

药物临床试验富集策略与设计指导原则

(征求意见稿)

2020年8月

目 录

一、概述	1
二、富集策略与设计的适用性	2
三、常用的富集策略与设计	3
(一) 同质化富集	3
(二) 预后型富集	4
(三) 预测型富集	6
(四) 复合型富集	11
(五) 适应性富集	11
四、富集策略与设计的相关考虑	12
(一) 标志物检测的灵敏度和特异度	12
(二) 是否纳入标志物阳性和阴性的受试者	13
(三) 入选人群和分析集	14
(四) 筛选富集人群对优效和非劣效试验的不同影响	14
(五) 控制 I 类错误	15
五、监管考虑	16
(一) 明确界定富集人群	16
(二) 不应忽视非富集人群的疗效	16
(三) 预先确定研究方案并与监管机构沟通	16
六、参考文献	17
附录 1: 词汇表	23
附录 2: 中英文对照表	25
附录 3: 富集设计的研究案例	28
示例 1: 预后型富集——心血管病研究	28
示例 2: 预测型富集——黑色素瘤研究	28
示例 3: 随机撤药设计——普瑞巴林用于治疗纤维性肌痛症的研究	29
示例 4: 预测型富集——MSI 研究	30

1 药物临床试验富集策略与设计指导原则

3 一、概述

4 临床试验的目的是在入组的受试者中验证试验药物的有效
5 性和安全性。但实际上，由于受试者病理生理学特点和药物作用
6 机理的复杂性，不同受试者的药物治疗效果不尽相同，从而影响
7 临床试验的效率。为了入组能够从试验药物中获益最大化的受试
8 者，以提高临床试验的效率，富集策略的概念应运而生。

9 富集是指在临床试验中根据受试者的某些特征（如人口学、
10 病理生理学、组织学、基因组和蛋白质组学等）前瞻性地精准定
11 义从试验药物中获益最大化的目标人群。在临床试验有多种选择
12 受试者的富集策略，例如，可以选择因具有一定特征而对研究药
13 物最有可能应答的受试者，也可以选择那些对现有药物治疗效果
14 不明显而可能对试验药物敏感的受试者，或者单纯选择更容易出
15 现终点事件的受试者，等等。

16 本指导原则阐述了常用的富集策略与设计的原理与方法、各
17 自的优缺点，并从实际应用和监管角度说明需要考虑的关键问题。
18 本指导原则中，“富集策略”主要是指随机对照临床试验中用于
19 选择最有可能获益的受试者的方法，但也可以扩展到使用外部
20 （历史或平行）对照的单臂试验。

21 本指导原则主要适用于以支持药品注册上市为目的的确证
22 性临床试验，也可供以非注册为目的的临床试验参考使用。

二、富集策略与设计的适用性

广义来讲，所有临床试验设计都在运用富集的概念，这主要反映在受试者的入选和排除标准上，其目的是尽可能入选对试验药物有应答的受试者，从而提高临床试验的效率。例如，研究降胆固醇药物降低心血管事件发生率时，临床试验只入组血液中总胆固醇浓度高于某一阈值的患者。实际上，根据疾病领域、药物作用机理以及受试者的应答情况等，可选择不同的富集策略和设计，其适用性主要从科学上的有效性、试验结果的可解释性和医疗实践中的可推广性等方面考虑。

1. 科学上的有效性：包括筛选受试者有科学依据，筛选工具的灵敏度和特异度符合一定要求，试验设计时使用避免偏倚的措施（如随机、盲法等），以及 I 类错误的控制等。

2. 试验结果的可解释性：指试验药物在富集人群中的疗效可以从疾病的病理生理学、基因组学、遗传学或者药物作用机理等方面进行解释；如果限于生物学、医学或者药理学等方面的知识而无法解释，则试验药物在富集人群中的疗效需要具有一定程度的重现性。

3. 医疗实践中的可推广性：包括富集策略能够在临床实践中被广泛的使用，以便及时、准确地识别对试验药物有应答或敏感的患者。有时，由于筛选患者的方法复杂、灵敏度偏低、成本高昂等原因而使其无法普及，或者筛选方法耗时较长而在治疗开始时无法富集患者，这些都会影响富集策略和方法的可推广性。

45 三、常用的富集策略与设计

46 根据临床试验关注的主要问题和实施过程,可以选择不同的
47 富集策略,主要包括同质化富集、预后型富集、预测型富集、复
48 合型(预后型和预测型相结合的)富集和适应性富集五种策略类
49 型。

50 实际应用中,通常依据与药物作用机理相关的标志物选择富
51 集策略与设计。这里的“标志物”定义为与受试者预后或药物治
52 疗应答有关的流行病学因素(如人口学)、既往病史、家族史、
53 临床观测变量(如疾病严重程度)、实验室检查(如病理生理学、
54 药物代谢)、基因组学和蛋白质组学等各种特征变量。根据标志
55 物的不同作用,可分为预后型、预测型和混合型标志物。另外,
56 在有些疾病领域,可能没有明显的标志物,这时一般根据受试者
57 在筛选期间对治疗的应答情况,或者其他临床试验的数据以及文
58 献报道选择富集的受试者。

59 (一) 同质化富集

60 同质化富集是指通过减少受试者间的异质性以提高临床试
61 验的检验效能的一种研究策略。减少异质性最简单且实用的方法
62 是尽量选择病情稳定的受试者,同时对入选受试者进行精准定义,
63 并对疾病的状态和有关变量进行精确测量。例如,在高血压药物
64 试验中,为了筛选出血压相对稳定的受试者,可以在入组前对受
65 试者的血压进行一段时间的观察,以排除血压变化较大的受试者。

