针对初始设计时除预期组间差异以外的其他参数（例
244 如事件发生率、变异等）的不确定性而进行的调整。期中
245 分析时因不涉及组间疗效比较，一般不需要调整 I 类错误率，
246 且因是在盲态下进行的，所以引入操作偏倚的风险较低。

247 2. 非盲态样本量重新估计

248 对于非盲态样本量重新估计，针对样本量重新估计的
249 期中分析使用试验分组信息，分析内容涉及组间疗效比较。
250 样本量重新估计时，通常都会结合可操作性和最小临床意
251 义差别设置最大样本量，调整后的样本量不应大于最大样
252 本量。

253 常见的样本量调整决策规则有：①基于条件检验效能
254 或试验成功率，当条件检验效能或试验成功率落在某个区
255 间时对样本量进行重新估计，否则不对样本量进行调整；
256 ②基于期中分析的组间治疗效应差异，当差异落在某个区
257 间时对样本量进行重新估计，否则不对样本量进行调整。

258 与常用调整决策规则相对应的计算调整后样本量的方
259 法通常有①基于条件检验效能或试验成功率，使得根据调
260 整后的样本量计算所得条件检验效能或试验成功率达到预
261 先设定的要求；②基于期中分析得到的治疗效应与方案初
262 始设计的治疗效应的比值调整样本量。

263 常见的 FWER 控制策略包括改变检验统计量（例如加
264 权统计量方法及基于独立 P 值的方法等）、限定样本量调整
265 的条件（例如希望区域法）以及调整假设检验拒绝域或名
266 义检验水准（例如极大似然比方法及条件误差函数方法）
267 等。

268 四、 试验实施过程中的考虑

269 申办者应按照计划的样本量完成临床试验，除了出于
270 伦理原因提前终止或出于检验效能不再可接受而放弃外，
271 原则上不得随意终止试验。试验过程中不得随意增加或减
272 少样本量，样本量调整也需事先在方案中明确并按照方案
273 进行实施。如果出于实施层面的原因，导致实际收集到的

274 样本量与方案预先设定的样本量存在较大不一致，应就其
275 对有效性和安全性等评价的影响进行评估。

276 盲态的样本量重新估计一般不会引入操作偏倚，可由
277 申办者或独立统计团队完成。非盲态的样本量重新估计需
278 由 DMC 及其独立统计团队完成。由于涉及非盲的数据和结
279 果，期中分析的执行应是一个完全保密的过程。参与试验
280 实施的所有人员（包括申办者所雇佣的人员）应当对这些
281 分析结果保持盲态，否则会导致招募患者的特征改变或产
282 生治疗组间比较的偏倚。研究者应仅被告知继续或终止试
283 验的决定，或实施修订试验程序的决定。DMC 向申办者提
284 出的建议中也应避免提供可以反推疗效的相关内容，以保
285 证试验的完整性。

286 考虑到样本量重新估计的复杂性，尤其是结合其他适
287 应性设计和多重性的场景时，通常需要模拟确定调整的时
288 间点和决策规则及其他运行特性，以证明这些规则满足样
289 本量调整所需满足和达到的要求，如控制 FWER、达到检
290 验效能、避免过早或者在入组结束后进行样本量调整等。

291 五、其他

292 （一）基于贝叶斯方法的样本量估计

293 基于贝叶斯方法的样本量估计除先验分布的设置外，
294 其余考虑因素与传统样本量估计考虑的因素一致。设计时

295 应在全面严谨的模拟研究基础上，充分评估先验信息的合
296 理性，确定合理的拒绝原假设的决策规则，使方法满足控
297 制 FWER 的要求。建议同时使用其他合理的先验信息作为
298 敏感性分析。先验信息的信息量可以通过先验的有效样本
299 量进行衡量。

300 (二) 基于其他目的的样本量估计

301 样本量一般基于主要目标来估计，如依据合理，也可
302 以根据安全性评价或重要的次要估计目标估计样本量，估
303 计方法须在方案中详细描述，且据此所确定的样本量不比
304 基于主要估计目标所需样本量小。

