

类风湿关节炎治疗药物 临床研究指导原则

**Clinical Development Programs for Drugs, Devices, and
Biological Products for the Treatment of
Rheumatoid Arthritis (RA)**

1999 年 2 月 美国 FDA 发布
2009 年 11 月 药审中心组织翻译
辉瑞制药有限公司翻译
百华协会审核
药审中心最终核准

目 录

I. 前言	1
II. RA 治疗的新声明	2
A. RA 体征和症状的减轻	2
B. 主要临床缓解	3
C. 完全临床缓解	3
D. 缓解	3
E. 预防残疾	3
F. 预防器质性损害	3
III. RA 产品的研发	4
A. 临床前研究	5
B. 药代动力学/药效学研究策略	8
C. I 期临床试验	8
D. II 期临床试验	10
E. 疗效试验	11
I. 样本量	20
F. 安全性分析	20
IV. 生物制品的特殊问题	21
A. 种属特异性	21
B. 剂量效应	21
C. 毒性反应	22
D. 产品均一性	22
E. 抗体的作用	22
V. 医疗器械的特殊问题	22
A. 背景	22
B. 有效性	22
C. 安全性	23
IV. 幼年类风湿关节炎的特殊问题	23
A. 背景	23
B. 儿科用药管理规定的实施及其对 JRA 研究设计的影响	25
C. 结果变量和结论声明	25
D. 试验设计问题	27
E. 合并使用的抗风湿药物	27
F. 多中心试验和中心效应	28
参考文献	28
附录 A: 回应率比较试验	31

类风湿关节炎治疗药物临床研究指导原则^①

I. 前言

本指南旨在帮助开发者从事类风湿性关节炎（RA）治疗药物、生物制品及医疗器械的研发工作。本文中讨论了此类产品应考虑说明书声明的描述类型，并为支持说明书声明内容的临床研究方案的设计提供了指导。

说明书所载内容的核心目的是告知开处方者及患者一种产品已证明的获益。由于 RA 是一种慢性的、有明显症状的疾病，不同患者的治疗结果在病程、严重程度及总体影响程度上亦不尽相同，任一治疗结果均可作为一项说明书声明的基础。

症状减轻——说明书中关于症状及体征的描述——是 1997 年前后上市的大多数 RA 治疗药物的核心疗效。然而这种描述的含义过于宽泛，涵盖了多种 RA 治疗结局，超出了之前被认可的可列入说明书 RA 治疗声明的范围。基于此，本指导原则规范了当患者的获益不仅为“症状减轻”的程度时，如何准确阐述患者的获益程度。例如，主要临床缓解、完全临床缓解和缓解（与完全临床缓解的标准相同，但是需在停用所有抗风湿药物情况下）反映了对于疾病体征及症状更强的治疗效果。预防器质性损害的声明是经过各种放射影像学检查证实的。预防残疾的声明反映了病程中的长期受益。本行业指南草案中讨论的说明书声明及临床研究方案代表了权威机构风湿病学家对目前临床试验中可获取的、可供评价的临床相关总体治疗结果的普遍观点。

传统上，RA 治疗药物被分为改善病程的抗风湿药物（DMARDs）和非甾体抗炎药物（NSAIDs）两大类。对 RA 自然病程具有积极影响且更有效的治疗药物仍在不断的探索中，某些具有良好前景的新治疗药物正在接受临床的验证。处于研究阶段的许多 RA 治疗新药依据推测的作用机制，无法被上述两种分类标准涵盖。因此传统的分类方法显示出一定的局限性。为此，说明书中会提供一些关于治疗起效时间、作用持续时间及疗效持久性等方面的信息，旨在反映临床试验中所收集到的数据。因此，本文中描述的某些说明书声明在其行文中加入了缓解持续时间。

在过去的十年里，人们一直在探索如何更好的评价 RA 临床试验中患者的治疗结果。许多学术组织包括欧洲抗风湿病联盟、国际抗风湿病联盟、美国风湿病学会（ACR）以及类风湿关节炎临床试验结果评价标准小组（OMERACT），均试图制定核心评价指标以及综合评价指标来描述患者的治疗结果。通过这些努力，目前数个新的评价指标已经被提出并被临床数据所验证。我们希望这些新的评价指标能够对患者治疗结果提供更有用的信息，

^① 该指南由药品审评及研究中心（CDER）的医疗政策协调委员会（MPCC）类风湿性疾病工作组、生物制剂审评及研究中心（CBER）、器械及放射卫生中心（CDRH）撰写。该指南代表了监管机构目前对于类风湿性关节炎治疗药物、医疗器械及生物制品临床研究方案的观点。其不为任何人建立任何权力或赋予任何人任何权力，并且不约束 FDA 或公众。本指南以外的其他方法也可以被使用，如果该方法满足相应的法令、规定（或两者都满足）。

因此本文对在临床试验过程中如何使用这些新评价指标进而支持说明书声明提供了指导。

有一项结果评价指标，即总体健康相关的生活质量（HR—QOL），虽不能完全依靠它来支持一项说明书声明，但由于 RA 影响患者生活的多个领域，我们期望 HR—QOL 指标可以为治疗干预手段对患者的长期影响提供一项综合评价。然而评价患者长期的生活质量需要较长期的关节炎试验，在这期间对于总体 HR—QOL 指标评价所需的信息往往是获取不够充分的。即便如此，我们仍鼓励在制定临床试验方案时纳入这类评价指标。

II. RA 治疗的新声明

目前，一些新的疗效声明正在药物研发过程中接受临床评估。这些疗效声明的描述方法以及支持这些声明的可被接受的治疗结果评价标准将在下述段落中讨论。

A. RA 体征和症状的减轻

该项声明旨在反映症状上获益的临床证据，包括疾病活动体征及症状的改善。体征和症状减轻通常是药品获准上市之初的疗效声明。通常情况下，该疗效声明需要至少为期 6 个月的临床试验予以支持，除非该药品属于前已述及的药理学分类的药物（如 NSAIDs），该情况下为期 3 个月的临床试验即足以证明药物在改善体征和症状方面的疗效。推荐进行为期 6 个月的试验有如下几个原因。首先，RA 是一种长期的疾病。仅进行短期的、有限时间内的干预，可能不会对患者有整体的治疗价值。另外，因为抗体会阻断药物疗效，所以可能引起抗体形成的药物应评价其疗效的持久性。在评价体征和症状时，推荐采用实时疗效评价（即重视治疗过程中的评价）方法，而不是仅仅评价治疗前的基线及治疗终点的症状，除非有充足的理由证明最后一次访视时对症状的评估明显优于过程中的评估。可接受能够支持疗效声明 A 的疗效评价指标包括：

1. 体征和症状的综合评价终点或指标

这些综合指标可用于构建患者成功或失败的分类标准。例如，Paulus 标准（1990）或 ACR 病情改善标准定义（ACR 20）^① 可用于评价患者的病情缓解情况。

说明：在为期 6 个月的试验中，患者治疗成功的标准是在为期 6 个月的观察中，至少有 4 个月满足相对于治疗基线有明显改善的评价标准，并且没有因为药物的毒性而脱落。

2. 众所接受的体征/症状指标集

在 1988 年 CDER 的抗炎及抗风湿药物临床评价指南（FDA 1988）中曾推荐了 4 个指标，即疼痛、压痛及肿胀的关节计数和医生、患者的整体评价；此外，ACR 核心集也是众所接受的体征和症状评价指标系统的例子。治疗成功的标准和统计分析方法应预先定义并达成一致。例如，历史上在使用关节计数和整体评价指标时，曾使用了在 4 个评价指标中至少 3 个指标在对照组和治疗组之间具有统计学显著差异作为试验成功的标准。但是，如前所述，只比较治疗前的基线和治疗后的最终观察点不是分析疗效的最佳方法，因为这种方法忽略了所有中间过程的疗效观察。

^① ACR 病情改善标准定义（ACR 20）为压痛和肿胀关节计数指标改善 20%，以及余下的 5 个核心指标中的 3 个指标改善 20%，这 5 个指标分别为：患者和医生总体评价、疼痛、残疾和急性期反应物（Felson 1993, 1995）。关于具体指标的详细描述（如残疾、急性期反应物），请参考 Felson 1995。

对于上述两种指标，适合使用 66 或 28 关节计数（Smolen 1995）。

B. 主要临床缓解

该项疗效声明旨在反映连续 6 个月治疗成功的临床证据，基于“ACR 70”评价标准。“ACR 70”的标准完全与 ACR 20 类似，除了是需要 70% 改善而非 20% 改善作为被评价部分。该项疗效声明是基于 ACR 70 标准体系下，治疗组较随机对照背景治疗组在连续 6 个月缓解率指标上的具有统计学意义的显著改善。本指南附录 A 中列出了在以往的两个历史数据库中满足各种 ACR 标准定义（从 ACR 20 到 ACR 70）的患者数量，供参考。对于预计能够快速起效的药物，试验周期应至少为 7 个月，而对于短期内不会产生疗效的药物，临床试验周期应更长。

C. 完全临床缓解

该声明旨在描述治疗效果优于主要临床缓解声明的治疗获益。完全临床缓解和缓解（见下文）定义相同，均为持续 6 个月的缓解，该处所指的缓解既要符合 ACR 对缓解的标准定义也需放射影像学资料证实疾病处于静止状态（即无放射影像学提示的疾病进展 [Larsen 1977] 或改良 Sharp 方法 [1985]）。完全临床缓解是指需要持续药物治疗才能获得的受益；而缓解定义为停用所有抗风湿药物后相同的受益。1981 年 ACR 对于缓解的标准（Pinals 1981）要求以下描述中的至少 5 项：晨僵持续时间小于 15 分钟，无疲劳，无关节疼痛史，无关节压痛或运动疼痛，关节或腱鞘无肿胀以及红细胞沉降率（ESR）小于 20（男性）或小于 30（女性）。支持该项疗效声明的临床试验周期需视受试药物的起效速度而定。对于大多数药物（除了快速起效药物），应计划为期 1 年的试验。极慢作用的药物可能需要更长期的试验。评价完全临床缓解的试验应使用分类终点（患者完全缓解或治疗失败）作为首要治疗效果评价指标。

D. 缓解

该项声明定义为停用所有抗风湿药物治疗后，符合 ACR 病情缓解标准且放射影像学资料提示疾病处于静止状态（即无放射影像学提示的疾病进展 [Larsen 1977] 或改良 Sharp 方法 [1985]）持续超过 6 个月。缓解不表示治愈，即使患者在 6 个月后或更长时间后复发均可称为缓解。支持缓解声明的临床试验周期视受试药物的起效速度而定。通常该类临床试验至少应计划为期 1 年的试验。慢速作用的药物可能需要更长期的试验。

