

药物临床试验亚组分析
指导原则
(征求意见稿)

2020 年 8 月

目 录

一、概述	1
二、亚组的识别和定义	2
(一) 亚组的识别	2
(二) 亚组的定义	3
三、亚组分析的类型	4
(一) 探索性亚组分析	4
(二) 支持性亚组分析	5
(三) 确证性亚组分析	5
四、一般考虑	6
(一) 生物学上的合理性及临床意义	6
(二) 异质性	8
(三) 可信度	9
(四) 主要终点指标与亚组分析	10
五、确证性亚组分析	11
(一) 亚组的选择	12
(二) 试验设计	12
(三) 多重性	14
(四) 结果的解释	16
六、支持性亚组分析	16
七、其他考虑	19
(一) 非劣效试验中的亚组分析	19
(二) 安全性亚组分析	19
(三) 与监管机构的沟通	20
八、参考文献	21
附录1. 词汇表	27
附录2. 中英文对照表	31
附录3. 亚组分析与设计研究案例	34
示例1:帕博利珠单抗免疫治疗小细胞肺癌的亚组分析	34
示例2:那他珠单抗治疗复发-缓解型多发性硬化症的获益-风险评估	35

1 药物临床试验亚组分析指导原则（征求意见稿）

3 一、概述

4 参与临床试验的患者由于受各种因素（如遗传学、人口学、
5 环境、病理生理学、合并症和合并用药等）的影响，往往具有不
6 同程度的异质性，从而可能导致试验药物在不同患者中的疗效不
7 同。临床试验中将具有不同特征的患者分组，是探索不同患者人
8 群之间疗效差异的直观方法，同时也是获益-风险评估不可缺少
9 的一部分。

10 对于事先未设定亚组分析的临床试验，在对结果进行解读和
11 下结论时需要特别慎重。事后根据数据驱动寻找有统计学意义的
12 亚组，会导致 I 类错误膨胀，其结果通常不能用于确证该亚组的
13 有效性。此外，因亚组的样本量较少而导致检验效能不足，可能
14 会影响试验药物在这些亚组人群中疗效的精确估计，或者无法得
15 出各亚组间疗效是否一致的结论。

16 本指导原则中将目标适应症人群称为总体人群，将通过入排

17 标准纳入临床试验的人群称为全人群。亚组人群（简称亚群）指
18 总体人群中具有某些特征的一个子集，亚组是全人群中的一个子
19 集。亚组分析是指对试验药物在亚组中的疗效和/或安全性进行试
20 验设计与统计分析的过程，亚组是亚群的一个样本，将亚组结果
21 推广到亚群时需要考虑亚组对亚群的代表性问题。

22 本指导原则主要阐述了亚组的识别和定义、亚组分析的类型、
23 一般考虑以及确证性临床试验中的亚组分析等方面的内容，旨在
24 为申办者能够在临床试验中对亚组分析进行正确地设计、实施和
25 评价提供指导性建议。本指导原则主要适用于以支持药品注册上
26 市为目的的确证性临床试验，也可供以非注册为目的的临床试验
27 参考。

28 二、亚组的识别和定义

29 （一）亚组的识别

30 亚组的识别一般基于早期临床试验或确证性临床试验的探
31 索性分析，可以使用定量的方法，如交互树和递归分割树等方法；
32 也可以根据文献报道或者医疗实践积累的知识进行识别。

33 亚组识别主要关注不同亚组间疗效的差异及其临床意义，本
34 质上属于探索性研究，一般无须控制 I 类错误。进行亚组识别时
35 应考虑的因素包括但不限于以下几个方面：①临床上的可解释性；
36 ②临床上的可操作性；③药物的作用机理；④定义亚组的变量个
37 数和类型（如连续变量、分类变量）；⑤误分或漏分亚组两种情
38 况带来的风险；⑥亚组识别模型的选择（不宜过于复杂，以避免
39 过拟合）。

40 （二）亚组的定义

41 亚组通常由患者的一个或多个内在和/或外在因素（见 ICH
42 E5）来定义，而且应具有一定的临床意义。这些变量通常是基线
43 变量，包括人口学特征（如年龄、性别等）、实验室检查指标、
44 基因组相关标志物、疾病的严重程度或分型、临床状况（如合并
45 症、伴随用药）、地区（如国家、试验中心）和环境因素等。

46 实际应用中，亚组大多由一到两个变量来定义，例如把研究
47 对象按照年龄分为 18-65 岁和 > 65 岁两个年龄组；在抗肿瘤药物
48 研究中，把患者按照 ECOG 分数和基因突变(或基因表达)水平，

49 分成不同的亚组。需要注意的是，使用多于两个变量定义亚组比
50 较少见，主要是因为：①定义亚组的变量过多，容易出现误分和
51 漏分等情况；②使用多个变量定义亚组时，对亚组分析结果的解
52 释往往出现困难。当亚组的定义确实需要涉及多个变量时，可以
53 考虑将这些变量通过一定的方法转换成单一的风险分数，并以此
54 来定义亚组。例如抗肿瘤药物试验中常用的 ECOG 评分反应了患
55 者自我保健、日常活动和体能状况等综合能力。