66 一般来说,为了更加准确地定义富集人群,除常规的入选和

67 排除标准外，还要考虑以下几方面问题：

68 (1) 入选标准：更仔细地定义入选标准，以确保入选受试
69 者间的基线特征具有较好的一致性。

70 (2) 排除标准：①对安慰剂过于敏感的受试者；②基线检
71 测结果不稳定的受试者，如在初筛期病情或症状不稳定的受试者；
72 ③伴随某种疾病可能导致过早死亡的受试者；④服用与试验药具
73 有相似治疗作用的药物的受试者；⑤可能无法耐受试验药物治疗
74 的受试者；⑥可能因并发症提前退出研究的受试者。

75 (3) 依从性：应尽可能入选依从性好的受试者，即选择不
76 会因为非医学原因（如不便前往研究地点）而退出的受试者，以
77 及能够坚持按照试验方案进行治疗的受试者，从而减少由于受试
78 者过多退出或使用不同的治疗方法而导致的差异。对患者的依从
79 性识别和选择必须在随机化分组之前进行。

80 (4) 培训：研究人员和临床试验协调员应接受相关培训，
81 确保严格按照方案入选和排除受试者，并按照方案进行研究。

82 (二) 预后型富集

83 预后型富集是指通过对预后型标志物的识别，入选更有可能
84 观察到终点事件或疾病进展的高风险人群（特指更容易出现预后
85 结局或疾病进展的人群），以增加检验效能的一种策略。该策略
86 主要增加试验的绝对效应，而非相对效应。例如，在一项降低终
87 点事件发生率的临床试验中，经过一段时间的治疗，高风险人群
88 的终点事件发生率由 10%降低至 5%，低风险人群的终点事件发

89 生率由 1%降低至 0.5%，虽然两者相对效应均降低 50%，但前者
90 显然需要较少的样本量或者较短的随访时间就可以观察到试验
91 药物的绝对疗效。常用的预后型富集设计有以下两种。

92 1. 基于终点事件的富集设计

93 在以降低终点事件发生率为主要评价指标的研究中，一般认
94 为有效的试验药物在高风险人群中能够减少或者避免发生更多
95 的终点事件。因此，应考虑在高风险人群中招募受试者。通常，
96 在样本量不变时，与低风险人群相比，高风险人群更容易发生终
97 点事件，经治疗后终点事件发生率降幅较大，因此检验效能更高。
98 这种策略经常用在抗肿瘤和治疗心血管疾病的药物研究中，例如
99 在基因 BRCA1/2 突变的女性人群中进行乳腺癌或卵巢癌预防的
100 研究；又如在降血脂药物的研究中，选择血液中高含量低密度脂
101 蛋白（LDL）、低含量高密度脂蛋白（HDL）和高含量 C 反应蛋
102 白（CRP）的患者试验。在有些疾病领域，如阿尔兹海默症和各
103 种癌症药物研究中，也可以通过基因组或蛋白质组学筛选高风险
104 的患者。

105 2. 基于疾病进展的富集设计

106 预后型富集设计也可用于研究能够减缓疾病进展的试验药
107 物，如在对阿尔茨海默症、帕金森病、类风湿性关节炎、慢性阻
108 塞性肺疾病和恶性肿瘤等疾病开展药物临床试验时，可以选择疾
109 病进展可能较快的受试者。例如，在类风湿性关节炎的患者中，
110 具有以下特征的患者疾病往往进展较快：类风湿因子阳性、具有

111 某些临床特征（如多关节受影响、关节以外的病症、皮下结节、
112 活动受限等）以及实验室指标异常（如血红蛋白降低）等；又如
113 在慢性阻塞性肺疾病（COPD）的患者中，具有近期发作史（过
114 去一年中至少发作一次）或血浆纤维蛋白原较高的患者疾病进展
115 较快。在抗肿瘤药物研究中，常见的预后标志物包括组织学分级、
116 血管浸润、分子亚型以及转移性肿瘤结节等指标。

117 需要注意的是，如果预后型标志物与试验药物之间存在交互
118 作用，即试验药物对该标志物阳性和阴性患者都有效果，但疗效
119 不同，则该预后型标志物也可以起到预测的作用，此类标志物通
120 常称为混合型标志物。

121 （三）预测型富集

122 预测型富集是指根据受试者的生理或疾病特征选择对试验
123 药物最可能有应答的受试者，以提高试验效率的一种研究策略。
124 例如在靶向治疗中，可根据药物相关的靶向基因或蛋白，或生理
125 机能（如肾素型高血压/低血压、慢性心力衰竭时的射血分数），
126 选择可能有应答的受试者。采用这一策略既能增加试验药物的绝
127 对效应，也能增加其相对效应，因此能够以较小的样本量获得较
128 高的检验效能。当患有某种疾病的受试者中只有一小部分对试验
129 药物有应答时（如只有部分受试者具有药物作用的受体），使用
130 这种富集策略十分有效。在实践中，既可以基于研究者对疾病的
131 认识（如各种标志物）选择受试者，也可以根据以往的试验数据
132 和结果。

133 1. 基于病理生理学特征的富集设计

134 疾病的病理生理学特征可以提示对试验药物有更好应答的
135 受试者。基于病理生理学的富集指标可以是生物标志物（如影响
136 肿瘤生长的基因突变、基因/蛋白表达水平）、影像学特征，以及
137 与疾病表型相关的一些人口学（如年龄、种族等）或临床特征（如
138 疾病分期、分型等）。根据富集标志物的性质，可将其分为以下
139 几类：

140 (1) 基因或蛋白标志物

141 治疗肿瘤的药物通常针对肿瘤细胞表面或细胞内相关的受
142 体、酶、激素或其他内源活性物质，对此可根据一个或多个相应
143 的基因或蛋白标志物选择富集人群。例如曲妥珠单抗主要针对人
144 表皮生长因子受体 2（HER-2）蛋白阳性的乳腺癌患者；也有一些
145 细胞受体最初作为蛋白标志物，但后来被确认为肿瘤基因标志
146 物（如 EGFR 和 BRAF 基因突变），并用该基因标志物定义病理
147 生理状态、选择最有可能获益的受试者。