E. 预防残疾

该项声明旨在鼓励开展 RA 的长期试验。目前，健康评价问卷（HAQ）（Fries 1982）和关节炎影响评估量表（AIMS）（Meenan 1982）已经证实足以应用于该类试验，作为主要的治疗效果评价指标。研究周期应为 2~5 年。寻求该项治疗声明的申办者应考虑之前已经证实了，或者随着试验的进展将会证实患者在体征和症状上的改善。另外，若不使用更全面的 HR—QOL 评估体系，则无法获知 RA 患者的全面影响，因此还应采集其他经验证的评价指标，诸如 SF—36，并且在试验期间患者在这些指标上不应表现出恶化。

F. 预防器质性损害

预防器质性损害是 RA 治疗的一个重要目标。评价该治疗结果的试验周期至少为 1 年。以下列举了可用于支持“预防器质性损害”疗效声明的结果评价标准：

1. X 射线检查提示疾病进展缓慢，使用 Larsen、改良 Sharp 或其他经验证的放射影像学技术

放射影像学声明应基于治疗 1 年后（以及随后的每年时间点）的影像资料与治疗前基线时的比较。所有随机患者均应具有这两个时间点上的影像学资料，无论其是否继续治疗。脱落患者应在脱落时具有影像资料。在这些试验中，预先规定脱落病例的处理方法尤为重要。

2. 预防新的侵蚀性病变（X 光片检查提示）——维持无侵蚀性病变状态或预防新的侵蚀性病变

评价该疗效声明的试验通常使用分类终点标准，通过比较治疗后最终状态与治疗前基线状态，来判断每个患者的状态为病情进展或无进展。

3. 其他评价工具（如 MRI）

其他技术，如 MRI（核磁共振成像）或超声影像学均可以采用。但是，这些技术可识别虽然具有统计学显著性，但差异却很小的病情变化，不利于反映实际患者之间受益程度的差异，需要进一步确定。

由于放射影像学检查提示的疾病进展缓慢从本质上说并未定义患者的受益，因此在“预防器质性损害”疗效声明提交时，我们希望该药物的疗效声明中不仅仅包括“预防器质性损害”，该药物最好是已经被证明或在试验过程中逐步被证明具有其他疗效的药物（如可预防残疾）。然而，实际情况是某些药物不是用来治疗急性期炎症反应的，而是被设计用来预防或延缓关节的破坏。这类药物其首要的临床功能主治可能就是延缓放射影像学进展。而延缓关节破坏的最终目标是改善症状，以及保护关节功能。因此，延缓疾病影像学进展被认为是反映 RA 患者总体获益的替代标志。

根据 21 CFR 314 下面的章节 H 和 21 CFR 601 下面的章节 E，FDA 可基于药物对替代标志的作用，批准治疗严重和致命性疾病的药物上市，前提条件是该药物满足既定的标准，并且承诺在上市后完成进一步评价药物的实际临床受益的研究。在病情严重的 RA 患者人群中，影像学进展显著减缓的证据可满足这些规定的要求。建议申办者在基于 FDA 这些规定开展临床试验之前，与 FDA 相关人员协商。

一个对放射影像学进展有显著影响的实例为，在随机对照试验中，治疗组大部分患者可维持无侵蚀状态，而对照组患者发生多处侵蚀。对放射影像学检查所见的评估方法、考虑为具有临床显著性（无论是患者自我评估还是治疗组评估）的变化程度以及统计分析方法均应在临床方案中预先定义，并且申办者应该寻求管理当局对于疗效评估方案的赞同。利用加速审批途径需要及时完成 4 期临床试验，采用可接受的临床终点评估标准来评价体征及症状的改善及对残疾的预防。申请快速审批需要进行为期 1 年的临床试验，而上述 4 期临床试验将作为快速审批申请后研究的扩展和延伸。

III. RA 产品的研发

下述关于非临床和早期临床产品研发的信息主要属于药品（药物和生物制品）范畴。

除了最初两部分外，以下描述的一般原则也适用于医疗器械。器械研发详细的信息请参考本文中标题为“关于医疗器械的特殊考虑”部分。复合治疗产品（如生物制品和器械）研发者可请求 FDA 的帮助，指定一个产品研发评估领导小组。此类请求应提交至：Office of Chief Mediator and Ombudsman (HF—7), Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857。

在 RA 产品研发中常遇到的问题包括：

1. 选择合适的体外（动物或人类系统）和体内动物模型来筛选具有潜在活性的药物。
2. 设计并进行合适的非临床安全性研究，以支持新分子实体在志愿者或患者中的使用。
3. 平衡患者在疾病早期对治疗干预的潜在需求，避免将轻度疾病患者过早的暴露于具有毒性或极少安全性记录的药物之下。
4. 发现联合用药的潜在风险，特别是对于那些具有相同的靶器官毒性或可能存在药代动力学相互作用的药物。

5. 设计充分且可行的长期安全性监测方法。

6. 设计可明确证明临床疗效的试验。

下述段落讨论了上述问题的解决方法。

A. 临床前研究

该部分着重于抗风湿治疗药物特有的临床研发相关的临床前研究问题。在设计毒性研究以及该类研究的时程时，建议对现行的药物、医疗器械及生物制品的建议及指南等问题咨询管理当局。非临床安全性试验指南由人用药物注册技术要求国际协调会议（ICH）制定，该指南阐述了毒代动力学、生殖毒性、遗传毒性和致癌性研究的要求和设计方法。由于生物制品在动物研究设计中面临特殊的挑战（例如种属特异性结合，或人源蛋白在动物体内潜在的免疫原性），因此 ICH 制定了专门的指导文件来阐述生物技术来源药物的安全性评价方法（ICH S6 1997）。^①

1. 药代动力学

在早期新药研究（IND）阶段开展药物吸收、分布、代谢和排泄的动物研究非常重要，其可帮助解释毒性研究的结果，但不需要在 I 期临床试验前全部完成。一般来说，为了开展人类早期临床试验，仅需要在动物试验中确定药代动力学（PK）参数，如曲线下面积（AUC）、血药浓度曲线峰浓度（C_{max}）和半衰期（t_{1/2}），则足以预测安全的临床暴露量提供基础。

过去，在早期临床试验前，联合用药（或生物制品）的疗效评价一般不在非临床试验中进行。然而，目前多种药物均用于 RA 患者的治疗，包括 NSAID、麻醉药、皮质激素和 DMARDs，进行常见伴随药物的非临床和临床试验是有帮助的。另外，为了评价潜在的药物相互作用，可能需要关于伴随治疗对于药代动力学影响的信息来优化给药方案，并发现潜在的安全性问题。代谢相互作用常可在动物或人类肝脏切片、微粒体制备物或纯化 P450

^① ICH 文档可通过 FDA 主页 (<http://www.fda.gov/cder> 或 <http://www.fda.gov/cber>) 获得。

酶的体外系统中评价（FDA 1997）。

相互作用也可能由于个体或疾病特异性因素引起，如类风湿因子可与各种单克隆抗体治疗药物结合；在这种情况下，鉴定高滴度患者的体外结合研究可能有助于发现独特药代动力学或临床缓解形式的患者。

2. 生物活性

潜在的抗风湿治疗药物的生物学活性应使用多种临床前模型系统（如体外、体内、离体）来确定。体外筛选可使用动物或人源的细胞或组织，一般用于筛选对某个分子靶点具有预期作用的候选药物。这种试验也能够用于为所选药物设计适当的生物检测方法。然后使用健康、具有风湿病（自发的或诱导的）、或经基因修饰的动物来确定生物活性是否能在体内得到验证。所使用的体内模型应模拟类风湿关节炎的某一个或几个方面或模拟其病因，但每种动物模型均具有局限性。

a. 体外

体外研究数据有助于阐明一种药物或生物制品可能的作用机制，并有助于为体内药物药效或毒性评价确定特殊动物种属间的相关性。当临床前研究可能发现潜在的替代标志时，那么这些体外研究数据将尤为重要。例如，如果产品是作用于淋巴细胞 CD4 受体的，则该受体既可作为活性评价的替代标志也可作为毒性评价的替代标志。

多种体外试验可以被采用，这取决于受试药物或生物制品的作用机制。例如，体外结合试验对于研发受体拮抗剂或单克隆抗体很有帮助。体外功能试验（如血小板和中性粒细胞聚集试验）对于发现炎性介质的抑制剂是很有帮助。酶试验（如环氧化酶、脂氧合酶及磷脂酶的体外或离体抑制试验）对于确定药物对同功酶抑制的选择性很有帮助。

b. 体内

动物模型的选择应根据药效（PD）反应、动物疾病与临床疾病病因的相似性，和/或确定毒性作用的机制。理想情况下，用于治疗关节炎患者某一方面症状的产品，应在与该目标症状最相关的实验模型中研发。例如，大鼠对 5-脂氧合酶抑制剂不敏感。因此，小鼠或家兔模型在评价白三烯抑制剂的抗炎活性时更为相关。

鼓励研究开发用于筛查潜在的 RA 候选药物的类风湿疾病模型。下文仅阐述了目前应用的一些模型，并不排除其他模型的使用。

胶原诱导性关节炎（CIA）：

胶原诱导性关节炎由于具有主要组织相容性、完全 II 型限制性 T 辅助细胞激活以及组织病理学病变的相似性，常被认为是研究治疗人类 RA 潜在治疗药物或生物制品的适宜模型。CIA 模型关节的 X 光影像学检查常显示出与人类风湿关节炎相似的侵蚀性病变。关节炎进行性发展常导致 RA 样关节畸形和功能障碍。在 CIA 模型中同样也产生在某些 RA 患者体内产生的抗胶原抗体。

CIA 模型可用于免疫抑制剂、甾体激素以及炎性介质抑制剂的评价。由于该模型能够在多种动物种属中建立，其在评价某些具有种属特异性的药物时（如白三烯拮抗剂和 5-脂氧合酶抑制剂）显得尤为重要。另外，虽然该模型并不常规进行功能试验，但加入移动能力和关节功能评价可增强该模型的预测价值。

自发性关节炎或自身免疫反应：

MRL/lpr 小鼠、Biozzi H 小鼠和 DBA/1 小鼠已用于研究药物诱导性耐受的发生，以及免疫抑制剂对自身免疫反应的影响。MRL/lpr 小鼠模型已用于免疫抑制剂和激素类药物的评价。