56 三、亚组分析的类型

57 根据研究目的，亚组分析分为探索性亚组分析、支持性亚组
58 分析和确证性亚组分析。对于探索性亚组分析，亚组既可以在设
59 计阶段事先定义、也可以在分析阶段事后定义（如根据数据驱动
60 划分的亚组分析）。对于支持性亚组分析，亚组一般应在临床试
61 验的设计阶段事先定义，并在试验方案中详细描述。对于确证性
62 亚组分析，亚组必须在临床试验的设计阶段事先定义，并在试验
63 方案中详细描述。

64 （一）探索性亚组分析

65 探索性亚组分析主要用于早期临床试验或在确证性临床试
66 验的探索性分析中，其目的是发现药物在不同亚组间疗效或安全
67 性方面的差异，进而提出研究假设，以待在后续的临床试验中进
68 一步探索和验证。因此，探索性亚组分析主要关注的是其结果在
69 生物学上的合理性或临床上的可解释性，一般无须进行多重性调
70 整。

71 (二) 支持性亚组分析

72 在确证性临床试验中，除了对全人群的主要终点进行分析
73 外，还需要对各亚组的疗效进行分析，目的是进一步考察试验
74 药物在各个亚组中疗效的一致性。如果试验药物在各亚组间的
75 疗效差异在可接受范围，可认为试验药物在亚组间的疗效一致，
76 并为药物适用于全人群提供进一步支持性证据。如果各亚组间
77 的疗效差异超出了可接受的范围，特别是方向相反时，则亚组
78 分析结果的解释可能会出现困难，需要对其做进一步的分析和
79 研究。

80 (三) 确证性亚组分析

81 确证性临床试验中，按照预先规定的亚组和多重性调整方
82 法，考察试验药物在目标亚组和/或全人群中的疗效，其结果应
83 同时具有临床意义和统计学意义，以支持药物说明书的撰写。
84 确证性临床试验也可以对目标亚组进行确证性亚组分析，而对
85 其它（非目标）亚组进行支持性或探索性亚组分析，以支持试
86 验药物在各亚组人群或全人群中的有效性和安全性的结论，或
87 者为进一步的亚组研究提供线索。

88 四、一般考虑

89 （一）生物学上的合理性及临床意义

90 1. 病理生理学和遗传学的差异

91 不同患者之间潜在的病理生理学或遗传学的差异可能导致
92 药物治疗效果的不同。一些临床综合征可能有多个潜在的病理学
93 机制，如原发性全身性癫痫有不同的分子病理学基础，因此使用
94 药物治疗该病的效果往往因不同的发病机制而异。另外，对于一
95 些基础疾病，由于其可预见的生物学反应的差异，也可能促进亚
96 组分析的使用。例如在减少肾移植排斥反应方面，对于术前免疫

97 致敏的患者在手术期间使用抗淋巴细胞抗体，可以使其获益更多。
98 此外，很多药物的疗效及不良反应与患者的基因型密切相关，例
99 如对于绝经后妇女补充雌激素，可以增加血浆中高密度脂蛋白胆
100 固醇含量，其效果与雌激素受体基因序列的变异密切相关。

101 2. 最佳治疗时机和合并症

102 亚组分析所关注的主要问题之一是在医疗实践中如何使药
103 物的疗效在不同的亚组或疾病的不同阶段达到最优化，例如确定
104 疾病治疗的最佳阶段、在某一临床事件发生后的最优治疗时间以
105 及治疗时机与合并症之间的关系等。药物的疗效通常取决于疾病
106 的严重程度或进展阶段，在医疗实践中经常据此对患者进行分组，
107 以确定药物的适应症和最佳时机。例如，急性心肌梗死的溶栓治
108 疗对 ST 段抬高的患者非常有效，而对 ST 段压低的患者无效甚
109 至有害；同时，溶栓治疗的合适时机是发病 6 小时以内，且 3 小
110 时以内效果更好。另外，治疗效果也可能与合并症有关，例如血
111 管紧张素转换酶抑制剂对合并非肾性高血压的心衰患者非常有

112 益，但对合并肾血管疾病的患者却有害。

113 3. 有效性和安全性亚组分析的差异

114 亚组通常是根据试验药物对亚组人群的有效性来定义的，这
115 样的分组对安全性分析可能不合适，因为：①药物的有效性和安
116 全性一般使用不同的临床终点或者替代终点，其机理可能不同；
117 ②定义有效性亚组的变量（例如靶点受体）和安全性亚组的变量
118 （例如年龄等因素）往往不同；③对可能导致潜在危害的药物，
119 不同亚组患者的耐受程度可能不同（例如病情严重的患者为了治
120 疗获益可能更容易耐受一定程度的副反应），此时亚组的获益-风
121 险评估非常重要。

122 （二）异质性

123 亚组分析的主要目的是为了更好地了解试验药物在各亚组
124 和全人群中的疗效，而是否需要和如何设计亚组分析关键在于临
125 床试验中目标人群的异质性，这与研究方案中制定的患者入排标
126 准有关。通常，入排标准的限制条件越严格，招募的患者异质性
127 就越小，亚组间疗效的差异也就相对较小；反之，宽松的入排标