148 当在富集设计中使用基因或蛋白标志物时，标志物检测的准
149 确性和精确性至关重要。如果诊断检测不准确，不仅会影响富集
150 的效果导致检验效能降低，而且在非劣效试验中可能会增加 I 类
151 错误。同时，还要尽可能地考虑标志物与疗效之间的关系，从而
152 有助于选择合适的受试者。

153 (2) 药物代谢物

154 不同受试者对试验药物的代谢能力不同，入组能够产生足够

155 数量活性代谢物的受试者，可以提高临床试验的效率。在某些情
156 况下，给予产生活性物质能力较弱的患者更高的剂量，有助于他
157 们产生足够的活性物质，从而更可能观察到药物的疗效。然而，
158 对于完全不能代谢出有效活性成分的患者，也应该纳入试验或事
159 先设定的分析计划中，目的是对指定的代谢物进行验证和获益-
160 风险分析。

161 (3) 肿瘤代谢物

162 抗肿瘤药物试验可以通过检测组织或血液中的肿瘤代谢物
163 含量选择受试者。例如，只入选那些代谢反应较强的受试者，或
164 者在肿瘤患者中按代谢反应程度进行分组，并对代谢反应较强的
165 受试者进行主要分析。

166 2. 基于对试验药物应答证据的富集设计

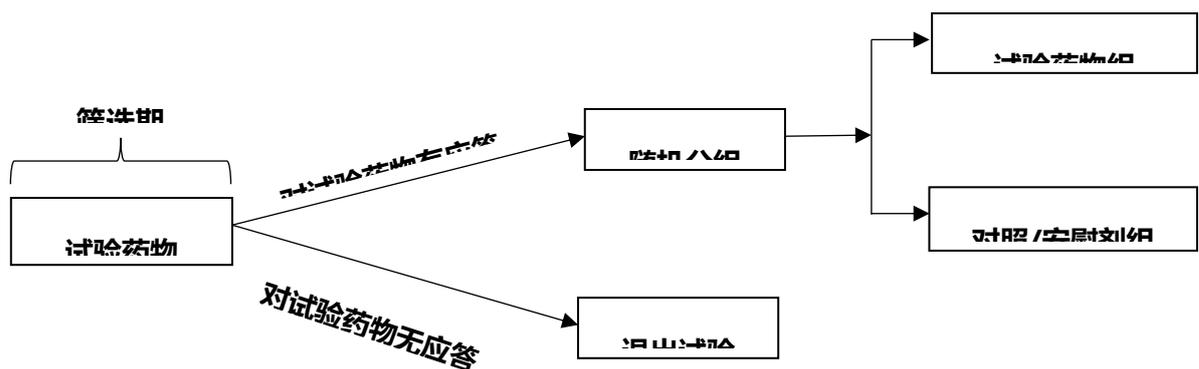
167 此类富集设计根据受试者在筛选期内对试验药物（或既往对
168 类似药物）的应答情况，选择可能合适的受试者。

169 (1) 筛选有应答的受试者

170 对于在研究开始前无法根据标志物识别出可能对试验药物
171 有应答的受试者的临床试验，需要设置合理的筛选期，将试验药
172 物用于所有受试者。根据事先确定的主要终点或替代终点来筛选
173 治疗有效的受试者；然后，将筛选期内对试验药物有应答的受试
174 者进行随机分组。例如，某降胆固醇药物的试验可分两个阶段进
175 行，第一阶段将试验药物用于入组的高胆固醇患者，经过一段时
176 间治疗后通过生化检查判断患者对药物是否有应答（即胆固醇降

177 低); 第二阶段则仅将对药物有应答的患者随机分配到试验组和
178 对照组。采用这种先筛选、后随机的方法可以首先入选较多的受
179 试者, 对这些受试者还可以按筛选期的应答程度进行分层, 将应
180 答程度相对较好的亚组作为主要分析人群, 同时可以进一步探索
181 应答程度相对较弱的受试者。然而, 这种设计并不适用于停药后
182 药物作用持续时间相对较长或停药后会对受试者造成伤害的药
183 物的研究。

184 选择有应答的受试者还可以使用随机撤药设计。该设计一般
185 分为两个阶段, 即第一阶段测试受试者是否对试验药物有应答
186 (可以使用单臂开放试验或随机对照试验), 第二阶段对试验药
187 物有应答的受试者, 随机分到试验组(继续使用试验药物)或安
188 慰剂组(将试验药物撤出), 无应答的受试者退出试验(图 1)。
189 判断患者对试验药物是否有应答, 可以根据某些替代指标如症状、
190 体征、实验室检查和疾病复发等进行评估。



191

192

图 1 随机撤药研究流程图

193

194

随机撤药设计由于筛选了对试验药物有应答的受试者, 从而
提高了临床试验的效率; 同时, 可利用已经入组的受试者研究药

195 物的长期疗效或安全性, 以及利用退出试验的受试者研究撤药效
196 应。另一方面, 这种设计更符合伦理学要求, 即一旦治疗失败可
197 及时终止试验, 可用于儿童药物研究。该设计的主要缺点是可能
198 具有残留效应以及研究时间相对较长。