大鼠角叉菜胶诱导的急性炎症模型：

该模型一直以来对于环氧合酶抑制剂抗炎活性的评价很有帮助。大多数爪部有炎症的动物模型均可用于评价药物的消炎作用。

大鼠佐剂性关节炎（AA）：

大鼠 AA 模型常用于筛查非甾体抗炎药物、炎性细胞因子抑制剂、以及抗代谢类免疫抑制剂。

链球菌细胞壁诱导的关节炎：

该模型已用于细胞因子抑制剂的研发。

实验性器官移植动物模型：

该模型已用于鉴定免疫抑制剂和抗代谢药物的活性，特别是那些直接作用于溶细胞性细胞免疫过程的药物。

转基因动物模型：

目前，很多转基因动物模型正处于开发阶段，其在类风湿关节炎研究中的作用很可能在未来的十年内被证明。包括携带人类 T 细胞白血病病毒 I 型基因组中 env—Px 结构域基因、人类 TNF、CD4 和 HLA B—27 基因的转基因小鼠。

3. 毒理学

药物或生物制品的临床前毒理学研究用于描述一般毒性及特殊毒性，使用的给药途径及方案应尽可能与计划的临床试验相似，并考虑目标患者人群的人口统计学特征和疾病状态。例如，RA 在女性中发生率较高。因此，生殖毒性研究应在研发早期阶段完成，以支持在临床试验早期阶段纳入育龄妇女病例。生物制品由于其免疫原性和种属特异性的复杂性，其生殖毒性研究的要求则需要具体情况具体分析。因此，标准化研究设计（如 ICH 生殖毒性研究指南中推荐的）对于生物制品可能不具备可行性或临床相关性（ICH S5 1994）。这些研究的要求及特殊设计方案可与监管机构的审评人员讨论。

作为单药治疗或联合治疗给予 RA 患者的免疫调节剂或免疫抑制剂，也带来了长期免疫功能低下等不良反应的担忧。例如，移植手术的患者长期接受非选择性免疫抑制剂，已知的风险是易患恶性肿瘤（如淋巴瘤）。处于研究阶段的药物相关的机会性感染和免疫抑制相关的死亡情况已在 RA 患者中发生。我们鼓励申办者发现并使用有助于筛选选择性抑制 RA 相关细胞及疾病过程的候选药物的动物模型。

抗风湿药物常联合使用，以期改善效果并减小毒性。然而，药物相互作用可导致毒性增加，即便在低于曾评价过的各个药物剂量的情况下。该问题对于具有长半衰期或非选择性活性的药物、或对于具有共同毒性靶器官的药物来说尤为明显。在临床前毒理学研究阶段进行联合用药的试验，可对预测临床安全性危害方面有所帮助。动物在毒性剂量下的给药持续时间通常取决于患者的给药方案。我们鼓励建立并验证体外或整体动物模型，用以

阐明药物短期或长期毒性的相关问题，也可为患者免疫功能的评价建立替代标志。

B. 药代动力学/药效学研究策略

体内药代动力学研究用于评价药物在体内的分布及代谢、线性程度及蓄积、剂量比例、以及对于口服制剂，还应评价与食物的相互作用（Peck 1992）。其中某些数据可以通过设计一个单独的可评价较多指标参数的试验中获得。在剂型研发期间，推荐进行与剂型相关的生物等效性研究。

对于生物制品有一个特殊的担忧，那就是体内可能产生抗体，从而加速了药物在体内的清除或改变药物的体内分布，继而导致随着时间的推移疗效发生改变或在治疗过程中需要重复给药。为了阐述该问题，建议申办者应在重复给药的临床方案中加入药物水平、受体饱和度、抗药抗体和临床反应等参数的综合评价。这些评价最好在治疗开始时以及治疗过程中的数个时间点上进行。抗药抗体的存在，以及它们在改变药物暴露量、临床疗效终点或药物不良反应中的作用均需要进行评价。综合分析这些参数的目的是为给药剂量的选择及治疗方案的变更提供支持数据，用以优化临床疗效。进行这些药代动力学研究的最佳时间是在 III 期临床试验之前、给药剂量及给药方案已经确定之前。

由于风湿病的治疗多采用联合用药，故采集处于疾病活动期的患者血液来进行体外结合试验，可作为初步筛选潜在置换反应的工具。

对于可能与类风湿因子相互作用的药物（如单克隆抗体），应评价类风湿因子阳性患者对于抗体的反应程度，以及相互作用对药物药代动力学的影响。

C. I 期临床试验

关于大多数药物及生物制品相关的一般临床研究信息，参见 CDER 指南药物临床评价的一般考虑（FDA 1978）。

术语 I 期具有 2 个含义：一个为最早期的、首次在人类中进行的试验，而另一个含义包括了前已述及的药代动力学、代谢、药物相互作用、特殊人群研究，及其他临床药理学试验。两种含义下的 I 期试验均在 RA 治疗药物的临床评价期间进行。本节主要讨论了人类首次暴露于药物（包括特定的剂量水平、联合用药及治疗周期）相关的问题。

1. 条件配置及研究者

首次在人类中进行的 I 期研究应在具备足够的临床和实验室设备的研究机构进行，患者应受到密切观察。试验最好应在具有早期药物研发及风湿病学经验的医生指导下进行，或者在具有风湿病学和临床药理学经验的研究者团队的指导下进行。

2. 受试者

首次在人类中进行的药物试验通常在健康志愿者中进行。这类研究的开展是基于一定可行性基础上的，并能够解释临床前动物试验的结果。如果临床前试验未显示药物潜在的致突变作用或对免疫系统的影响，或者在或接近预期治疗剂量范围内未出现潜在的严重影响，则可以开始在健康志愿者中进行试验。但是，对于具有潜在严重毒性的药物，最好选择有一定可能从中受益的患者来进行。在这种情况下，如何选择合适的初始受试患者则成为一项挑战。

对于已经在相关临床前毒性评价中显示出相对安全（在拟应用剂量下无潜在致突变、

免疫系统影响或其他严重影响)的药物,可以在健康志愿者中开始试验。然而,对于已经证明存在或可能存在显著不良影响的药物,则有必要选择适当的患者人群。我们推荐被选择的患者应在诊断和 RA 活动程度方面均满足 ACR 标准,且无其他严重合并症。疾病程度较轻的患者有时也不宜选择,因为其与健康受试者的状态接近,而该类试验不宜选择健康志愿者进行。具有严重器质性损伤的 RA 患者因伴有临床合并症亦不宜被选为初始受试者。而且,由于疾病程度过于严重的患者通常对治疗不太可能有明显的反应。

目前,正在进行流行病学工作以期鉴别出预示 RA 风险增加的相关标记。这可能有助于鉴别出预后较差的患者,而这些患者可被考虑使用潜在毒性较大的侵袭性治疗(如免疫清除性治疗结合干细胞移植)。流行病学研究结果的应用,可使这些侵袭性治疗被限制在特定的 RA 患者中进行,这些患者基于其自身的病情已被证明具有较短的存活期限(如大于 30 个受累关节或 HAQ 评分中回答“无困难”的问题小于 75% 的患者)。

当药物显示出可能对性腺功能具有长期的影响时,应选择近期无生育计划的男性和女性进行 I 期研究。

3. 试验设计

一般来说,早期 I 期临床研究采用剂量逐步递增的方法给药,这样某一特定剂量下的安全性和耐受性可在其他受试者暴露于更高剂量之前即可确定。几乎总是先进行单次给药研究,然后进行多次给药研究。但是这种设计也受所用药物类型的影响。虽然剂量递增至最大耐受剂量(MTD)将有助于下一步临床试验的设计,但在某些情况下,试图充分观察到 MTD 剂量下的表现在医学上是不明智的。另外,对于某些药物来讲其 MTD 可能无法定义。

所选起始剂量常为无不良反应剂量(通过物种间毫克(mg)/平方米/天的剂量换算,从动物换算至人类来确定)。对于生物制品,所选的初始剂量常为无不良生物学反应的剂量,并需充分考虑相关的种属特异性,并比较受试动物与人类的受体亲和力的差异。通常推荐使用保守的剂量递增(如半对数或更低)方案。

4. 伴随治疗

小剂量皮质激素(相当于强的松 10mg/天)和 NSAIDs 通常在 I 期试验中仍然持续使用。在所有新的抗风湿药物、生物制品及医疗器械的初始 I 期临床试验中,应避免联合使用甲氨蝶呤及类似的药物,因为这样将难以将新药的毒性和合并用药的毒性区分开来。

目前,医生在类风湿关节炎治疗早期就开始给予患者甲氨蝶呤及类似药物治疗。故招募足够数量的从未服用过这些药物的患者可能存在困难。允许在 I 期试验后期使用甲氨蝶呤及类似药物的情况包括(a)从联合用药的相关动物模型中获得了药物无毒的可靠证据,(b)以显著低于单个药的无不良反应(通过早期的 I 期研究或临床前研究证明,或二者均证明)的剂量开始试验。此类研究应在方案制定阶段与监管机构人员进行讨论。

5. 观察

a. 安全性

安全性观察的标准指标已在其他文件中描述(ICH S5A 1994)。然而,其他安全性观察可能也有必要进行(如对细胞和体液免疫功能的影响或宿主防御功能影响的试验)。对于

那些给药后的效应持续时间较长、或可能具有延迟毒性的药物，应设计适当的随访。例如，对用于去除或改良 T 细胞亚群功能药物的 I 期研究，应在治疗期间和随访期间详细评估其对细胞群体数量和功能状态（例如迟发型超敏反应），以及对其他相关的药效学指标的短期和长期影响。

在方案设计中建议加入个体患者不良事件的终止/退出规则。另外，我们也建议在试验方案中加入当观察到不良事件发生时，试验停止或试验方案修订的规定。例如，在剂量探索试验中应明确定义剂量递增规则，还应规定如果观察到可能的严重不良事件，还要在小于或等于引起毒性的剂量下入组额外的患者。

我们推荐基于产品以及该类药品已知的和预知的毒性，建立标准的毒性评级量表，应用于该产品的所有临床试验中。这样可以提高不良事件报告的一致性，使得试验全程的数据可进行更准确的比较。

b. 疗效

在早期试验中了解药物的疗效潜力，对于有效的产品研发很有帮助。这可在 I 期试验中开始尝试，但通常只有通过进行对照试验后才能完成。开放性试验中观察到的 RA 治疗反应在证明疗效方面基本上没有参考价值。应考虑更合适的目标，确定基于临床前研究所预测的药理学作用是否在临床观察中也同样存在（理念的证明）。

D. II 期临床试验

II 期试验期间，通常进行更大规模、更长周期的试验来充分的确定与剂量和暴露量相关的活性和毒性。从 II 期试验中应产生足够的信息，以确保 III 期试验能够安全、成功地进行。另外，II 期试验应为总的药物研发策略提供坚实的基础，确保在 III 期安全性/疗效试验完成后，所有注册所需的信息都是完整的，包括适当的安全性数据库、临床药理学、剂量反应数据、特殊人群中的探索（如肾功能衰竭患者、肝功能衰竭患者、儿科患者），以及与预期联合使用的药物相互作用的信息。