128 准可能导致入组患者的异质性增加，亚组间的疗效差异也会增加，
129 不易显示试验药物在全人群中的疗效，此时进行亚组分析就显得
130 非常必要。

131 (三) 可信度

132 亚组分析结果的可信度评估包括但不限于以下几个方面：①
133 定义亚组的变量具有生物学上的合理性，即对患者的预后或治疗
134 应答的预测有科学依据；②亚组分析结果的可重现性，即在相同
135 或相似条件下的其他临床研究中，具有相同或相似的亚组效应；
136 ③先验知识的可靠性，即源于外部数据和文献报告的亚组分析结
137 果具有一定的科学依据。

138 在研究设计阶段，有时基于先验知识指定用于分层随机化和
139 亚组分析的变量。这种预先指定亚组变量的方法，通常用在确证
140 性和支持性亚组分析中，蕴含了亚组之间疗效有差异的推测，因
141 此得到的亚组分析结果具有一定的可信度。然而，即使亚组变量
142 不是被预先指定，也要予以重视，尤其是对在安全性亚组分析中
143 发现的亚组之间的差异，要特别关注其生物学上的合理性和结果

144 的可重现性。

145 当亚组的样本量不足、无法准确估计药物在亚组中的疗效时，
146 应主要考察其生物学上的合理性和结果的可重现性。当亚组疗效
147 在试验条件（如研究设计、目标人群、亚组定义、治疗方案、结
148 局测量等）相似的一系列临床试验中一致时，即使没有明确的临
149 床和生物学方面的解释，亚组结果也具有一定的可信度。另外，
150 虽然随机化可以使不同治疗组间入组患者的基线变量分布趋于
151 平衡，但由于亚组内样本量的减少，可能会存在基线不均衡，因
152 此必须检查药物在各亚组间疗效差异是否是由于基线分布不均
153 衡所致。

154 （四）主要终点指标与亚组分析

155 当临床试验的主要终点为复合终点时，有些情况下药物的疗
156 效可能只反映在构成复合终点的某一或某些组成终点上，而其余
157 的组成终点则受到试验药物以外其他因素的影响，此时，不宜用
158 复合终点进行亚组分析。例如，在一项治疗中轻度高血压的试验
159 中，以全因死亡率作为主要终点，但对亚组进行分析时发现，试

160 验药物降低了男性的全因死亡率，但却增加了女性的全因死亡率，
161 原因是女性非心血管因素导致的死亡过多，此时对性别的亚组分
162 析采用全因死亡率这一复核终点可能并不合适，而使用心血管因
163 素的死亡专率较为恰当。

164 当使用多个主要终点指标时，整体治疗效果通常是对其综合
165 分析的结果，如果这些指标具有不同的风险因素，需要考虑这些
166 因素在各亚组的分布，并阐述如何进行亚组分析及其结果的临床
167 意义。

168 五、确证性亚组分析

169 确证性亚组分析是在确证性临床试验中对事先指定的目标
170 亚组进行假设检验的分析，目标亚组可以作为主要或共同主要分
171 析人群。亚组分析的结果提示不同亚组人群可能的获益与风险，
172 因此直接影响决策和产品使用说明书的撰写。确证性亚组分析的
173 临床试验需要考虑的关键问题包括但不限于以下几个方面：①亚
174 组的选择；②试验设计类型（如固定样本设计、适应性设计或者
175 适应性富集设计）；③多重性；④亚组分析结果的解释。需要注

176 意的是，把亚组作为主要或共同主要分析人群的试验设计应考虑
177 亚组的样本量以及是否使用分层随机等关键问题。

178 (一) 亚组的选择

179 确证性亚组分析应在方案中预先规定目标亚组。关于亚组的
180 选择，如果是基于医疗知识或实践，例如按照疾病严重程度、人
181 口学特征（性别、年龄等）或已知的能够精确分辨亚组的生物标
182 志物（如基因突变）进行分类，通常具有一定的临床意义。另一
183 方面，亚组的定义是否合理取决于亚组分类器（如标志物）是否
184 能够可靠地识别最有可能从药物中获益的亚组人群。分类器的确
185 定通常基于早期临床研究数据，由于样本量往往不足，因此分类
186 器的性能有限，在研究设计时要考虑到这一问题。

187 (二) 试验设计

188 在进行确证性亚组分析的样本量估计时，除了常用的试验设
189 计参数如期望治疗效应大小及其变异度、I 类和 II 类错误概率之
190 外，还必须考虑亚组分类器分辨亚组的准确度，以及亚组人群在
191 符合入组条件的目标人群中所占的比例。

192 如果有可靠的证据表明试验药物在不同亚组中的疗效不同，
193 且其具有生物学上的合理性和重要的临床意义，可以采用固定样
194 本量设计，以验证药物的疗效。此时，临床试验的目标人群可以
195 是亚组人群和/或总体人群，其至少有三种设计方案（以标记物举
196 例）：