199 (2) 基于历史数据或文献报道确定入选受试者

200 根据既往研究中确定的亚组特征入选受试者, 即在总体人群
201 中很少或几乎没有观察到明显的治疗效果, 但通过分析特定的亚
202 组人群可能获得显著疗效, 最后仅在亚组人群中招募受试者。例
203 如, 复方硝酸异山梨酯/盐酸肼屈嗪是一种治疗严重心力衰竭的
204 药物, 之前的研究发现其对非裔美国人的治疗效果明显优于高加
205 索人, 在后续设计的随机安慰剂对照试验中, 入选了 1050 名心
206 力衰竭的非裔美国人患者, 由此验证了该复方在该心力衰竭患者
207 中的有效性。

208 3. 基于对现有药物无应答的富集设计

209 在富集设计中, 除上述选择对试验药物有应答的受试者外,
210 还可以考虑选择对现有药物无应答的受试者, 目的是更好地显现
211 出试验药物的治疗效果。

212 无应答的富集设计适用于满足一定条件的临床试验, 即试验
213 药物与现有药物具有不同的作用机制, 或试验药物的疗效至少略
214 优于现有药物。如果对受试者未加选择, 则需要较大的样本量才
215 能显示出试验药物的疗效; 相反, 如果只选择对现有药物无应答
216 的受试者, 由于对照组的应答率很低, 则可能只需要较小的样本

217 量，就能得出试验组优于对照组的结论。必须指出的是，对于某
218 些可能危及生命、并不断进展的疾病来说，采用无应答的富集设
219 计会将受试者随机分配到无应答的对照组，可能存在伦理问题。

220 (四) 复合型富集

221 复合型富集指同时使用多个标志物(如预后型和预测型标志
222 物)以减少受试者异质性的富集策略。对有些疾病领域，疾病的
223 发生、发展和预后机制复杂、个体异质性高，使用单一标志物不
224 大可能富集最有可能获益的受试者，而使用复合标志物(例如综
225 合评分)进行富集可以有效地降低受试者的异质性，从而提高试
226 验效率。

227 需要注意的是，使用复合标志物评分时应列出其构成的单个
228 标志物并阐明它们之间的关系；如果对不同的单个标志物赋予不
229 同的权重，应详细说明其生物学上的原理。

230 (五) 适应性富集

231 适应性富集策略是指按照预先制定的计划，根据临床试验期
232 中分析结果，在保证试验的合理性和完整性的前提下，对临床试
233 验方案进行修改，如改变入组标准或仅纳入一个亚组的受试者，
234 并对样本量进行调整等。

235 当试验药物在标志物阳性和阴性的受试者中的疗效不确定
236 时，试验可以同时入组标志物阳性和阴性的受试者，根据期中分
237 析结果适应性地调整需要入选的受试者。当主要终点规定为标志
238 物阳性受试者的疗效时，如果期中分析结果显示标志物阴性受试

239 者的疗效远低于标志物阳性受试者, 则应减少或完全停止标志物
 240 阴性的受试者入组 (图 2)。假定标志物阳性受试者的疗效高于
 241 标志物阴性受试者的疗效, 也可以考虑首先入组标志物阳性的受
 242 试者, 如果期中分析结果表明试验药物在该标志物阳性的受试者
 243 中有疗效, 则考虑入组标志物阴性的受试者; 否则, 终止试验。

244

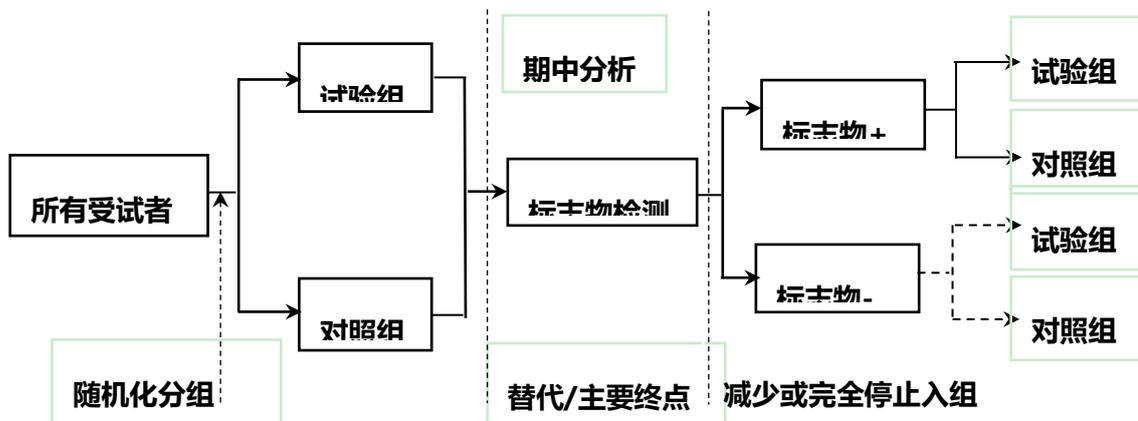
245

246

247

248

249



250

图 2 适应性富集设计流程图

251

252

253

254

255

256

257

一般来说, 如果标志物与疗效的关系越不确定, 越需要包含标志物阴性的受试者, 此时可以评估药物在全人群中使用时的获益与风险。当不确定一个标志物的预测性时, 主要终点可以是全人群中的疗效; 如果标志物阳性人群和全人群的疗效同时作为主要分析时, 需要按照一定的规则将检验水准 α 进行分配。无论何种情况, 都应事先在方案中明确规定检验假设, 并需要对 I 类错误进行控制。

258

四、富集策略与设计的相关考虑

259

(一) 标志物检测的灵敏度和特异度

260

当采用筛检试验选择受试者时, 必须考虑检测方法的可靠性,

261 以便能更准确地选择高风险或者对试验药物有应答的受试者。理
262 想情况下，用于筛选受试者的检测方法，应该对选择高风险或对
263 试验药物有应答的受试者有较高的灵敏度，同时对鉴别低风险或
264 对试验药物无应答的受试者有较高的特异度。