在 III 期试验过程中，也可同时开展补充的 I 期临床药理试验和 II 期试验。

下述问题对于 RA 的 II 期试验非常重要：

1. 试验设计

剂量探索是 II 期试验的核心挑战。一旦确定了合理的安全剂量范围，通常推荐进行随机、平行剂量对照试验。需求使用安慰剂是基于多种原因的。首先，如果各个研究剂量间无差异，则通常没有其他方法能确定是所有剂量都同样有效还是都同样无效。第二，如果在临床试验中观察到了剂量效应关系，通过与安慰剂组对比可以判断所观察到的作用的程度。如果无法使用安慰剂，方案中应包括较宽的给药剂量范围、较长的给药周期或重复。已知的疗效明确的阳性对照设计也很有帮助。

体征和症状指标可用于剂量探索研究，但是有时可能需要设计单独的剂量探索试验来满足长期观察终点的需要。

对于具有即刻起效和快速失效的药物，包括交叉和滴定在内的其他替代方案设计可能有所帮助，虽然以往并未如此。通常，在 ≥ 2 个剂量组内允许根据患者的疗效程度而进行自由滴定的试验无法清晰的阐明剂量效应关系，因为这种滴定设计会导致不同剂量组之间

的差异不明显。

我们非常希望通过 III 期临床试验，能够确定适宜的临床给药剂量范围，在该范围内药物具有可以接受的毒性和合理的疗效。

2. 安全性

每个处于研究阶段的 RA 药物均存在安全性担忧。当药物存在潜在的严重毒性、在体内消除缓慢或起效延迟时，应在进行较大规模的 III 期研究前的 II 期研究中，设计一组患者接受更长期的随访。提供长期的随访资料将有助于解释药品上市前评估过程中提出的问题（如潜在的免疫抑制作用、机会感染、肿瘤形成以及诱导自身免疫疾病的可能）。在 II 期研究中推荐采用标准的毒性分级量表，并制定何种情况下需要停药的规则。

3. 研发相关的其他方面

a. 伴随治疗

开始 III 期试验前，需评价供试产品与其他药物（应用于目标患者的可能的其他药物）之间的相互作用。通过代谢途径研究、体外系统的研究、动物或人类药理学研究或药物相互作用研究可建立初步的信息。该类信息有助于指导那些需要临床试验来进行评价的领域。在临床试验中，拟与受试药品联合给药的产品，其与受试药品之间的相互作用及安全性相关的充分的信息应在 II 期试验中研究并获取。

b. 性别影响

大多数 RA 试验入组的患者多为女性。申办者应评估所观察到的安全性及疗效结果是否仅局限于女性，或者也可以外推至男性受试者。这可通过试验中的亚集分析、PK 数据或其他信息来确定（FDA 1993）。

E. 疗效试验

III 期试验工作的总体目标，是通过严格的对照试验来证明产品疗效，并积累足够的安全性数据。疗效试验方案中应包括分析方法，在分析方法中明确规定进行初步比较的具体条件，试验成功的标准、以及统计检验的分析方法。方案设计时应考虑所获取的临床数据应可充分支持拟定的说明书声明。任何其他计划开展的、正在进行的或已经完成的试验，只要也是用来获取数据来支持说明书声明的，也必须在方案中明确阐述。

1. 整体考虑

a. 患者选择

疾病活动程度：除非以某些其他特殊的患者亚群为目标，否则疗效试验中入组的患者应至少满足疾病的诊断标准，并符合 ACR 标准定义的疾病活动程度。关于说明书声明（源自具有严格病例纳入标准的临床试验）是否具有普遍性和代表性的问题，推荐与监管机构协商。

为了加强试验的效力，常使用某些策略来提高患者对治疗的反应机率。某些方案中试图纳入那些通过停掉所有正在进行的治疗继而诱发出疾病突发状态的患者。只有具有足够高疾病评分的患者才被入选。这种观察到的“疾病突发”类型，其相关性是令人质疑的，用该类患者能否预测正常的疾病活动过程尚不明确。在这类研究中，很多患者被随机分到

安慰剂对照组，在无进一步治疗的情况下，快速表现出回复至治疗基线状态。另外，当患者在这类试验盲态下退出治疗时，类似的“疾病突发”情况无法被观察到。这就带来一个问题，即是否在被告知疾病突发过程的患者方面存在预期偏倚，还是在希望患者满足入选标准并入组的研究者方面存在确定偏倚。当实施这类方案时，应考虑到这类试验中所用的结果评价指标的不确定性和不稳定性。

在不同于上述使用“疾病突发”策略的试验中，患者通常被要求必须达到某一最小评分标准后才可考虑入组，但我们也观察到这些患者的关节评分逐步渐小但仍显而易见且快速的回归至平均值。这意味着患者，总体来讲，实际的疾病活动程度与被期望的入组标准规定的疾病活动程度不同。

根据疾病标记将患者分为亚组：RA 很可能包括了一些或多或少的不同的疾病，这些疾病可具有共同的遗传背景、相关的临床表现、相似的血清学、以及相似的对治疗的反应和预后。将这些对患者预后具有明确的重要价值的疾病标记加入到入组标准中，可提高入选患者的均一性，从而可使 RA 研究能够更加有效的进行。新的流行病学和分子遗传学方法可鉴别出更多的临床亚组。但是，首先需要进行前瞻性研究确定新报告的预后因子的临床价值。如果现有数据确实可以支持某些标记可以作为预后因子时（风险因素），如类风湿因子、侵蚀性或血管炎性疾病，以及 DR4 纯合子，则可在研究设计中考虑加入这些标记，因为这些因子被认为可能影响患者对治疗的反应。虽然在某些情况下，这类研究可能使研究结果的普遍性和代表性受到限制，并可能影响最终产品的说明书声明，但这种努力对研究药物的风险/受益特性仍然很有帮助。

b. 抗风湿病的联合用药治疗

RA 患者的研究（除病情极轻微的患者外）通常采取联合用药的方式进行，联合用药的药物包括皮质激素、NSAIDs、羟氯喹等。这种联合治疗的方式给患者的选择、毒性监测和临床试验设计方面带来很多挑战。例如，甲氨蝶呤用于很多 RA 患者的治疗，因此在临床实践中新药经常不可避免地需与甲氨蝶呤联用，除非存在禁忌症。所以，除非不与甲氨蝶呤联合给药是患者可以承受的，否则需要研究药物与甲氨蝶呤联用的数据，来评估联合治疗潜在的免疫抑制作用。其他联合给药的药物也应进行类似的评价。

另外，根据患者对之前所接受的标准治疗的反应，可将患者分为不同的类型。不同类型的患者需要采用不同的试验方案来研究其对所研究的新治疗药物的反应。例如依据甲氨蝶呤的使用效果，RA 患者可分为 5 种类型：（1）非甲氨蝶呤治疗候选人——病情太轻或处于疾病过于早期阶段，无需使用甲氨蝶呤；（2）甲氨蝶呤治疗候选人——疾病的活动程度已经（或将要）达到需要使用甲氨蝶呤的程度；（3）甲氨蝶呤治疗成功者——疾病减轻至可忽略的程度；（4）甲氨蝶呤治疗失败者——由于无效或产生药物耐受，从而导致治疗失败；（5）甲氨蝶呤部分有效者——尽管使用甲氨蝶呤治疗，仍有一定程度的疾病活动。上述任一类型的患者对所研究的 RA 治疗新药的反应性，均应采用适宜的方案单独考察。如果在特定的试验中只研究了某个 RA 患者亚组人群（如对甲氨蝶呤无反应者），该结果只是严格反映了药物在该类患者中的疗效，当然这些信息对其他类型患者的疗效也可以有一定的提示作用。针对各种反应类型所进行的单项试验，可从其他类型人群试验的阳性结果

中得到支持。任何拟定的目标患者亚群必须都是临床上可鉴别的。当申办者计划开展的临床试验仅涉及 RA 患者中的特殊亚群，从而支持该亚群患者的疗效声明时，应与监管机构人员进行协商。

c. 其他联合治疗

大多数 RA 患者也正在同时服用其他药物。不太可能影响治疗结果的药物（如抗高血压药物）应简单记录，虽然研究者应注意可能存在的药物相互作用。在临床研发期间，应获得试验药物和预期的伴随药物的联合用药信息。在关节炎药物或止痛药物的处理中，可考虑下述几种方法。

禁用：该策略可能导致依从性差或脱落数量增加。

方案中加入具体的用法规定，并予监测：通过这种策略，其他合并使用的止痛药物（及其他可能的关节炎药物）可依据方案中的具体规定使用。另外对于长期研究，方案应说明（1）关节内使用激素是否被允许，以及如果允许，多长时间的被累及关节的评价不予分析；（2）因手术等需要给与应激剂量皮质激素的处理方式；（3）这种给药方案多久后允许进行评价。

将止痛药的使用，或其使用量设计成为一项疗效终点（或疗效终点的一部分）评价指标。

将其他关节炎治疗手段的使用设计成为一项疗效终点（或疗效终点的一部分）评价指标。

d. 分层

随机化用于平衡人群的混杂因素；但是，总是存在随机化未能达到平衡的可能，特别是在较小规模的试验中。根据已知的（或高度怀疑的）主要危险因素将患者分层，确保其在各组间的平衡是明智之举。当某个因素对研究结果的影响被认为有可能与治疗手段带给患者的影响程度相当时，应考虑将该因素作为分层因素（如侵蚀性病变、类风湿因子阳性）。一个常被忽视的危险因素是患者的既往治疗史。（进一步讨论参见统计部分）。

e. 设盲

由于大多数 RA 结果评价指标都具有较强的主观性，应力求在患者和评价者中保持高度的盲态保密性，以期获得可靠的结论。如果阳性药物和对照药物组之间在起效时间、患者反应特点及毒性特点并不近似平行时，则盲态可能会受到影响。试验应尽可能在所有治疗组中保持平行（如“双模拟”）给药，以使频繁的给药操作不会危及到盲态。如果必须对试验组的治疗进行特殊调整（如按照监测的药物水平），可通过第三方揭盲来完成（第三方同时被隔离），以保持患者及评价者仍处于盲态。在这种情况下，应在对照组也应模拟相应的调整，同时进行平行改变，以期保持盲态。类似的，如果盲态可能被输注相关的事件或其他治疗方案中的因素影响，那么关键治疗终点（如关节计数）应由不了解受试者病史的独立方进行评价。

f. 脱落和不依从的影响

试验中尽可能减少脱落和参与者信息丢失是非常重要的。传统上，推荐的 RA 试验设计着重于排除患者之间的差异，例如额外使用止痛药物，以及接受关节内注射，这些干预

措施常常被认为显著违背方案，要求相应的患者退出研究。在 RA 试验的设计中，患者是应该保留还是基于存在的差异使其退出研究需要进行权衡。方案要求过于严格时，可能因为患者和研究者对方案的要求无法容忍而导致患者脱落和不依从，继而导致研究结果无法解释。另一方面，方案中允许任何额外的干预措施存在又可能令人困惑，以至于同样难以解释研究结果。