197 ①如果只有标志物阳性的患者才能从试验药物中获益，入组
198 患者可以仅限于这个亚组。

199 ②如果标志物阳性和阴性的患者都能从试验药物中获益，但
200 阳性患者获益高于阴性患者，而试验的主要目的是验证药物在阳
201 性患者中的疗效，则样本量和药物疗效的估计可以主要针对标志
202 物阳性的亚组，但建议标志物阴性的亚组也纳入试验，以便更好
203 地了解试验药物在该人群中的疗效，用于获益-风险评估或后续
204 研究设计。

205 ③如果标志物阳性和阴性的患者都能从试验药物中获益，但
206 不能确定哪个亚组人群的获益更大，而试验的主要目的是验证药

207 物在总体人群中的疗效，则患者可以在总体人群中招募，也可以
208 在两个亚组人群中进行分层随机。如果是后者，则需要注意各亚
209 组样本量的比例应与目标人群中各亚组人群患者的比例相似，以
210 避免由于过多纳入疗效较好的亚组人群而夸大了药物在总体人
211 群中的疗效。

212 如果没有充分的证据表明不同亚群间的疗效差异具有临床
213 意义时，可以采用适应性设计的方法进行亚组的选择。例如，当
214 不清楚试验药物在标志物阳性和阴性患者中的疗效时，可以考虑
215 两阶段适应性设计（即适应性亚组选择和富集）。第一阶段试验
216 的数据可用于估计药物在亚组中的疗效，然后据此调整第二阶
217 段的入组人群及其样本量。对于适应性设计中的亚组的选择，应重
218 点考虑试验设计和统计分析方法的有效性（如是否控制了 I 类错
219 误）。

220 (三) 多重性

221 多重性确证性亚组分析中需要重点关注的问题之一。对于
222 将亚组作为共同主要分析的临床试验，由于要对全人群和亚组的

223 疗效进行多次检验，因此如果不进行多重性调整，会增加总 I 类
224 错误。为了将总 I 类错误控制在预设的水平，有多种多重性调整
225 方法，如 Bonferroni 单步检验法、逐步多重检验法和回退法等，
226 多重性调整的方法应在方案和统计分析计划中事先指定。

227 不同的多重性调整方法各有其优缺点。如果一个亚组和全人
228 群作为共同主要分析人群，则可以用 Bonferroni 法将 α 分配给这
229 两个主要分析人群，即以 α_1 水平检验亚组，以 α_2 水平检验全人群，
230 并且 $\alpha_1 + \alpha_2 = \alpha$ 。这里也可以使用回退法。

231 另外，在没有足够的证据表明药物在全人群和亚组的疗效差
232 异时，可以考虑使用逐步多重检验法（如 Hochberg、Hommel 和
233 Simes 法），即对多个假设按照一定的顺序进行检验。如果有足够
234 的证据表明亚组的疗效优于全人群的疗效时，可以考虑使用固定
235 顺序检验策略，即先检验亚组疗效，如果有统计学意义，再检验
236 全人群的疗效，否则停止检验。

237 值得注意的是，由于亚组属于全人群的一部分，使用统计量

238 的联合分布来确定检验界值，可以提高检验效能，但由于数据有
239 限，估计统计量的相关系数往往不可靠，并可能增加 I 类错误。
240 因此，使用基于统计量的联合分布来确定检验界值的多重性调整
241 方法需要特别谨慎。

242 (四) 结果的解释

243 亚组分析结果的解释必须具有临床意义，以支持药物的上市
244 和说明书的撰写。例如，在一项临床试验中预先以某一目标亚组
245 和全人群共同作为主要的分析人群，经过多重性调整后的统计分
246 析可以得出下述三个结论之一：①与对照组相比，试验药物在全
247 人群中的疗效差异有统计学意义而目标亚组的疗效差异无统计
248 学意义；②目标亚组的疗效有统计学意义而全人群的疗效无统计
249 学意义；③全人群和目标亚组的疗效都有统计学意义。需要注意
250 的是，如果结论③是由于药物在目标亚组中的疗效较大所致，而
251 其余亚组很少甚至不获益，此时，将其使用限定于该目标亚组可
252 能更加合适。

253 六、支持性亚组分析

254 一般情况下，确证性临床试验的目的是验证药物在全人群中的
255 的有效性以及各主要亚组之间疗效的一致性，后者通过支持性亚
256 组分析来实现。常用的支持性亚组分析方法包括：①描述性分析
257 （如疗效的点估计）和区间估计；②图形显示（如森林图）；③
258 模型法（如在主要分析模型中加入药物与定义目标亚组变量之间的
259 交互作用）。使用何种方法应在方案及统计分析计划书中充分
260 考虑和说明。

261 需要指出的是，当各亚组的样本量较少或亚组间样本量分布
262 不均衡时，药物与亚组变量交互作用的检验效能往往不足。为此，
263 可以考虑选取宽松的检验水准（例如 $\alpha=0.10$ ）进行检验，同时要
264 考虑各亚组间疗效差异的临床意义。如果交互作用显著，则提示
265 试验药物在各亚组间的疗效可能不同。