305 (三) 样本量的敏感性计算

306 影响样本量估计的因素众多，而历史数据通常相对有
307 限，在进行样本量估计时，通常需进行各种偏离假设的计
308 算，通过一系列敏感性计算，使各参数在一定合理范围内
309 取值，或使参数取值来自某一分布，提供偏离假设的样本
310 量范围，可在一定程度上降低样本量估计的不确定性。出
311 于谨慎的考虑，必要时申办者可以基于敏感计算结果，结
312 合保守的原则，用以指导确定合理的样本量。

313 (四) 样本量重新估计与其他适应性设计的结合

314 当样本量重新估计与其他适应性设计相结合或与其他

315 多重性问题相结合时，由于设计的复杂性，尤其需要考虑
316 FWER 控制问题和较高的破坏完整性的风险，需要慎重考
317 虑是否有必要进行多重适应性调整，并在确认必要性后，
318 严格遵守适应性设计合理性、完整性和可行性的原则进行
319 设计和执行，须进行全面严谨的理论或模拟研究，并与监
320 管机构进行全面细致的沟通。样本量重新估与其他适应性
321 设计相结合常见以下两种情形。

322 (1) 与成组序贯设计相结合

323 该方法是指在期中分析时，预设优效性终止试验和/或
324 无效性终止试验决策规则的同时预设样本量重新估计的决
325 策规则。一般不建议在样本量重新估计后设置有效性提前
326 终止的期中分析，若基于合理充分的理由，需慎重评估所
327 采用的样本量重新估计方法及分析时间点设定是否仍然可
328 控制 FWER，并评估统计分析方法的合理性。

329 (2) 与 II/III 期无缝剂量选择相结合

330 该方法是指根据 II 期期中分析结果对于选中的剂量预
331 设样本量重新估计的决策规则。需事先在方案中明确 I 类错
332 误控制措施，并就统计分析与监管机构进行沟通。

333 (五) 与监管机构的沟通

334 鼓励申办者在研究开始前，与监管机构就关键性临床
335 试验中的样本量进行沟通。建议申办者在沟通时对试验设

336 计、检验水准及要达到的检验效能、统计分析方法、预期
337 效应及方差参数设置和样本量计算方法等进行详细说明。
338 如采用模拟的方法进行样本量估计，建议递交模拟代码和
339 模拟方法的详细说明，包括但不限于模拟参数的假设及依
340 据、模拟的种子数、模拟次数等。

341 对于非盲的样本量重新估计，建议在沟通方案时一并
342 递交（单独的）样本量重新估计计划，该计划至少应该包
343 括进行样本量重新估计的原因、样本量计算方法、样本量
344 重新估计的决策规则以及必要的计算或模拟代码及结果、
345 样本量重新估计的时间节点和最大样本量、FWER 控制、
346 实施中盲态保持和为避免根据调整样本量而反推期中分析
347 疗效结果的相关措施。