下述策略可能有助于减少信息的丢失：

i. 使用筛查或清洗期的方法，以确保患者符合要求而且愿意参加试验后，才被随机分到各治疗组。

ii. 对研究者和研究人员进行全面系统的培训，减少不合适的病例入组、方案违背和其他可能降低评价试验结果能力的偏差。

iii. 在终点的定义中引入脱落情况，正如在“明确的治疗失败”或“明确的由患者引起的治疗成功或失败”中引入脱落情况一样。应选择足够长的时间点作为终点，避免两个治疗组之间对药物的反应时程存在差异，从而产生判断误差。

采用这种方法的案例之一，是采用在方案中明确定义的“缓解率”作为主要终点，而由于缺乏疗效而脱落的患者被分类为无反应者。在该终点类型下，无反应者的分类标准应明确地预先定义。当有可靠的 II 期数据，证明了在治疗开始后规定的时间点上药物的疗效以及药物反应的持续时间，则适宜使用这类评估终点。在这种情况下，研究者可以将治疗 6 个月时符合 ACR 20 缓解标准的患者比例进行对比分析，以此作为首要的分析方法。方案中应规定，如果依照方案治疗 2 个月后，在连续的两次研究访视中未观察到相对于治疗基线的改善，则患者被视为无反应者。实践将检验这种方法能否有效地限制由于脱落造成的信息丢失。

iv. 制定对已停止实验治疗的患者进行随访的规定。对于符合预先规定的“治疗失败标准”的患者，可考虑在方案中做出特殊规定，允许其进行交叉试验，重新接受标准的治疗。

v. 研究期间允许更灵活的治疗选择。某些已经被采用的方案中，允许比较治疗组进行剂量调整（评价者和患者均处于盲态），对于那些满足方案中预先定义的“疗效不充分”标准的患者，允许其使用额外的其他治疗，以及允许有限的少量患者进行关节腔注射，在评价时将该受累关节排除在外。

2. RA 试验设计

RA 临床试验可采用多种设计方法。同一个试验可支持一种以上的疗效声明，疗效声明可单独提交或同时提交。试验可被设计用来检验某项差异的存在——证明研究药物优于对照（安慰剂，较低剂量，另一种活性药物），或设计用来检验不存在差异——证明产品与具有活性的对照药物在疗效方面足够相似。安慰剂对照、剂量对照、浓度对照或活性药物对照的设计均可被使用。

一般来说，显示出存在差异的试验其说服力优于等效试验，因此一项疗效声明应至少在一个试验中充分显示出受试药物的疗效优于安慰剂或阳性对照药。如果寻求的是优于某特殊对照药的疗效声明而非药物单独的直接疗效，该疗效声明应通过两个充分且具有良好对照的、显示出优越性的试验证明。这类试验也可以为证明产品疗效奠定基础。

a. 优效试验

标准的两组设计，即研究药物组与安慰剂对照组是最常见且最直接的 RA 试验设计。试验方案的具体内容取决于受试人群。具有轻度疾病活动性的 RA 患者只服用 NSAID，而未接受过其他类型的治疗，这类患者可入选至安慰剂对照组，同时继续使用 NSAID 背景治疗；然而，NSAID 单独治疗效果差的患者通常不适于被入选安慰剂对照组。对其他不同的治疗方法显示为部分缓解或治疗失败的患者，也采取相同的考虑。

除了上述两组差异设计方案外，还可选择标准的剂量反应研究和优于阳性对照假设研究。当患者被随机分至安慰剂对照组不可行时，需给与活性药物治疗，本类设计方案可满足该种情况的需要。

b. 等效试验

等效试验被用来证明受试药物与活性对照药物充分相似。使用预先定义的等效检验方法，设置 95% 置信区间水平，使受试药物组与活性对照药物组的实际差异小于某个预先确定的量。两种药物达到相似的疗效估计值，并不能作为两者等效的证明。

等效试验可实现两个目的。首先，其能够为单纯的疗效声明提供证据。第二，能够支持特定的与药物 X 等效的声明。两个目的可在同一个试验中实现。应注意到，对于与药物 X 等效的声明，其证据强度可能需要大于单纯疗效声明所需的程度。因此，基于不同的疗效声明需要，等效试验可能存在差异。目前，适合于某特定的 RA 试验的等效标准需视具体情况而定。如上所述，如果需要获得与药物 X 等效的疗效声明证据，该试验的要求可能会更加严格。另外，如果设置安慰剂组，等效性试验的设计则有所不同，因为安慰剂可用来估计绝对的和相对的药物作用程度。

在任一情况下，等效性的统计检验方法均需要在方案中给出定量的描述。无论单独寻求疗效声明，还是寻求与药物 X 等效的疗效声明，基本上，确定适宜的统计学检验方法的基础仍是临床。方案代表了一种共识，即在特定的情况下为了支持特定的疗效声明，小到何种程度的差异可被认定为无临床显著性，从而可被认为两种治疗方法在临床上等效的。

有一种情况值得考虑，即活性药物对照组的有效率非常高。例如，由抗感染药品部门 (CDER, FDA) 评价过的某些抗生素产品。对于这些产品，等效检验标准中允许的潜在差异大小取决于标准治疗的有效率。例如，当标准治疗药物有效率 $\geq 90\%$ 时，受试新药只能被允许有 10% 的空间来证明两者之间有效率的差异不超过 10%。从技术上看，这意味着差异的 95% 置信区间只能位于 10% 的右侧。如果已知标准治疗具有 80 ~ 90% 的有效率，则可使用 15% 的范围。这些等效检验适用于无安慰剂组的试验、以及已知安慰剂有效率极低的情况。

RA 的治疗有效率常常只在 50% 左右（当然，依赖于所使用的评估终点），而安慰剂有效率约为 20%，因此在等效试验中，临床上在制定“允许的差异程度”决策时，与前述使用抗生素治疗的情况不同。此时，临床决策必须根据各个试验的具体情况而定。

在未设置安慰剂组的等效试验中，一个主要的问题是确保两种治疗同样地有效，而非均同样无效。很多被批准用于 RA 治疗的药物其疗效很有限，在与安慰剂比较时可能无法显示出疗效。当与这些药物进行等效试验时，若不使用安慰剂对照组则可能缺乏可信度。

因此，在等效试验设计中应选择疗效较高的参比药物。如果可能，推荐在等效试验中设置第三组（安慰剂组或较低剂量组）来观察治疗的差异。该第三组的患者数量或治疗持续时间不需要与对照药物组相同。如果设置安慰剂组，试验组和活性药物对照组的疗效均需要在统计学上超过安慰剂组，这样“等效性”才真正具有意义。

密切关注试验设计和实施的各个环节，对于确保等效试验结论的准确非常重要。方案设计时，关于目标患者人群、给药方法、以及有效性和安全性的评价，应采用与对照相比无偏倚的方式进行，以确保公平地比较。此外，应注意试验执行过程中的某些具体问题，如如何减少脱落、不依从性和数据丢失，这对于试验结论的可靠性也很重要。这些试验执行方面的问题可能掩盖差异，继而导致错误的等效结论。这将导致试验结论与药物的实际疗效相悖，而真实的药物疗效差异原本可以在“优效设计”的试验中体现出来，从而导致试验的失败。

在任何特定的研究中，试验设计的选择依赖于很多因素。由于通过设置对照组从而证明疗效差异的试验证据，相比较等效试验的证据更具说服力，所以通过使用试验各组间差异最大化的策略，能够获得更高的试验效率（更少的患者数量或更短的暴露时间）。理想情况下，这意味着给予所有患者安慰剂对照（安慰剂组患者需给与必要的背景治疗）。

c. RA 研究中的新试验设计方法

虽然以往并没有在 RA 新治疗药物的研究中被应用过，但“退出设计”可在某些情况下予以考虑。退出设计有时可用于评价药物的疗效。在该种设计中，试验中的两组患者均接受受试药物治疗，然后一组患者在盲态下退出治疗，接下来比较两组患者的治疗结局。若患者停用受试药物后病情恶化则证明药物有效。退出设计的自然评估终点为从“停药到病情恶化的时间（预先在方案中明确定义）”，使用标准的时间—效应统计检验或进行两试验组结果比例的单纯比较。退出研究可以在两组患者均使用背景治疗的情况下进行。

关于退出设计，有很多需要注意的地方。如果产品毒性很大，以至于最开始被入选的患者中仅有小部分（耐受人群）可以坚持到试验终点，可进入双盲状态下的退出阶段，那么任何由于患者退出治疗而得出的结论仅适用于该耐受群体。另外应注意到，如果药物诱导成瘾或耐受，退出或反弹现象可能导致退出治疗的患者不但没有因治疗获益，反而病情加重。

3. 结果分析

a. 处理脱落

以往 RA 试验一直面临着由于患者脱落而导致的信息丢失的问题。脱落通常不会不随机发生，也很少完全与受试药物的治疗无关，因此脱落通常会带来试验结果判断上的偏倚。该问题在很多随机试验中常见。一些方法已经被提出用来分析脱落带来的影响，但用这些方法来分析脱落的影响都不够充分。

通过使用意向治疗（即包括所有随机分组并接受过治疗的患者）分析方法，当出现脱落时，采用之前最后一次观测数据结转（ITT/LOCF）的方法进行补充；或将 ITT/LOCF 和 PP/OC（仅为依照方案完成者/观察到的病例）两种方法一并使用，仍无法解决脱落的问题，虽然这些方法可以增加结果的可信度。还应注意到，还有其他的方法可帮助模仿建立

缺失的数据，参见 Little 和 Rubin (1987)。使用这类重建方法需要满足一定的前提条件，但是现有的数据无法被证实满足这些条件。

在所有的试验中均需要阐述脱落带来的影响，以证明试验结论的可靠性。一种试验设计方法是随访所有患者（包括脱落患者）直至计划的试验终点，即使脱落后患者接受的新治疗会干扰随访获得的信息，然后进行包括这些患者的试验结果分析。另一种方法采用“最差情况规则”：即分配给所有脱落的安慰剂组患者可能的最好评分，而给予所有脱落的治疗组患者最差评分，然后进行包括这些患者的试验结果分析。

b. 与治疗基线比较的结果评价

在 RA 以及其他疾病中常观察到的一种现象是，留在试验中的患者比那些脱落的患者对治疗的反应更好，即响应者优于无响应者。安慰剂组和治疗组的情况均如此。如果仅从临床试验中进行疾病观察，则可能得出结论——疾病的自然病程被显著改善。但这种现象归咎于恶化患者有可能提前脱落（而该现象在 LOCF 分析中无法充分得到补偿）以及向均数回归。在“疾病突发”类方案中，这种问题更加严重，此时患者均表现出显著改善，无论治疗状态如何。该事实使得与“治疗基线进行对比”的结果评价难以进行，因为所观察到的很多病情改善经常与治疗作用无关。基于上述原因，在未设置安慰剂仅设置活性药物对照组的试验中，使用与基线进行比较的分析方法可能极难解释，特别是在使用“疾病突发”设计的情况下。在任何情况中，任何试验中的成功均意味着其对疾病的改善程度超过对照组。