266 在确证性临床试验中，如果试验药物的疗效在全人群中有统
267 计学意义，通常需要报告亚组分析的结果，包括各亚组疗效的点
268 估计及其置信区间。需要注意的是，如果对全人群疗效和各亚组

269 的疗效同时进行分析，则各亚组疗效与全人群疗效的方向应该一
270 致，否则需要进一步分析其不一致的原因。

271 如果各亚组间的疗效差异具有临床意义，应考虑按照以下几
272 个步骤探索其可能的原因：

273 ①定义亚组的变量是否与相应的预后或预测因素有关。这些
274 因素通常从早期临床试验、文献报告或医疗实践中获知，可以是
275 内在因素（如遇药物代谢酶相关的基因多态性等遗传因素）、外
276 在因素（如疾病严重程度、吸烟状况和 BMI 等）或治疗方法（如
277 不同剂量的合并用药）等。

278 ②如果定义亚组的变量与相应的预后或预测因素有关，则应
279 进行进一步分析。如首先分析预后因素是否具有预测作用，如果
280 有预测作用说明该因素会影响患者对药物的应答或疗效；然后，
281 分析预后因素在各亚组的试验组和对照组中的分布是否均衡，如
282 果不均衡则其可能是亚组间疗效差异的原因。

283 ③如果通过上述分析，亚组间疗效的差异仍无法解释，则需

284 考虑进一步的探索性分析以识别造成此差异的其他可能因素，对
285 试验结果给出合理的解释。

286 七、其他考虑

287 （一）非劣效试验中的亚组分析

288 亚组分析在只有阳性对照的双臂非劣效临床试验中要十分
289 谨慎。当在各亚组间对试验药疗效进行非劣效评价时，其结果依
290 赖于阳性对照药相对于安慰剂的疗效在各亚组中的一致性。若阳
291 性对照药相对于安慰剂的效应与亚组变量有关时，则非劣效界值
292 的选择应考虑这些特征变量在研究人群中的分布。由于非劣效界
293 值的确定通常使用历史数据，因此，如果亚组变量在试验人群中
294 的分布与历史数据不同，则会影响非劣效检验结论的正确性。要
295 强调的是，非劣效界值的确定应当基于阳性对照药疗效（相对于
296 安慰剂）的尽可能多的历史数据的 meta 分析结果，而不是基于
297 主观选择部分试验数据或者部分亚组人群数据的分析结果。

298 （二）安全性亚组分析

299 安全性亚组分析主要用于研究与药物安全性相关的风险因

300 素，即探究具有一定特征的亚组人群可能对药物产生的不良反应。

301 获益-风险评估是亚组安全性分析不可或缺的一部分。安全
302 性亚组分析与有效性亚组分析略有不同。安全性亚组分析虽然可
303 以由预后因素（如年龄、是否有并发症等）或预测因素（如疾病
304 分期、基因分型等）定义亚组，但前者更为常见。

305 亚组人群的安全性分析具有一定的挑战，特别是对低发生率
306 或者潜伏期较长的安全性事件，由于样本量较小或随访时间较短，
307 较难在试验期间发现和验证。如果现有数据表明试验药物在特定
308 亚组人群中与某一严重不良事件（SAE）有潜在的关联，则可以
309 考虑有针对性地设计一个大型的具有足够检验效能的临床试验，
310 验证试验药物在该亚组人群中是否会引起严重的不良反应。

311 （三）与监管机构的沟通

312 当临床试验设计包含确证性亚组分析时，鼓励申请人就临床
313 试验设计中的关键问题与监管机构进行沟通。沟通的内容包括但
314 不限于试验设计的类型、亚组的选择、I 类错误控制、一致性和
315 可信性等方面的评价策略。

316 八、参考文献

- 317 [1] ICH. ICH E5: Ethnic Factors in the Acceptability of foreign
318 Clinical Data E5 (R1). ICH Harmonised Tripartite
319 ([https://database.ich.org/sites/default/files/E5_R1 Guideline.pdf](https://database.ich.org/sites/default/files/E5_R1_Guideline.pdf)),
320 1998.
- 321 [2] EMA. Guideline on the investigation of subgroups in
322 confirmatory clinical trials. European Medicines Agency (EMA)
323 ([https://www.ema.europa.eu/en/documents/scientific-guideline/g](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-subgroups-confirmatory-clinical-trials_en.pdf)
324 [uideline-investigation-subgroups-confirmatory-clinical-trials_en.](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-subgroups-confirmatory-clinical-trials_en.pdf)
325 [pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-subgroups-confirmatory-clinical-trials_en.pdf)), 2019.
- 326 [3] Julien Tanniou, Ingeborg Van Der Tweel, Steven Teerenstra, and
327 Kit CB Roes. Subgroup analyses in confirmatory clinical trials:
328 time to be specific about their purposes. *BMC Medical Research*
329 *Methodology*, 16(1):20, 2016.
- 330 [4] Mohamed Alish, Mohammad F Huque, Frank Bretz, and Ralph
331 B D'Agostino Sr. Tutorial on statistical considerations on
332 subgroup analysis in confirmatory clinical trials. *Statistics in*
333 *medicine*, 36(8):1334-1360, 2017.
- 334 [5] Ilya Lipkovich, Alex Dmitrienko, and Ralph D'Agostino Sr.
335 Tutorial in biostatistics: data-driven subgroup identification and
336 analysis in clinical trials. *Statistics in medicine*, 36(1):136-196,
337 2017.
- 338 [6] James F Burke, Jeremy B Sussman, David M Kent, and Rodney