265 当利用生物标志物筛选入组受试者时，如果不能准确给出预
266 测标志物的阈值，可以通过受试者诊断特征（ROC）分析，即对
267 标志物不同阈值点的灵敏度和特异度进行分析，并用 ROC 曲线
268 下面积衡量其筛检效果。关于预测标志物阈值的确定，一般可以
269 先在早期研究阶段给出一个初步的阈值，然后通过较大样本的试
270 验对其进行调整，获得更可靠的阈值。

271 （二）是否纳入标志物阳性和阴性的受试者

272 富集设计既可以只纳入标志物阳性的受试者，也可以同时纳
273 入阳性和阴性的受试者。然而，富集设计的关键问题是纳入标志
274 物阴性受试者的比例。一般来说，可以考虑如下的富集设计：

275 1. 只纳入生物标志物阳性的受试者

276 如果作用机制或已有数据表明，试验药物在标志物阳性的受
277 试者中有明显的疗效，而在标志物阴性的受试者中疗效较小或者
278 完全没有疗效，尤其在试验药物有较大毒性的情况下，不应纳入
279 标志物阴性的受试者。

280 2. 同时纳入生物标志物阳性和阴性的受试者

281 如果作用机制或已有数据表明，标志物阳性的受试者较阴性
282 受试者的疗效更好，则在试验药物毒性相对较小的情况下，应同

283 时纳入标志物阳性和阴性的受试者。这种策略的优点是，能在非
284 富集人群中提供合理的获益-风险估计。

285 如果在试验开始之前就能够明确标志物，则可以通过对其分
286 组实施分层随机化，主要分析可限制在生物标志物阳性的受试者
287 中。实际中，也可以在全人群中进行主要分析，或者在全人群和
288 标志物阳性受试者中同时进行主要分析，并适当地控制 I 类错误。

289 一般来说，如果标志物的阈值或标志物阴性的受试者应答程
290 度不确定，则有必要纳入标志物阴性的受试者。

291 (三) 入选人群和分析集

292 使用富集策略的主要问题是研究结果的适用性和可外推性，
293 即采用富集设计时，要重点考虑这种富集策略是否能够在医学实
294 践中用于识别对研究药物应答的人群，以及该药在更广泛的患者
295 人群中是否也有类似的疗效。因此，对不符合富集入选标准的患
296 者人群进行研究同样重要。需要注意的是，试验确定的入选受试
297 者和主要分析集可以不同（后者可以是前者的子集），但这些必
298 须在研究方案中明确定义。在基因或其它检测结果不能立即获得，
299 而患者需要及时接受治疗时，选择以总体人群入组，以提供更多的
300 安全性信息，但主要疗效分析可以是其中的一个子集。

301 (四) 筛选富集人群对优效和非劣效试验的不同影响

302 使用标志物选择受试者时，对优效和非劣效试验有着不同的
303 影响。对于优效试验，当采用富集策略时，如果筛检方法的灵敏
304 度不高，则需要招募更多的受试者从中进行筛选，才能获得规定

305 的富集样本量；如果特异度不高，则需要增加富集样本量或延长
306 试验时间才能获得足够的终点事件数。尽管如此，但不会增加优
307 效试验的 I 类错误。

308 然而，对于非劣效试验，筛检的准确度不仅会影响研究所需
309 的样本量或持续时间，还可能增加 I 类错误。例如，采用预后型
310 富集策略进行非劣效试验，如果阳性对照的筛选方法与以往研究
311 不同，则可能导致阳性对照组的疗效低于以往研究的疗效，从而
312 增加 I 类错误。另外，对于基于预测型富集策略的非劣效试验，
313 对 I 类错误的影响更为复杂，它取决于标志物是与试验药物和阳
314 性对照药物的治疗相关，还是仅与其中一种治疗相关。因此，非
315 劣效试验中选择受试者的检测方法最好与阳性对照以往研究的
316 筛检方法一致，或者两种筛检方法有相似的灵敏度和特异度。

317 (五) 控制 I 类错误

318 对于同时入选富集人群和非富集人群的富集设计，可根据筛
319 检方法的准确度和受试者对治疗的应答情况，选择不同的假设检
320 验策略。如果有多个假设检验，如在标志物阳性人群和总体人群
321 中分别进行假设检验，则需要考虑多重性调整的问题；如果仅有一
322 个假设检验，如在标志物阳性人群中进行假设检验，则无需考
323 虑这一问题。在不同假设下 I 类错误 α 的分配，可以根据标志物
324 阳性人群对药物的应答程度、阳性人群在总体人群中的比例以及
325 按照预先设定的检验效能所需要的样本量进行设置。对总体人群
326 和富集人群进行假设检验时，可以采取独立或者序贯策略进行假

327 设检验。

328 五、监管考虑

329 (一) 明确界定富集人群

330 临床试验是否使用、何时使用以及使用何种富集策略，主要
331 取决于能否准确界定富集人群，这显然会对产品说明书的撰写和
332 后续的医疗实践产生影响。如果使用富集策略和设计无法准确界
333 定富集人群，则可能导致在说明书中无法准确地定义对治疗有效
334 的患者人群，进而无法准确指导临床合理用药。

335 (二) 不应忽视非富集人群的疗效

336 试验药物在富集人群中的有效性和安全性得到确证后，其在
337 非富集人群中相应的信息也应受到重视。通过非富集人群中开
338 展进一步研究，可以更全面地描述药物的获益-风险状况，为药
339 物能够在更广泛的患者人群中使用时提供依据。

340 对基于高风险人群预后型富集分析获批上市的药物，在随后
341 的低风险人群中试验，可能会使用不同的结局指标，如在高风险
342 人群中使用病死率，而在低风险人群中可以使用一个复合结局指
343 标，有助于提高试验效率。