4. 疗效试验设计的统计学

建议在着手开始研究前，与 FDA 审评小组讨论方案设计和结果分析的问题。另外，FDA 制定的新药申请的临床及统计部分格式、内容指南（1988）也包含了较多有用的信息。

a. 随机/分层

随机旨在将患者分配到治疗组，避免偏倚并确保合理地使用统计分析程序。

在某些临床试验中存在一些已知因素，它们在影响疾病严重程度方面至少与研究药物相当。分层设计可避免仅依赖随机法来平衡这些因素。通过分层法，可将患者由简单随机化分组变为基于所选择的分层因素被均衡分至各治疗组。

每个 II 期和 III 期研究方案中均应包括关于随机的内容。所有关于随机化过程的约束条件应明确阐述。当分层因素或样本量的要求未在方案中提及时，可断定无相应的随机化要求。这适用于患者是否在进入研究后未被均衡分配至组内接受治疗，以及适用于基于中心和基线的更明显分层。

由于分层意味着对随机的约束，因某些因素分层的研究应在统计分析部分中考虑这些因素。各个研究均应执行方案定义的分析。

可使用某些统计程序来处理治疗组间差异的偏倚，这些程序中引入了经某些因素（或者相关变量）校正的方法，这些因素包括分层因素等，研究方案中需预先明确这些因素，或者明确如何确定一定数量的相关变量。预先在方案中确定相关变量（或选择相关变量的标准）是非常重要的。另外，这些统计校正程序应谨慎使用，且需验证方法所需的前提条

件是否具备。

在决定是否在所有临床试验中均采用分层随机时，需要根据具体情况进行判断。选择分层和选择统计校正都是有理由的。

分层的首要优势是其避免了可能需要进行的基于不同治疗影响的大量统计校正。结果分析开始前，分层将从本质上排除这种校正的影响。第二，虽然分层和统计校正程序均用于去除预期治疗效应的偏倚，但分层通常更具说服力。这是因为分层可使预期治疗效应的变异程度小于没有实施分层的统计学校正。最后，在统计分析模式中加入分层因素，将有助于提高我们对药物疗效检测的能力。

随着层级数量的增加，以及继而导致的每层中可进行随机化的患者数量下降，分层变得越来越不灵活。在这种情况下，逻辑上更简单的方法就是不进行分层，而依赖统计学方法在这些因素上进行校正。

最佳方法是将分层（适用于有限数量的对预后影响最大的因素）与统计学模型相结合。这种方案中定义的统计学模型将既考虑分层，也基于余下的极少数重要因素的影响而考虑进行统计学校正。

b. 主要疗效变量的确定

各个Ⅱ期或Ⅲ期研究方案均应确定主要和次要疗效变量。主要疗效变量对于确定产品是否有效非常关键。次要疗效变量对主要变量的正确性起到支持的作用，但在决定受试产品是否有效方面，其重要性次于主要变量。统计学上能够证明受试药物在次要疗效变量方面仍然有效，这在临床上是非常有帮助的，尽管并不是必须的。

c. 统计分析方法的预先阐述

临床终点的统计分析是获得一致的且可信的产品疗效证据过程的一部分。这些统计分析方法不应根据数据来设计。在各个研究方案中，应在获得分析所需的数据前充分且详细的制定主要疗效变量统计分析方法，以便独立的统计工作者可根据方案进行统计分析。统计分析方法的简要描述应包括但不限于下述内容：（1）使用的统计学显著性水平；（2）统计学假设检验或可信区间是进行单侧还是双侧；（3）是否计划进行中间时间点的分析，如果进行，假设检验和置信区间如何根据间期数据进行调整；（4）所用统计模型的数学表达；（5）证明治疗成功所需的最小统计结果；（6）缺失值和脱落的处理；（7）用于控制多重主要疗效变量 I 类错误率的方法；以及（8）进行多重治疗比较所使用的方法。

d. 多终点

在 RA 治疗中，是否使用多终点来评价疗效的主要证据常存在临床争议。这种复合终点的理论基础是目前正在进行的统计学研究的焦点。例如，对于之前 FDA 指南（1988）推荐的 4 个指标——医生整体评价、患者整体评价、肿胀关节计数和疼痛关节计数，若在某项试验中至少 3 个指标具有统计学疗效证据，则该试验结果被认为可以支持药物有效的结论。

多变量统计方法也可用于分析一系列的主要疗效变量。用于从多终点结果中获得推论的统计程序正处于研究开发中。

疗效变量可在患者内进行合并（复合终点）。这种已经固定下来的合并的疗效指标应

在研究方案中详细定义。多终点评价体系可能存在一些统计和推理上的困难，复合疗效终点在避免这些困难方面具有一定的优势。

e. 脱落病例

脱落病例是指在试验开展一定时间后，未能按照方案如期提供临床上治疗有效数据的患者。脱落的原因常与所服用的试验药物（基于不良反应或缺乏疗效）相关。由于脱落并非随机发生，那么余下的患者，相对于试验之初原本是随机化的患者群体来讲，则构成了一个偏倚的亚样本群体。除非能够假设脱落是随机发生的（这种情况极为罕见），否则没有一种模型能够被证实可用于补偿其所带来的影响。

用于处理脱落的方法（如 LOCF 和完成试验者分析）均不完全令人满意，尽管这些方法通常作为判断临床上提供的治疗有效的证据在统计学上是否充分的基础。LOCF 方法通常不保留检验样本含量的大小，无论对于最终观察值比较还是对于变化率的比较。其他方法包括生长曲线分析和随机影响回归。这些方法也对信息缺失较为敏感，即脱落依赖于对治疗反应的评价。如果能够证明试验结果是经过各种方法分析后得到的结论，这通常是非常有用的（即结果是可靠的）。

f. 具有数个治疗组/多重比较的试验

在涉及两个以上治疗组的临床试验中，应使用多重比较统计程序，以控制实验误差率在 5% 或更低。实质上，在试图进行特定的药物间比较研究（用来支持拟定的试验药物说明书疗效声明）前，应获得主要治疗效应上的总体统计学证据，以证明受试药物具有肯定的疗效。

g. 被用来同时证明 1 个以上疗效声明的试验

单一试验可被用来同时寻求 1 个以上的疗效声明；基于多重分析的显著性水平调整并非总是必要的。如果预先定义了检验假设的顺序，则不需要采取补偿措施。例如，当试验同时寻求 6 个月的体征和症状改善声明以及 12 个月的 x 射线影像学改善声明时，如果试验在第一个假设检验（体征和症状改善）上证明是成功的，那么 x 射线影像学的假设可在无需调整补偿的情况下进行计算。

h. 间期分析

间期分析是指出于任何目的的，在累积临床试验疗效数据过程中重复进行的分析工作。多重检验（包括间期分析）可改变真实的显著性水平，目前多种方法已经被制定出来用于补偿这一事实。研究方案应说明是否计划进行这种间期分析。如果计划进行间期分析，那么分析计划及其执行过程应在方案中描述。描述应包括谁将具有访问间期数据的权限、间期分析的具体时间安排、用于调整显著性水平的方法、以及显著性水平（在该水平下，具有统计学显著性的结果将被声明）的相应时间顺序。

虽然间期分析不会被认为对后续的疗效数据采集有任何影响，但间期分析仍然会对研究的盲态或执行过程带来额外的风险。为了进行间期分析而进行的数据采集工作可导致破盲和偏倚，而目前尚无统计学方法可以补偿任何破盲及偏倚带来的后果。最后，如果方案有必要进行某些重要的改动（如出现新的治疗方法），该修订内容不应被那些了解数据的人员所影响。

I. 样本量

未能招募足够数量的患者，是导致原本有效的产品不能满足已建立的判定产品有效的统计学标准的主要原因，无论试验目的是证明治疗作用的优效还是治疗作用具有可比性。确定样本量的方法应充分、详细的描述，以便能够进行独立的计算验证。分析方案中应该包括确定的疗效变量（样本量确定的基础）、假设的临床差异程度、标准偏差、统计效力、显著性水平、以及统计程序为单尾还是双尾。此外，所选临床差异的大小应合理，并且应讨论用于确定样本量的疗效变量选择的原理。考虑到某个试验与下一个试验结果的可比性，使用与前一个试验相同的疗效变量较为适宜。

j. 证明临床等效性的试验

临床等效性，较其对于普通读者的含义来讲，目前使用的是其非常狭义的概念。首先，通常不会去证明两种或更多药物在广谱药理作用中是等效的。关注的焦点是，在预先明确的一个或可能的多个变量上，观察到无临床相关的差异。如果一种药物与另一种药物的差异足够明显，则认为两种药物在该变量上不再为等效，即等效的概念在任何一项结果分析中均具有双面性。

为了证明等效性，用于评价这些治疗作用的变量应在方案中明确规定。当在某个疗效变量上寻求临床等效的结果时，还应确定基于该疗效变量的不合理的差异程度。临床数据应证明，在 95% 可信度范围内，未超过预先明确定义的差异程度。

显示等效的试验与显示存在差异的试验相比，前者的试验结论更加令人信服。临床试验通常并不检测已知存在的治疗差异。在这种情况下，统计方法则转而貌似用来提供等效作用的证据（如对于安慰剂）。

k. 统计方法的适用性

应评价统计方法的适用性，包括检查离群值、确定是否满足分布假设（通常为正态分布），以及公共方差假设是否成立。

1. 研究基地影响

如果患者在某研究基地进行了分层及随机，那么结果分析方法中应考虑研究基地影响。可能存在研究基地—治疗交互作用，反映了各研究基地之间治疗差异的程度。当各研究基地之间入选的患者变异较大时尤为明显。在结果分析时需考虑基地—治疗的交互作用。

F. 安全性分析

传统上，评价不良事件数据以及实验室检测值的方法与用于评价疗效的方法不同。安全性评价的目的通常不是检验某项特定的假设，而是检验药物作用的模式以及发现异常或延迟的事件。使用累积发生率、实验室检测值散点图（基线与治疗比较）分析、或者其他技术可能会有帮助。药物的安全性应阐述不良事件（药物反应或实验室值）对药物暴露持续时间、剂量水平、伴随疾病或可能的药物相互作用的依赖程度。不良反应发生率应采用相同标准来计算，该标准能反映危险人群的药物暴露时间的。累积发生率（危险率，即刻可能性）在反映药物作用的暂时性特点方面优于总体发生率；累积发生率的比较表格（药物与活性药物对照、安慰剂对照比较）也对实践者非常有帮助。严重事件发生率应使用