- 339 A Hayward. Three simple rules to ensure reasonably credible
340 subgroup analyses. *BMJ*, 351:h5651, 2015.
- 341 [7] Wei-Yin Loh Loh, Luxi Cao, and Peigen Zhou. Subgroup
342 identification for precision medicine: a comparative review of
343 thirteen methods. *Wiley Interdisciplinary Reviews: Data Mining
344 and Knowledge Discovery*, 9(5):e1326, 2019.
- 345 [8] Peter M Rothwell. Subgroup analysis in randomised controlled
346 trials: importance, indications, and interpretation. *The Lancet*,
347 365(9454):176-186, 2005.
- 348 [9] Alex Dmitrienko, Christoph Muysers, Arno Fritsch, and Ilya
349 Lipkovich. General guidance on exploratory and confirmatory
350 subgroup analysis in late-stage clinical trials. *Journal of
351 Biopharmaceutical Statistics*, 26(1):71-98, 2016.
- 352 [10] Alex Dmitrienko and Ralph D'Agostino Sr. Traditional
353 multiplicity adjustment methods in clinical trials. *Statistics in
354 Medicine*, 32(29):5172-5218, 2013.
- 355 [11] Yan D Zhao, Alex Dmitrienko, and Roy Tamura. Design and
356 analysis considerations in clinical trials with a sensitive
357 subpopulation. *Statistics in Biopharmaceutical Research*,
358 2(1):72-83, 2010.
- 359 [12] Lynda Anne Szczech, Jesse A Berlin, and Harold I Feldman.
360 The effect of antilymphocyte induction therapy on renal allograft
361 survival: a meta-analysis of individual patient-level data. *Annals*

- 362 of Internal Medicine, 128(10):817-826, 1998.
- 363 [13] David M Herrington, Timothy D Howard, Gregory A Hawkins,
364 David M Reboussin, Jianfeng Xu, Siqun L Zheng, K Bridget
365 Brosnihan, Deborah A Meyers, and Eugene R Bleecker.
366 Estrogen-receptor polymorphisms and effects of estrogen
367 replacement on high-density lipoprotein cholesterol in women
368 with coronary disease. New England Journal of Medicine,
369 346(13):967-974, 2002.
- 370 [14] Fibrinolytic Therapy Trialists. Indications for fibrinolytic
371 therapy in suspected acute myocardial infarction: collaborative
372 overview of early mortality and major morbidity results from all
373 randomised trials of more than 1000 patients. The Lancet,
374 343(8893):311-322, 1994.
- 375 [15] Morris J Brown. Matching the right drug to the right patient in
376 essential hypertension. Heart, 86(1): 113-120, 2001.
- 377 [16] Heart Protection Study Collaborative Group . MRC/BHF Heart
378 Protection Study of cholesterol-lowering with simvastatin in 5963
379 people with diabetes: a randomised placebo-controlled trial. The
380 Lancet, 361 (9374):2005-2016, 2003.
- 381 [17] PROGRESS Collaborative Group . Randomised trial of a
382 perindopril-based blood-pressure-lowering regimen among 6105
383 individuals with previous stroke or transient ischaemic attack.
384 The Lancet, 358 (9287):1033-1041, 2001.

- 385 [18] Janet Woodcock and Lisa M LaVange. Master protocols to
386 study multiple therapies, multiple diseases, or both. *New England*
387 *Journal of Medicine*, 377(1):62-70, 2017.
- 388 [19] Nigel Stallard, Susan Todd, Deepak Parashar, Peter K Kimani,
389 and Lindsay A Renfro. On the need to adjust for multiplicity in
390 confirmatory clinical trials with master protocols. *Annals of*
391 *Oncology*, 30(4): 506-509, 2019.
- 392 [20] Lu Cui, HM James Hung, Sue Jane Wang, and Yi Tsong. Issues
393 related to subgroup analysis in clinical trials. *Journal of*
394 *biopharmaceutical statistics*, 12(3):347-358, 2002.
- 395 [21] Dennis J Slamon, Brian Leyland-Jones, Steven Shak, Hank
396 Fuchs, Virginia Paton, Alex Bajamonde, Thomas Fleming,
397 Wolfgang Eiermann, Janet Wolter, and Mark Pegram. Use of
398 chemotherapy plus a monoclonal antibody against her2 for
399 metastatic breast cancer that overexpresses her2. *New England*
400 *Journal of Medicine*, 344(11):783-792, 2001.
- 401 [22] Edward H Romond, Edith A Perez, John Bryant, Vera J Suman,
402 Charles E Geyer Jr, Nancy E Davidson, Elizabeth Tan-Chiu,
403 Silvana Martino, Soonmyung Paik, and Peter A Kaufman.
404 Trastuzumab plus adjuvant chemotherapy for operable
405 her2-positive breast cancer. *New England Journal of Medicine*,
406 353 (16):1673-1684, 2005.
- 407 [23] Conor A Bradley. Pembrolizumab improves os across pd-11