344 (三) 预先确定研究方案并与监管机构沟通

345 一般来说，在研究开始之前，应预先计划并确定受试者的选
346 择方案。如果已知可用于富集的特征变量或标志物，可在筛选受
347 试者时对其进行测量。而当特征变量或标志物在研究人群中富集
348 的效果或分布不确定时，可以考虑进行适应性富集，即在试验过

349 程中依据积累的数据，对试验设计的要素进行调整。无论采用何
350 种策略和设计，调整方法和过程应在研究方案中事先说明，确保
351 其合理性和正确性，并与监管机构进行充分的沟通。

352 六、参考文献

- 353 [1] FDA. Enrichment Strategies for Clinical Trials to Support
354 Determination of Effectiveness of Human Drugs and Biological
355 Products: Guidance for Industry. The US Food and Drug
356 Administration: Silver Spring, 2019.
- 357 [2] EMA. Points to consider on multiplicity issues in clinical trials.
358 European Medicines Agency: EMA, 2002.
- 359 [3] EMA. Guideline on the investigation of subgroups in
360 confirmatory clinical trials. European Medicines Agency: EMA,
361 2019.
- 362 [4] ICH. ICH E5. Ethnic Factors in the Acceptability of foreign
363 Clinical Data E5 (R1). ICH Harmonised Tripartite: ICH, 1998.
- 364 [5] ICH. ICH E5. Implementation Working Group Questions &
365 Answers (R1). ICH Harmonised Tripartite: ICH, 2006.
- 366 [6] Wang S, Hung HMJ, O Neill RT. Genomic Classifier for Patient
367 Enrichment: Misclassification and Type I Error Issues in
368 Pharmacogenomics Noninferiority Trial. Stat Biopharm Res,
369 2011, 3: 310-319.
- 370 [7] Temple RJ. Special study designs: early escape, enrichment,

- 371 studies in non-responders. *Communications in Statistics -*
372 *Theory and Methods*, 1994, 23: 499-531.
- 373 [8] Taylor AL, Ziesche S, Yancy C, et al. Combination of isosorbide
374 dinitrate and hydralazine in blacks with heart failure. *N Engl J*
375 *Med*, 2004, 351: 2049-2057.
- 376 [9] Temple R, Stockbridge NL. BiDil for heart failure in black
377 patients: The U.S. Food and Drug Administration. *Ann Intern*
378 *Med*, 2007, 146: 57-62.
- 379 [10] Institute OM. *Small Clinical Trials: Issues and Challenges*.
380 Washington, DC. The National Academies Press; 2001.
- 381 [11] D'Agostino RB. The Delayed-Start Study Design. *New Engl J*
382 *Med*, 2009, 361: 1304-1306.
- 383 [12] Priscilla Velengtas, Penny Mohr, Messner DA. Making
384 informed decisions: Assessing the strengths and weaknesses of
385 study designs and analytic methods for comparative
386 effectiveness research. National Pharmaceutical Council ed.
387 Washington, DC; 2012.
- 388 [13] Freidlin B, Korn EL. Biomarker enrichment strategies:
389 matching trial design to biomarker credentials. *Nat Rev Clin*
390 *Oncol*, 2014, 11: 81-90.
- 391 [14] Wang SJ, Hung HM, O'Neill RT. Adaptive patient enrichment
392 designs in therapeutic trials. *Biom J*, 2009, 51: 358-374.

- 393 [15] Temple R. Enrichment of clinical study populations. *Clin*
394 *Pharmacol Ther*, 2010, 88: 774-778.
- 395 [16] Simon R. Clinical trial designs for evaluating the medical
396 utility of prognostic and predictive biomarkers in oncology. *Pers*
397 *Med*, 2010, 7: 33-47.
- 398 [17] Wang SJ, O'Neill RT, Hung HM. Approaches to evaluation of
399 treatment effect in randomized clinical trials with genomic
400 subset. *Pharm Stat*, 2007, 6: 227-244.
- 401 [18] Liu A, Liu C, Li Q, et al. A threshold sample-enrichment
402 approach in a clinical trial with heterogeneous subpopulations.
403 *Clinical trials (London, England)*, 2010, 7: 537-545.
- 404 [19] Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold
405 design: a procedure for evaluating treatment with possible
406 biomarker-defined subset effect. *J Natl Cancer Inst*, 2007, 99:
407 1036-1043.
- 408 [20] D'Amico AV, Chen MH, Roehl KA, et al. Preoperative PSA
409 velocity and the risk of death from prostate cancer after radical
410 prostatectomy. *N Engl J Med*, 2004, 351: 125-135.
- 411 [21] Fan C, Oh DS, Wessels L, et al. Concordance among
412 Gene-Expression–Based Predictors for Breast Cancer. *New Engl*
413 *J Med*, 2006, 355: 560-569.
- 414 [22] Scandinavian Simvastatin Survival Study Group. Randomised

415 trial of cholesterol lowering in 4444 patients with coronary heart
416 disease: The Scandinavian Simvastatin Survival Study (4S).
417 Lancet, 1994, 344: 1383-1389.

418 [23] Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to
419 prevent vascular events in men and women with elevated
420 C-reactive protein. N Engl J Med, 2008, 359: 2195-2207.

421 [24] EMA. Qualification Opinion of Alzheimer’s Disease Novel
422 Methodologies/biomarkers for BMS-708163. European
423 Medicines Agency: EMA, 2011.

424 [25] EMA. Qualification opinion of low hippocampal volume
425 (atrophy) by MRI for use in regulatory clinical trials - in
426 pre-dementia stage of Alzheimer’s disease. European Medicines
427 Agency: EMA, 2011.

428 [26] Amur S, LaVange L, Zineh I, et al. Biomarker Qualification:
429 Toward a Multiple Stakeholder Framework for Biomarker
430 Development, Regulatory Acceptance, and Utilization. Clin
431 Pharmacol Ther, 2015, 98: 34-46.