348 六、参考文献

349 [1] 国家药品监督管理局药品评审中心. 药物临床试验多
350 重性问题指导原则（试行）. 2020

351 [2] 国家药品监督管理局药品评审中心. 药物临床试验非
352 劣效设计指导原则. 2020

353 [3] 国家药品监督管理局药品评审中心. 药物临床试验盲
354 法指导原则（试行）. 2022

355 [4] 国家药品监督管理局药品评审中心. 药物临床试验随
356 机分配指导原则（试行）. 2022

- 357 [5] 国家药品监督管理局药品评审中心. 药物临床试验适
358 应性设计指导原则 (试行) . 2021
- 359 [6] 国家药品监督管理局药品评审中心. 药物临床试验数
360 据管理与统计分析计划指导原则. 2021
- 361 [7] 国家药品监督管理局药品评审中心. 药物临床试验协
362 变量校正指导原则 (试行) . 2020
- 363 [8] 国家药品监督管理局药品评审中心. 药物真实世界研
364 究设计与方案框架指导原则 (试行) . 2023
- 365 [9] CCTS 工作组, 陈平雁 执笔. 临床试验中样本量确定
366 的统计学考虑, 中国卫生统计, 2015, 32(4):727-733.
- 367 [10] Campbell MJ, Donner A, Klar N. Developments in
368 cluster randomized trials and statistics in medicine. *Statistics in*
369 *Medicine*, 2007, 26(1): 2–19.
- 370 [11] Chen JYH, DeMets DL, Lan GKK. Increasing the
371 sample size when the unblinded interim results is promising.
372 *Statistics in Medicine*, 2004, 23(7):1023-1038.
- 373 [12] Chow SC, Chang M. *Adaptive Design Methods in*
374 *Clinical Trials*. CRC Press, 2nd edition 2011.
- 375 [13] Chow SC, Shao J, Wang H, Lockhnygina Y. *Sample*
376 *size calculations in clinical research*. Chapman & Hall/CRC, 3rd
377 edition 2018.

378 [14] Chuang-Stein, C. Sample size and the probability of a
379 successful trial. *Pharmaceutical Statistics*, 2006, 5(4): 305–309.

380 [15] Chuang-Stein C, Kirby S, Hirsch I, et al. The role of
381 the minimum clinically important difference and its impact on
382 designing a trial. *Pharmaceutical Statistics*, 2011, 10(3): 250–
383 256.

384 [16] Ciarleglio MM, Arendt CD, Makuch RW, et al.
385 Selection of the treatment effect for sample size determination in
386 a superiority clinical trial using a hybrid classical and Bayesian
387 procedure. *Contemporary Clinical Trials*, 2015, 41:160–171.

388 [17] Cook JA, Julious SA, Sones W, et al. DELTA2
389 guidance on choosing the target difference and undertaking and
390 reporting the sample size calculation for a randomised
391 controlled trial. *BMJ*, 2018, 363: k3750.

392 [18] Copay AG, Subach BR, Glassman SD, Polly DWJr,
393 Schuler TC. Understanding the minimum clinically important
394 difference: a review of concepts and methods. *The Spine Journal*,
395 2007,7(5):541-6.

396 [19] Cui L, Hung HMJ, Wang SJ. Modification of sample
397 size in group sequential clinical trials. *Biometrics*, 1999,
398 55(3):853-857.

399 [20] EMA. Guideline on the choice of the non-inferiority
400 margin. 2006.

401 [21] Friede T, Kieser M. On the inappropriateness of an
402 EM algorithm based procedure for blinded sample size re-
403 estimation. *Statistics in Medicine*, 2002, 21(2):165-176.

404 [22] Friede T, Kieser M. Blinded sample size reassessment
405 in non-inferiority and equivalence trials. *Statistics in Medicine*,
406 2003, 22(6):995–1007.

407 [23] Friede T, Kieser M. Blinded sample size re-estimation
408 in superiority and noninferiority trials: bias versus variance in
409 variance estimation. *Pharmaceutical Statistics*, 2013, 12(3):
410 141–146.

411 [24] Friede T, Pohlmann H, Schmidli H. Blinded sample
412 size reestimation in event- driven clinical trials: Methods and
413 an application in multiple sclerosis. *Pharmaceutical Statistics*.
414 2019,18(3): 351–365.

415 [25] Gould AL. Interim analyses for monitoring clinical
416 trials that do not materially affect the type I error rate. *Statistics*
417 *in Medicine*, 1992,11(1):55-66.

418 [26] Gould AL. Planning and revising the sample size for a
419 trial. *Statistics in Medicine*. 1995, 14(9):1039-51.

420 [27] Gould AL. Shih WJ. Sample size re-estimation
421 without unblinding for normally distributed outcomes with
422 unknown variance. Communications in Statistics - Theory and
423 Methods, 1992, 21(10):2833-2853.