- 408 subgroups. *Nature Reviews Clinical Oncology*, 16(7):403, 2019.
- 409 [24] Richard M Ransohoff. Natalizumab for multiple sclerosis. *New*
410 *England Journal of Medicine*, 356(25): 2622-2629, 2007.
- 411 [25] Jie Chen, Joseph Heyse, and Tze Leung Lai. *Medical Product*
412 *Safety Evaluation: Biological Models and Statistical Methods*.
413 Chapman & Hall/CRC Press, 2018.
- 414 [26] Robert Hemmings. An overview of statistical and regulatory
415 issues in the planning, analysis, and inter-pretation of subgroup
416 analyses in confirmatory clinical trials. *Journal of*
417 *Biopharmaceutical Statistics*, 24(1):4-18, 2014.
- 418 [27] Van Cutsem, E., M. Peeters, S. Siena, Y. Humblet, A. Hendlisz,
419 B. Neyns, J.-L. Canon, J.-L. Van Laethem, J. Maurel, and G.
420 Richardson (2007). Open-label phase iii trial of panitumumab
421 plus best supportive care compared with best supportive care
422 alone in patients with chemotherapy-refractory metastatic
423 colorectal cancer. *Journal of Clinical Oncology* 25(13),
424 1658-1664, 2007.
- 425 [28] Prenen, H., S. Tejpar, and E. Van Cutsem (2010). New
426 strategies for treatment of kras mutant metastatic colorectal
427 cancer. *Clinical Cancer Research*, 16(11): 2921-2926, 2010.
- 428 [29] Amado, R. G., M. Wolf, M. Peeters, E. Van Cutsem, S. Siena, D.
429 J. Freeman, T. Juan, R. Sikorski, S. Suggs, and R. Radinsky
430 (2008). Wild-type kras is required for panitumumab efficacy in

431 patients with metastatic colorectal cancer. *Journal of Clinical*
432 *Oncology*, 26(10): 1626-1634, 2008.

433 [30] Cappuzzo, F., T. Ciuleanu, L. Stelmakh, S. Cicens, A. Szcz'
434 esna, E. Juhász, E. Esteban, O. Molinier, W. Brugger, and I.
435 Melez'inek (2010). Erlotinib as maintenance treatment in
436 advanced non-small-cell lung cancer: a multicentre, randomised,
437 placebo-controlled phase 3 study. *The lancet oncology*, 11(6):
438 521-529, 2010.

439 **附录 1. 词汇表**

440 **重现性 (Reproducibility)**: 临床试验的结果可在类似的研究条件
441 下重现的能力。

442 **分类器 (Classifier)**: 使用患者的一组特征 (如基因表达) 和分
443 类标签 (如药物应答或耐药) 数据, 拟合一个分类函数或模型,
444 通过相应的特征对患者进行分类或定义亚组, 并可对无标签的患
445 者进行预测。

446 **风险分数 (Risk Score)**: 将多个患者基线特征变量通过一定的方
447 法转换成能够预测患者结局的单一变量, 后者通常称为风险分数。

448 **I类错误 (Type I Error)**: 当原假设为真而检验结果拒绝了原假
449 设的错误, 其概率通常用 α 表示, 并用其作为检验水准。

450 **内在因素 (Intrinsic Factors)**: 指患者个体的遗传、生理和病理
451 特征, 是一个人“内在”的生物特征, 而不是由个人所在的环境决
452 定。

453 **确证性亚组分析 (Confirmatory Subgroup Analysis)**: 按照预先

454 制定的计划，以特定的亚组为研究的主要目标人群，能够对I类
455 错误进行控制的亚组分析，以确证目标亚群患者的疗效。

456 **生物学合理性 (Biological Plausibility)**: 生物学特征与研究终点
457 (如主要终点、不良事件等) 之间因果关联推理的合理性。

458 **适应性设计 (Adaptive Design)**: 按照预先制定的计划，在期中分
459 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
460 设计，这种修改又称为适应性修改。适应性修改计划必须在临床
461 试验开始前的试验方案和统计分析计划中预先设定。

462 **适应性富集设计 (Adaptive Enrichment Design)**: 按照预先制
463 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
464 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
465 ，选择最有可能从治疗中获益的患者入组的自适应设计。

466 **探索性亚组分析 (Exploratory Subgroup Analysis)**: 通过亚组分
467 析探索药物对不同亚组患者可能存在的疗效或者安全性上的差
468 异，此类分析通常不对I类错误进行控制，主要关注的是其结果