432 [27] FDA. Qualification of Biomarker–Plasma Fibrinogen in Studies
433 Examining Exacerbations and/or All-Cause Mortality in Patients
434 with Chronic Obstructive Pulmonary Disease. US Food and
435 Drug Administration: Silver Spring, 2016.

436 [28] FDA. Qualification of Biomarker–Total Kidney Volume in

- 437 Studies for Treatment of Autosomal Dominant Polycystic
438 Kidney Disease. US Food and Drug Administration: Silver
439 Spring, 2016.
- 440 [29] Early Breast Cancer Trialists' Collaborative Group (EBCTCG).
441 Effects of chemotherapy and hormonal therapy for early breast
442 cancer on recurrence and 15-year survival: an overview of the
443 randomised trials. *Lancet*, 2005, 365: 1687-1717.
- 444 [30] Loo E, Khalili P, Beuhler K, et al. BRAF V600E Mutation
445 Across Multiple Tumor Types: Correlation Between DNA-based
446 Sequencing and Mutation-specific Immunohistochemistry. *Appl*
447 *Immunohistochem Mol Morphol*, 2018, 26: 709-713.
- 448 [31] Chapman PB, Hauschild A, Robert C, et al. Improved survival
449 with vemurafenib in melanoma with BRAF V600E mutation. *N*
450 *Engl J Med*, 2011, 364: 2507-2516.
- 451 [32] Kowanetz M, Zou W, Gettinger SN, et al. Differential
452 regulation of PD-L1 expression by immune and tumor cells in
453 NSCLC and the response to treatment with atezolizumab
454 (anti-PD-L1). *Proc Natl Acad Sci USA*, 2018, 115:
455 E10119-E10126.
- 456 [33] Havel JJ, Chowell D, Chan TA. The evolving landscape of
457 biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev*
458 *Cancer*, 2019, 19: 133-150.

- 459 [34] Schrock AB, Ouyang C, Sandhu J, et al. Tumor mutational
460 burden is predictive of response to immune checkpoint inhibitors
461 in MSI-high metastatic colorectal cancer. *Ann Oncol* 2019, 30:
462 1096-1103.
- 463 [35] Hughes RA, Donofrio P, Bril V, et al. Intravenous immune
464 globulin (10% caprylate-chromatography purified) for the
465 treatment of chronic inflammatory demyelinating
466 polyradiculoneuropathy (ICE study): a randomised
467 placebo-controlled trial. *Lancet Neurol*, 2008, 7: 136-144.
- 468 [36] Parkinson Study Group. A controlled, randomized,
469 delayed-start study of rasagiline in early Parkinson disease. *Arch*
470 *Neurol*, 2004, 61: 561-566.
- 471 [37] Singh BN. Comparative efficacy and safety of bepridil and
472 diltiazem in chronic stable angina pectoris refractory to diltiazem.
473 The Bepridil Collaborative Study Group. *Am J Cardiol*, 1991, 68:
474 306-312.
- 475 [38] Barker AD, Sigman CC, Kelloff GJ, et al. I-SPY 2: an adaptive
476 breast cancer trial design in the setting of neoadjuvant
477 chemotherapy. *Clin Pharmacol Ther*, 2009, 86: 97-100.

478

附录 1：词汇表

479

480

481

482

灵敏度 (Sensitivity)：评价诊断试验和筛检试验准确度的基本指标之一。在药物临床试验的富集研究中，灵敏度表示对于发生终点事件具有高风险或对药物有应答的受试者，能够将其正确识别出的概率。

483

484

485

486

487

适应性富集设计 (Adaptive Enrichment Design)：按照预先制定的计划，根据临床试验数据的期中分析结果，在保证试验合理性和完整性的前提下，允许在试验过程中适应性地更新入排标准，选择最有可能从治疗中获益的受试者入组的自适应设计。

488

489

490

491

随机撤药设计 (Randomized Withdrawal Design)：在此类设计中，所有受试者都在最初的开放标签期间接受试验药物的治疗，之后对药物无应答的受试者退出试验，有应答的受试者（富集亚群）在试验第二阶段中随机接受试验药物或安慰剂。

492

493

494

495

特异度 (Specificity)：评价诊断试验和筛检试验准确度的基本指标之一。在药物临床试验的富集研究中，特异度表示对于发生终点事件具有低风险或对药物无应答的受试者，能够将其正确识别出的概率。

496

497

498

499

异质性 (Heterogeneity)：在临床试验中，异质性体现在个体和群体两个水平，前者通常是指受试者间具有不同的特征，个体性质或状态的不同可能会导致不同的受试者对治疗有不同的应答；后者通常是指不同中心、种族、地域等受试者具有不

500 同的特征，有可能导致不同受试者对治疗有不同的应答。

501 **预测型富集 (Predictive Enrichment)**：是指选择性地纳入可
502 能会对治疗有应答的受试者的一种研究策略或设计，这些受试
503 者具有共同的有预测意义的生物学和组织病理学特征，可以更
504 敏感地显示试验药物的。

505 **预后型富集 (Prognostic Enrichment)**：是指选择性地纳入更可
506 能发生终点事件（如死亡或疾病恶化）的受试者，从而降低为达
507 到统计学显著疗效所需样本量的一种研究策略或设计。

中文	英文
靶向治疗	Target Therapy
病理生理学	Pathophysiology
蛋白质组学	Proteomics
单臂试验	Single-arm Trial
低风险人群	Low-risk Population
多重性	Multiplicity
非富集人群	Non-enriched Population
非劣效性试验	Non-inferiority Trial
分析集	Analysis Set
复合结局指标	Composite Endpoint
富集策略	Enrichment Strategies
富集人群	Enriched Population
高风险人群	High-risk Population
复合型富集策略	Mixed Enrichment Strategies
获益-风险比	Benefit-risk Ratio
基因组学	Genomics
基因突变	Gene Mutation
假设检验	Hypothesis Test
可推广性	Generalizability
灵敏度	Sensitivity