424 [28] Hemming K, Eldridge S, Forbes G, Weijer C, Taljaard
425 M. How to design efficient cluster randomised trials. BMJ 2017,
426 358:j3064.

427 [29] ICH E8: General Considerations for Clinical Trials.
428 1997.

429 [30] ICH E8(R1) step 5: General Considerations for
430 Clinical Trials. 2022.

431 [31] ICH E9: Statistical Principles for Clinical Trials. 1998.

432 [32] ICH E9R1: Addendum on estimands and sensitivity
433 analysis in clinical trials to the guideline on statistical principles
434 for clinical trials. 2019.

435 [33] Kevin K, Michael JG, Kim ML, et al. A review of
436 Bayesian perspectives on sample size derivation for
437 confirmatory trials. The American Statistician, 2021, 75:4, 424-
438 432.

439 [34] Kieser M, Friede T. Simple procedures for blinded
440 sample size adjustment that do not affect the type I error rate.

441 Statistics in Medicine, 2003, 22(23): 3571-3581.

442 [35] Stephen Lake S, Kammann E, Klar N, Betensky R.
443 Sample size re-estimation in cluster randomization trials.
444 Statistics in Medicine, 2002; 21(10):1337–1350.

445 [35] Mehta R, Pocock SJ. Adaptive increase in sample size
446 when interim results are promising: A practical guide with
447 examples. Statistics in Medicine, 2011, 30(28): 3267-3284.

448 [36] Michael A. Proschan MA. Two-Stage Sample Size Re-
449 Estimation Based on a Nuisance Parameter: A Review, Journal
450 of Biopharmaceutical Statistics. 2005, 15(4):559-574

451 [37] O’Hagan A, Stevens JW, Campbell MJ. Assurance in
452 clinical trial design. Pharmaceutical Statistics 2005, 4(3):187–
453 201.

454 [38] Robertson DS, Choodari Oskooei B, Dimairo M,
455 Flight L, Pallmann P, Jaki T. Point estimation for adaptive trial
456 designs I: A methodological review. Statistics in Medicine 2023,
457 42(2): 122-145.

458 [39] Robertson DS, Choodari Oskooei B, Dimairo M,
459 Flight L, Pallmann P, Jaki T. Point estimation for adaptive trial
460 designs II: Practical considerations and guidance. Statistics in
461 Medicine 2023, 42(14): 2496-2520.

462 [40] Rutterford C, Copas A, Eldridge S. Methods for
463 sample size determination in cluster randomized trials.
464 International Journal of Epidemiology, 2015, 44(3):1051-67.

465 [41] Schulz KF, Grimes DA. Sample size calculations in
466 randomised trials: mandatory and mystical. Lancet, 2005,
467 365:1348-53.

468 [42] Shih WJ. Sample size re-estimation – a journey for a
469 decade. Statistics in Medicine, 2001, 20:515-518.

470 [43] Shih WJ, Li G, Wang Y. Methods for flexible sample-
471 size design in clinical trials: Likelihood, weighted, dual test, and
472 promising zone approaches. Contemporary Clinical Trials, 2016,
473 47: 40-48.

474 [44] Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian
475 Approaches to Clinical Trials and Health- Care Evaluation.
476 John Wiley & Sons, 2004.

477 [45] US FDA. FDA Demonstrating substantial evidence
478 guideline – guidance for the industry. 2022.

479 [46] Waksman JA, Assessment of the Gould-Shih
480 procedure for sample size re-estimation, Pharmaceutical
481 Statistics, 2007, 6(1): 53-65.

482 [47] Wang J, Ren Q. A Note on the Promising Zone

483 Approach in Adaptive Trial Design, Statistics in
484 Biopharmaceutical Research, 2022, 14(1):132-137.

485 [48] Zucker DM, Wittes JT, Schabenberger O, et al.
486 Internal pilot studies II: comparison of various procedures.
487 Statistics in Medicine, 1999, 18(24): 3493-3509.