95% 置信区间描述。

确保药物、生物制品或医疗器械的安全性，可通过充分的临床前评价以及建立标准的临床安全性评价体系来实现。成功的安全性评价体系应包括预先定义的标准术语（例如注册事务医学术语字典），定义及评价不良事件（AE）的标准、优化 AE 检测的方法、以及试验中适宜的出于安全性考量的试验终止规定。获得如下信息也将对试验非常有帮助，包括 AE 严重程度（1 级为轻度、2 级为中度、3 级为重度、4 级为致命性）、最终结果（如治疗需要及是否缓解或死亡）、医生评价的不良反应与受试药物之间的相关性（轻度的、可能的或很可能相关）、以及对于试验的影响（无影响、药物给药延迟或剂量变化、患者退出后续的研究）等。推荐制定关于试验终止的规则（根据受试人群对药物的风险/获益比来确定），该规则既应用于个体患者（通常使用单独的 3 ~4 级 AE 标准）也应用于整个临床试验，尤其对于剂量递增研究。

1. 内在试验设计

应努力阐明那些对不良药物反应较为敏感的患者人群的特点。某些外部因素可使安全性数据复杂化，例如不同研究者在收集和报告不良事件上的差别，以及不同实验室间实验室检测值的正常值范围定义差异等。由于校正这些因素带来的影响可能较为困难，因此应在试验设计阶段采取预防措施，通过制定明确详细的数据采集指导方法，并监督研究者和实验室严格按照方案的要求来执行，从而将这些因素的影响降至最低。例如，可采取某些措施将实验室数据标准化。如前所述，制定适用于所有研究中的标准毒性分级量表也很有帮助。

2. 足够的患者数量

检测不良用药经验的能力依赖于临床试验中以及临床使用中评价的患者数量。在 300 名或更多患者充分暴露于研究药物的研究中，预期（95% 可信度）至少 1 名患者出现发生率为 1% 或更高的不良事件。更小规模的研究不能满足该最小发生率的检测标准。然而在大多数情况，可合并具有相等周期的研究来确定不良事件发生率。

对于任何长期给药的产品，安全性数据库应包括至少 300 名患者接受最大推荐剂量治疗至少 6 个月，以及至少 100 名患者接受治疗至少 12 个月（ICH E1A 1995）。对于已知或可能存在安全性问题的药物，建议建立更大样本量和/或更长周期的安全性数据库。

IV. 生物制品的特殊问题

尽管药品与生物制品在 RA 试验设计上有许多类似之处，但是生物制品在研发过程中有其特有的特点和问题。

A. 种属特异性

传统的用于确定人类起始使用剂量的方案可能不适用于生物制品。生物制品在动物模型和人体内的作用方式存在差异，差异的程度取决于动物和人的生理相关性，以及配体与受体的亲和性在人和动物之间的差异大小。另外，免疫原性可能也存在种属差异。

B. 剂量效应

剂量—效应曲线可能很陡，甚至可能是双曲线，而且药物水平在刚超过预期起效水平

的时候就可能有很强的毒性。

C. 毒性反应

毒性反应可能非常难以预计，而且可能会非常危险，甚至有可能造成疾病恶化。药物的治疗窗可能很窄。生物制品还可能会扰乱免疫学和生理过程。例如针对免疫系统细胞表位的单克隆抗体，或者如 TNF 受体的单抗，使用剂量若稍微高于有效剂量，就会造成严重的并发症，而在有效剂量下使用的毒性则明显减小。

D. 产品均一性

产品的均一性通常在产品活性和毒性方面起关键作用。产品的改变可能显著影响生理学活性。如此，在研究阶段的生物制品必须最大限度地证明批与批之间的一致性，并且已进行了充分的正确的评估。

E. 抗体的作用

如果 II 期临床试验数据显示，药物诱导产生的中和抗体随着时间的推移可以影响到生物制品的有效性，那么就需要正式的研究这种可能性，在随机一对照设计中。如果有中和抗体的产生，那么药物剂量或者给药方案需要重新考虑。非中和抗体也可能对药代动力学产生深远的影响，因而也有可能与中和抗体同样重要。

V. 医疗器械的特殊问题

A. 背景

用于治疗 RA 的医疗器械其治疗作用的差异是非常大的，从作为某个特定受累关节的一线治疗方法，到作为一线治疗（例如超声波或热疗与药物联合使用来缓解症状）的辅助措施，。由于疾病和反应的异质性而造成的治疗效果的差异可能在某些器械治疗过程中被放大，因为这些器械的治疗常仅针对单个病人的一个或少数几个关节。

医疗器械在临床前研究中，多项检测数据被要求进行评估，例如，用于评估某器械化学成分、机械可靠性或电学性质等。每款产品都经过一系列独特的检测，用以评估并确保特定产品的设计被理解透彻，且该产品会给患者带来与说明书声明内容一致的治疗效果。

RA 医疗器械的临床研究方案设计时，可能会遇到在药物试验中不经常遇到的特殊问题。其中的一个例子是，治疗和未治疗的关节之间的相互作用可能会影响功能和生活质量的评估。当被研究的有效治疗方法是植入物，而可供选择的对照的活性更多是全身作用效果时，如何选择适当的对常常会遇到困难。理解如何设计这些研究、分析数据和评估每项治疗的作用可能是有难度的。

B. 有效性

1. 对于医疗器械的研究，选择对照组可能是相当困难的，因为要使病人和研究者处于盲态常常是不可能的，而且设置安慰剂的可行性也较差。由于以往所选择的对常常不令人满意，评估局部治疗和具有全身治疗活性的对照之间的等效性本身就存在问题，所以通常需要研究出新的设计方案。一些新的设计方案包括随机将病人分为早期使用此器械和晚期使用此器械组，或者将此器械作为药物治疗失败之后的挽救措施。尽管对评价治疗 RA

的大多数产品来说，使用假的（安慰剂）器械，是最合适的对照，但是如果器械需要植入的话，就不适合了。另外，使用假器械，常常不能完全对病人和/或医生隐瞒。如果器械需要通过手术或其它有创操作才能使用的话，也无法使病人处于盲态。但是，在某些情况下，在评估产品有效性时，能够做到对评估者设盲。

2. 拟用于 RA 辅助治疗的器械，在设计和分析时需要考虑到患者之间疾病状态和严重程度差异，从而将结果评价时潜在的偏倚减少到最低。其中一个解决策略是对所有病人都使用同一种药物来避免混杂效应。在研究设计阶段，应当明确阐明所使用的额外的可能有混杂作用的辅助治疗（热/冷治疗、夹板、理疗和矫正术）。

3. 如果器械用于康复治疗的目的，那么生活质量（QOL）的评估就非常重要。评估干预措施对生活质量的改善，应当基于其治疗的准确性、舒适性和方便性，以及改善生活质量的同时对产品也满意。评估过程中应当由独立的评估者来评估产品的有效性，并且对评估者应当设盲，以避免评估时的偏倚。

4. 如果侵入性器械需要在医院或者诊所内使用，那么建议临床评估时包括使用的方便性和使用时的疼痛和不舒适感。早期对这方面进行评估可以为最终是否能够耐受提供重要信息，并且对是否继续该治疗的决定产生影响。

C. 安全性

1. 医疗器械的临床试验，通常没有必要像大型的药物临床试验一样进行药物早期安全性评估以获得短期不良事件发生率的详细资料。器械的试验常常规模很小。仅依靠早期数据不足以支撑对该医疗器械后续效应的预测。如果研究的器械是植入型的，评估的时间需要足够长，以评估迟发型不良事件发生的可能性，包括器械使用的失败和与器械材料相关的不良事件。在某些情况下，在上市前无法完成安全性的全面评估，需延至上市后进行安全性监测。

2. 如果器械是和其他药物或者手术治疗联合使用，那么将很难明确区分某项不良事件是与器械相关，还是常见于药物治疗和手术过程。不良事件的特征、持续时间和严重程度等因素可以帮助确定该不良事件的原因。由于这些结论都是基于临床判断，那么对此治疗方法下结论的时候就会存在潜在的偏倚。在制定方案时要充分考虑到这一点，方案中应该给研究者详细的指导，从而使研究者得出尽可能一致的和无偏倚的判断决定。

3. 器械（例如，放射性器械）如果存在间断使用或者长期使用的可能，需要评估是否存在总暴露时间的上限值或者最大的暴露频率。这些问题常常通过动物模型数据和长期临床评估来阐明。

IV. 幼年类风湿关节炎的特殊问题

A. 背景

幼年类风湿关节炎（Juvenile rheumatoid arthritis, JRA）是一类异质性疾病，它们的共同特点是慢性、原发性滑膜炎，在 16 岁之前发病。该疾病根据关节受累的范围和关节外的表现分为不同的临床亚型：单关节、多关节和系统性 JRA，以及与 HLA—B27 相关的单关节炎⁴，这些亚型根据临床病程再进一步细分（C 进一步 assidy 1986）为不同的类型。遗传

免疫亚型似乎与这些临床病程亚型相关，并且与成人 RA 也不相同（Nepom 1991）。这些不同的疾病亚型中，多关节型 JRA 在许多方面与成人 RA 类似，特别是在临床症状和体征、滑膜炎方面，另外多关节型 JRA 对一些已知的治疗 RA 药物（NSAIDs、甲氨蝶呤和强的松）也存在类似的治疗反应。由于仅有 3—5% 的类风湿关节炎的病人在儿童时期发病，应用于这一小部分人群的许多处于研究阶段的治疗药物，根据 21CFR 第 316 部分——孤儿药，将按照“孤儿药”来对待。成人 RA 临床试验所采用的原则，大体上也适用于 JRA，本章仅列出那些与成人 RA 的区别之处。对于即将被批准用于成人 RA 治疗的药物，通常鼓励申办者也提供尽可能多的用于治疗 JRA 的信息。至少应提供给药剂量和安全性的数据。

在儿童中进行药物研究通常是必须的，而且是与该病治疗方案的预期是相一致的。由于儿童受试者属于脆弱人群，研究过程中将风险减到最小非常重要。美国儿科协会药物委员会发布了儿童中药物研究的伦理实施指南（AAP 1995a），以及婴儿和儿童中临床药物评估的总体考虑（AAP 1982），这两个方案都要参照执行。同时也要依照美国儿科协会生物伦理委员会关于知情同意和儿科病人同意的指南进行（AAP 1995b）。对 JRA 病人进行临床研究以及，尤其是评估全身的疾病活动状态和对治疗反应的研究，需要已经接受广泛深入的儿科风湿病学培训，且被证明具备处理风湿病患者能力的儿科风湿免疫学家和成人风湿免疫学家共同参与。