中文	英文
目标人群	Target Population
偏倚	Bias
确证性临床试验	Confirmatory Clinical Trial
筛检试验	Screening Test
生物标志物	Biomarkers
生物标志物阳性	Biomarkers Positive
生物标志物阴性	Biomarkers Negative
适应性富集策略	Adaptive Enrichment Strategies
受试者诊断特征	Receiver Operating Characteristic (ROC)
随机撤药	Randomized Withdrawal
特异度	Specificity
替代指标	Surrogate Marker
同质化富集策略	Reducing Heterogeneity Strategies
外部对照	External Control
亚组分析	Subgroup Analysis
亚组人群	Subpopulation
I 类错误	Type I Error
依从性	Compliance
优效性试验	Superiority Trial
预测型富集策略	Predictive Enrichment Strategies
预后型富集策略	Prognostic Enrichment Strategies

中文	英文
终点事件	Endpoint
肿瘤代谢物	Tumor Metabolite
主要疗效分析	Primary Efficacy Analysis

509

附录 3：富集设计的研究案例

示例 1：预后型富集——心血管病研究

在心血管病研究中，选择高风险人群（如 AMI、中风、极高的胆固醇水平、非常严重的 CHF 及接受血管成形术等）进行临床研究，能够获得更多的结局事件。斯堪的纳维亚辛伐他汀生存研究（4S）是一项降脂药物的试验，主要目的是评估辛伐他汀能否通过降低血清胆固醇改善冠心病患者的生存率。该研究为随机双盲安慰剂对照的多中心临床试验，招募了 4444 名患有心绞痛或先前有心肌梗死（MI）的患者，这些受试者都具有较高的总胆固醇（TC）水平。在平均 5.4 年的随访期内，辛伐他汀相对于安慰剂的心血管病死亡率能够得到显著降低（相对危险度 RR 为 0.70，95% CI: 0.58-0.85）。

示例 2：预测型富集——黑色素瘤研究

BRAF 激酶抑制剂是一种治疗黑色素瘤的靶向药物，BRAF 基因第 15 外显子（V600E）可以作为预测型生物标志物。已知 BRAF 基因编码细胞质丝氨酸/苏氨酸激酶，该酶调节丝裂原活化蛋白激酶信号转导通路，控制包括细胞生长和分裂（增殖）在内的几个重要细胞功能。现已发现，BRAF V600E 在多种肿瘤中发生突变，如黑色素瘤、大肠癌、甲状腺乳头状癌、毛细胞白血病（Hairy Cell Leukemia）和朗格汉斯细胞增生症等。一项黑色素瘤 III 期临床试验研究，入选了 675 例转移性或不可切除、BRAFV600E 突变的受试者，分别给予 BRAF 激酶抑制剂维罗非

532 尼或化疗药物达卡巴嗪，结果发现：给予维罗非尼靶向药治疗的
533 受试者其应答率为 48%，而使用达卡巴嗪化疗的受试者其应答率
534 仅为 5%；两者相比较，经维罗非尼治疗的受试者相对死亡风险
535 降低了 63%。

536 示例 3：随机撤药设计——普瑞巴林用于治疗纤维性肌痛症 537 的研究

538 一项研究普瑞巴林（Pregabalin）对治疗纤维性肌痛症受试
539 者的疗效的临床试验使用了两阶段随机撤药设计，其目的是比较
540 普瑞巴林相对于安慰剂在接受治疗到疗效丧失的时间上的差异
541 （Time to Loss of Therapeutic Response, TLTR）。第一阶段是开
542 放性试验，患有纤维性肌痛症受试者全部接受普瑞巴林治疗，并
543 观察 6 周。其中，在 1-3 周，受试者接受递增剂量的普瑞巴林以
544 决定其最佳剂量；在 4-6 周，受试者维持在此最佳剂量。第一阶
545 段开放性治疗结束后，受试者必须具有至少 50% 以上的疼痛减轻、
546 且在 PGIC 量表上的自我评价至少是“显著改善”才能进入第二
547 阶段的双盲、安慰剂对照的试验。在入组的 1051 名受试者中，
548 经第一阶段的治疗，符合以上条件进入第二阶段的有 566 人，其
549 中 287 人被随机分配到安慰剂组、279 人被分配到普瑞巴林组。
550 经过第二阶段 26 周的治疗，两组间在治疗应答丧失时间（Time
551 to Loss of Therapeutic Response, LTR）上有显著差异（ $p < 0.0001$ ）。
552 试验结束时，安慰剂组有 61%（178）、普瑞巴林组有 32%（90）
553 达到了 LTR。

554 示例 4：预测型富集——MSI 研究

555 微卫星不稳定性 (MSI) 是对免疫检查点抑制剂产生应答的
556 生物标志物。PD-1/PD-L1 通路是调节 T 细胞活化的信号通路，
557 在肿瘤发生和进展中起重要作用。实际中，通常用免疫组化方法
558 检测 PD-L1 蛋白的表达水平，将其作为预测标志物并选择高表
559 达的受试者，但其对 PD-1/PD-L1 抑制剂的应答率仅为 10-20%。
560 然而，在高度微卫星不稳定型 (MSI-HIGH) 的肿瘤受试者中的
561 应答率却能达到 50%。基于此，FDA 批准帕博利珠单抗适用于
562 MSI-HIGH 型或存在错配修复缺陷的结直肠癌和子宫内膜癌受
563 试者的治疗。