488 [49] Friede T, Kieser M. Blinded sample size reassessment
489 in non-inferiority and equivalence trials. Statistics in Medicine,
490 2003, 22(6): 995-1007.

491 [50] ICH E1a: The Extent of Population Exposure to
492 Assess Clinical Safety: For Drugs Intended for Long-term
493 Treatment of Non-Life-Threatening Conditions. 1995.

494 附录 1 名词解释

495 **I类错误 (Type I Error):** 指原假设 (或称无效假设) 正确但检验结果拒
496 绝了原假设的错误, 相当于把实际上无效的药物经统计推断得出有效结论的错
497 误。其概率需控制在某一水平, 该水平称为检验水准, 或称显著性水准, 用 α 表
498 示。

499 **II类错误 (Type II Error):** 指原假设不正确, 但检验结果未能拒绝原假
500 设的错误, 相当于把实际上有效的药物经统计推断得出无效结论的错误。

501 **名义检验水准 (Nominal Level):** 对于多重检验中某一假设检验的检验水
502 准称之为名义检验水准, 又称局部检验水准, 用 α_i 表示。

503 **总I类错误率 (Familywise Error Rate, FWER):** 是指在同一临床试验所
504 关注的多个假设检验中, 至少一个真的原假设被拒绝的概率。其应控制在合理
505 水平。

506 **多重性问题 (Multiplicity Issues):** 指在一项完整的临床试验中, 需要经
507 过不止一次统计推断 (多重检验) 对研究结论做出决策的相关问题。

508 **估计目标 (Estimand):** 对治疗效应的精确描述, 反映了针对临床试验目
509 的提出的临床问题。它在群体水平上汇总比较相同患者在不同治疗条件下的结
510 局。

511 **伴发事件 (Intercurrent Event):** 治疗开始后发生的事件, 可影响与临床
512 问题相关的观测结果的解释或存在。在描述相关临床问题时, 需解决伴发事件,
513 以便准确定义需要估计的治疗效应。

514 **预期效应/差异 (Treatment Effects of Interest, Treatment Difference to**

515 **be Detected**): 研究旨在有一定检验效能去发现的关于分析指标的组间差异大小。

516 **样本量重新估计 (Sample Size Re-estimation)**: 是指依据预先设定的期中
517 分析计划, 利用累积的试验数据重新计算样本量, 以保证最终的统计检验能达
518 到预先设定的目标或修改后的目标, 并同时能够控制I类错误率。

519 **真实世界研究 (Real-World Research/Study, RWR/RWS)**: 指针对预设
520 的临床问题, 在真实世界环境下收集与研究对象健康状况和/或诊疗及保健有关
521 的数据 (真实世界数据) 或基于这些数据衍生的汇总数据, 通过分析, 获得药
522 物的使用情况及潜在获益-风险的临床证据 (真实世界证据) 的研究过程。

中文	英文
准确性	Accuracy
条件检验效能	Conditional Power
数据监查委员会	Data Monitoring Committee, DMC
效应量	Effect Size
有效样本量	Effective sample size, ESS
估计目标	Estimand
总 I 类错误率	Familywise Error Rate, FWER
风险率	Hazard Rate
完整性	Integrity
伴发事件	Intercurrent Event
期中分析	Interim Analysis
群内相关性	Intraclass correlation
名义检验水准	Nominal Level
最小临床意义差别	Minimum Clinically Important Difference
多重性	Multiplicity
试验成功率	Probability of Success
等比例风险	Proportional Hazards
检验效能	Power
真实世界研究	Real World Research/Study, RWR/RWS
可靠性	Reliability
检验水准	Significance Level
样本量调整	Sample Size Adjustment

中文	英文
样本量重新估计	Sample Size Re-estimation
敏感性计算	Sensitivity calculations
统计分析计划	Statistical Analysis Plan, SAP
预期效应/差异	Treatment Effects of Interest, Treatment Difference to be Detected
时间-事件数据	Time-to-event data
I 类错误	Type I Error
II 类错误	Type II Error
合理性	Validity

524

525