469 在生物学或者临床上的合理性。

470 **一致性 (Consistency)**: 在亚组分析中, 不同亚组间显示出相同
471 或相似的治疗效果。

472 **异质性 (Heterogeneity)**: 在临床试验中, 异质性体现在个体
473 和群体两个水平, 前者通常是指患者间具有不同的特征, 个体
474 性质或状态的不同可能会导致不同的患者对治疗有不同的应答
475 ; 后者通常是指不同中心、种族、地域等患者群体具有不同的
476 特征, 有可能导致不同患者群体对治疗有不同的应答。

477 **外在因素 (Extrinsic Factors)**: 指与患者居住地的环境和文化相
478 关的因素。外在因素往往与遗传和基因无关, 更多是由文化和行
479 为决定。

480 **亚组分析 (Subgroup Analysis)**: 将患者根据其特征变量值分成
481 不同的亚组、并估计各亚组的疗效和/或安全性的分析策略。

482 **支持性亚组分析 (Supportive Subgroup Analysis)**: 通过亚组分
483 析探讨药物的疗效和/或安全性在各亚组中的一致性, 为试验全人

484 群的研究结论提供进一步的支持。

中文	英文
安全性亚组分析	Safety Subgroup Analysis
变异度	Variability
病理生理学	Pathophysiology
Bonferroni 单步检验法	Bonferroni Single-step Procedure
重现性	Reproducibility
递归分割树方法	Recursive Partitioning Tree Method
多重性	Multiplicity
多重性调整	Multiplicity Adjustment
非劣效界值	Noninferiority Margin
分层随机化	Stratified Randomization
固定样本量设计	Fixed Sample Design
回退法	Fallback Procedure
获益-风险评估	Benefit-risk Evaluation

中文**英文**

I 类错误

Type I Error

交互树方法

Interaction Tree Method

可信度

Credibility

联合分布

Joint Distribution

描述性分析

Descriptive Analysis

内在因素

Intrinsic Factors

确证性亚组分析

Confirmatory Subgroup Analysis

生物标志物

Biomarkers

生物学合理性

Biological Plausibility

适应性设计

Adaptive Design

适应性富集设计

Adaptive Enrichment Design

探索性亚组分析

Exploratory Subgroup Analysis

外在因素

Extrinsic Factors

原假设

Null Hypothesis

中文**英文**

异质性

Heterogeneity

支持性亚组分析

Supportive Subgroup Analysis

逐步多重检验法

Stepwise Multiple Testing Procedure

主要终点

Primary Endpoint

准确度

Accuracy

总 I 类错误

Familywise Error Rate (FWER)

附录 3. 亚组分析与设计研究案例

示例 1: 帕博利珠单抗免疫治疗小细胞肺癌的亚组分析

在一项 PD-1 免疫制剂帕博利珠单抗 (Pembrolizumab) III 期临床试验中, 共纳入 1274 名既往未经治疗的局部晚期或转移性非小细胞肺癌患者 (EGFR 和 ALK 基因均无突变), 患者 PD-L1 表达水平 $TPS \geq 1\%$ 。将其按照 1:1 的比例随机分配至两组, 分别接受帕博利珠单抗或铂类化疗, 主要终点是预先定义三个亚群 (依次为 $TPS \geq 50\%$, $20\% \leq TPS < 50\%$ 和 $1\% \leq TPS < 20\%$) 的总生存率 (OS)。分析结果显示, 在中位随访期为 12.8 个月期间内, 三个亚组中接受帕博利珠单抗治疗的 OS 均明显长于接受铂类化疗的患者 (第一组: $HR=0.69$, $95\%CI=0.56-0.85$, $P=0.0003$; 第二组: $HR=0.77$, $95\%CI 0.64-0.92$, $P=0.0020$; 第三组: $HR=0.81$, $95\%CI=0.71-0.93$, $P=0.0018$)。试验组和对照组的中位 OS 在三个亚组中分别为 20.0 vs 12.2 (月)、17.7 vs 13.0 (月) 和 16.7 vs 12.1 (月)。据此, FDA 将帕博利珠单抗的适应症扩展到 PD-L1 表达水平 TPS 低至 1% 的非小细胞肺癌患者。

503 **示例 2:那他珠单抗治疗复发-缓解型多发性硬化症的获益-风**
504 **险评估**

505 那他珠单抗 (Tysabri) 是第一个由 FDA 批准用于治疗多发
506 性硬化症 (MS) 的单克隆抗体。2004 年批准后不久该药就被迫
507 退市, 其原因是两名患者在服用了该药物后发生了进行性多灶性
508 白质脑病 (PML), 这是一种罕见的不可治愈且易于致命的脑部
509 感染疾病 (18 个月的随访期内, 约有 1/1000 的患者并发感染
510 PML)。一年后, FDA 咨询委员会开会, 重新对患者使用该药的
511 获益与风险进行了评估, 认为与现有的其他治疗方法相比, 该药
512 更有效, 患者完全可以接受用药带来的风险。于是, FDA 于 2006
513 年宣布重新上市 Tysabri, 但只限复发-缓解型多发性硬化症
514 (RRMS) 亚型患者在其他治疗失败或不耐受时使用。这一限制
515 主要是出于安全性考虑, 即对活动性不高的非 RRMS 亚型患者可
516 能受益有限但却面临发生 PML 的风险。