作为一项一般性原则，未在成人中验证过安全性的药物不得在儿童中进行研究。但是，当依据现有的知识，预期的收益要大于预期的风险时，试验可以从儿童中开始。专门研发用于治疗 JRA 的药物（例如，一个专门针对 JRA 特有的而成人 RA 不具有的病理学过程而设计的生物制品）可能需要首先在儿童中进行检测，因为在成人 RA 患者或者甚至健康成人志愿者中的研究可能无法获得准确结果。但如果受试药物可能既用于成人 RA 也用于 JRA，那么在开始儿童试验之前，至少要获得成人 I 期临床 PK—PD（包括 MTD）的数据。对于某些药物的 JRA 临床试验，如果预期的临床疗效与现有药物效果类似，或者与基本治疗措施相比没有显著优势也不能作为替代治疗药物，那么试验应当推迟到在成人或其他儿童人群中已获得足够的有效性和安全性数据之后进行。

为了获得可靠的试验结论，不一定需要安慰剂对照，但是仍然需要随机和对照设计。根据对所研究药物的了解程度和可供受试者选择的其它治疗方法，选择对照组。如果在等效性试验中只设置了活性药物对照组，那么该活性药物的有效性需要有充分的证据支持，并且该等效性试验方法需要详细阐述。如果之前没有在成人患者上进行研究，或者研究药物有新的作用机制或者代表了一类全新的药物，那么需要设置安慰剂对照，或者设置与受试药物具有类似预期疗效的活性对照药物组，进行随机双盲试验。扩充说明书内容，增加安全性和疗效持续性的数据是非常有价值的。在 JRA 研究中可以用如下方式替代无活性的安慰剂对照，可以设置活性药物对照（标准治疗），采用剂量—效应设计（对照组给予较低剂量的受试药物），交叉、随机退出（增显设计），或者如果药物效果短暂，可以使用随机安慰剂阶段研究设计（Temple 1994, Feldman 1995）。作为一般性原则，建议使用退出机制来保证对受试药物反应不佳的儿童尽早接受传统治疗或者其他治疗。申办者需要指明在结果分析时如何处理脱落病例，无论是出于退出机制还是其他原因。

B. 儿科用药管理规定的实施及其对 JRA 研究设计的影响

说明书标签管理规定（21 CFR 201.57 (f) (9)）中关于“儿童用药”的章节指出，药品和生物制品可以在药品说明书和标签中标明供儿科使用，如果已在成人中被证实了有效性和安全性，并且药物在成人体内和儿童体内的作用机制是类似的。该儿科用药管理规定仅适用于在说明书里声明该药物可以改善 JRA 症状和体征；其它方面的疗效包括预防器质性损害、缓解和防止残疾等，应当在单独的 JRA 有效性研究中评估。尽管该管理规定允许成人的有效性数据可以外推至儿童，但通常仍然需要进行额外的儿科用药剂量探索 and 安全性评估。

一般而言，若要将成人 RA 产品申请用于 JRA 患者，通常会鼓励申办者在多关节 JRA 患者身上获取给药剂量和安全性数据，列入说明书及标签中的“用法用量”和“儿童使用”部分。用于 PK 研究所采集的标本数量可以大大减少，如果已有数据显示儿童和成人中系数是相似的。这些研究中应采用小样本量采样技术（Hsahimoto 1991）。安全性检测的范围依赖于药物之前使用的情况和在其它儿童人群中揭示的安全性情况。毒性分级量表也需要根据儿科人群进行调整。如果在产品批准上市前的安全性数据有限，那么强烈建议进行 IV 期临床研究来进一步评估安全性。在评估儿童用药剂量和安全性的过程中，需要收集尽可能多的药物有效的证据。

对于目前已批准的传统 NSAIDs（环氧合酶抑制剂）和皮质类固醇药物，已经有足够的疗效数据来支持说明书中将所有 JRA 和 JRA 亚型列入“适应症”中。对于甲氨蝶呤和柳氮磺胺吡啶，已有充分的疗效数据来支持将多关节受累的 JRA 患者列入说明书的“适应症”中。对于这类药物，仅需要在 JRA 患者中的 PK、PD 和安全性数据即可支持说明书中的疗效声明，虽然也鼓励提交补充的 JRA 患者有效性数据。

对于新药（未被批准用于成人 RA 治疗），且该药不属于药理学分类上的新类别时，成人的有效性数据可以支持其用于多关节型 JRA 患者的症状和体征控制，当有生物学证据提示该药物对 JRA 患者可能有类似的作用时。儿科用药管理规定对于支持说明书中的多关节型 JRA 适应症的适用性，是基于每个药物在成人 RA 中的有效性数据。如果没有生物学理论上的证据，那么需提供证据支持适用于儿科用药管理规定。（需要咨询相关部门来决定是否有充足的生物学理论依据来申请使用儿科用药管理规定。）需要提交关于儿科用药安全性和剂量研究的资料。

如果药物属于新的药理学分类，需要在 JRA 中进行有效性研究，以获得 JRA 患者使用的指导。适应症将反映出在有效性研究中涉及到的 JRA 亚型。申办者如果寻求批准用于所有的 JRA 患者，那么就要在有效性研究中把所有 JRA 亚型都包括进去。假如研究数据没有显示此产品对某个 JRA 亚型无效，那么该数据可以支持此药适用于 JRA（亚型可以不指定）。说明书中应反映出此产品的有效性已被证实，且已被批准用于 JRA（指定的亚型是根据在有效性研究中包括的亚型而确定）。

C. 结果变量和结论声明

申办者可以申请批准用于所有的 JRA 亚型或其中的某一种亚型。在之前的案例中，说明书中需要记录试验中每个亚型入选的病例数以及每个亚型的反应特点。除了如前所述申

请使用儿科用药管理规定的情况外，其他所有疗效声明均需得到目标亚型患者的有效性证明。

1. JRA 症状和体征的改善

所有的 JRA 试验都需要根据临床终点（经确证可用来评估病情改善程度的终点）来评估病情改善的情况。目前，公认的有效方法之一是采用依据 JRA 核心指标系统建立起来的对病情改善的定义：6 种评价指标（医生总体评价，家长/患者总体评价，活动关节数目，活动受限的关节数目，关节功能和 ESR）中的 3 种改善至少达到 30%，并且不超过 1 种加重程度超过 30%（Giannini 1997）。由于对 JRA 患者“病情改善的定义”是通过甲氨蝶呤临床试验来验证的，该试验主要入选多关节受累的 JRA 患者，所以对于其它亚型的 JRA 患者，在反应测试方面需要改进，方案需要进行个体化设计。例如，对于只有一侧膝关节受累和正常 ESR 的单关节 JRA 患者，评估受累关节及其功能，以及评估眼葡萄膜炎的情况作为共同的主要终点，可能也是非常有价值的（Lindsley 1996）。如果病人有系统性 JRA，额外评估发热、关节外表现和血小板增多/白细胞增多症也是非常有价值的终点（Silverman 1994）。结果变量要适当并与研究药物种类相一致。研究者在试验开始前，需要确定对每个结果变量，多大程度的改变量具有临床意义。

试验一般需要进行至少六个月，除非已有成人 RA 的六个月的有效性数据，并且没有理由推测有效性会随时间而降低。在这种情况下，研究周期可以是 3 个月的盲法/随机试验，但是仍需要获得 6 个月开放性的安全性数据。与成人 RA 一样，NSAIDs 类药物建议试验周期为 3 个月。

2. 主要临床缓解

与成人 RA 类似，“主要临床缓解”是指受试药物可带来切实的临床获益的疗效声明，包括对治疗不能完全应答的病人和疾病缓解的病人。目前此疗效声明只是理论上的，因为目前还没有能够充分支持“主要临床缓解”的临床 JRA 试验数据库。

3. 完全临床缓解

“完全临床缓解”的疗效声明包括服药后连续 6 个月晨僵时间少于 15 分钟，没有活动性滑膜炎（痛、红、压痛、肿胀、僵直或活动范围受限），没有关节外症状（包括发热、浆膜炎、淋巴结肿大、肝脾大、皮疹和葡萄膜炎），正常的实验室检测指标（包括 ESR、血小板和白细胞）以及，如果可行的话，在治疗过程中没有器质性损害的进展。需要进行至少 1 年的试验。之前疾病的残留损害，包括关节外表现，不影响对完全临床缓解标准的判断。由于 JRA 中自发的完全临床缓解发生率相对较高，所以这类研究需要设置对照。

4. 缓解

缓解的定义如上所述，但是需在所有抗风湿药物停用之后。缓解不代表治愈。

5. 预防残疾

此疗效声明反映的是在为期 1—2 年的研究中，药物显示出持续的对躯体功能以及残疾情况的改善作用，与此同时症状和体征也得到明显改善。目前被证实适用于 JRA 试验评估的工具包括儿童健康评估问卷（CHAQ）、幼年关节炎自我报告指数（JASI）和幼年关节炎

功能评估报告 (JAFAR)。HR—QOL 同样也需要评估, 并且在研究期间不能加重。评估终点应根据研究中入选的不同患者亚型来调整 (例如, 在单关节 JRA 患者中需评估膝关节的功能, 因为对这些患者而言膝关节炎是最主要的关节表现)。上述这些评价工具需要在试验中进一步验证其对不同年龄患者的适用性。

6. 预防器质性损害

与成人 RA 类似, 此项疗效声明反映的是持续超过 1 年的试验, 在此试验期间伴随有症状和体征的明显改善。目前, 只有零散的资料涉及到仅一个 JRA 影像学指标: 腕关节炎患者的腕掌关节间隙。其它有希望的临床指标包括评估病变侵蚀情况 (在伴有多关节炎的系统性 JRA 患者中), 系统性 JRA 患者髌关节受累情况, 以及单关节 JRA 患者的膝关节受累情况。

D. 试验设计问题

对药物有效性研究的建议是基于受试药物的特点。成人 RA 的原则基本上也适用于 JRA。这类临床研究中入选的病例其发病时间可能不同, 或属于不同的亚型。如果根据药物的作用机制推测受试药物可能不适用于所有的 JRA 亚型或对不同亚型患者的治疗效果可能不同, 则建议针对每个 JRA 亚型单独开展研究。另一种方法是开展单一的大样本研究, 分层入选病人, 以充分评估药物对每个亚型的有效性和安全性。相关的协变量包括疾病类型、发病持续时间以及对之前甲氨蝶呤的治疗无反应。由于 JRA 是罕见病, 试验设计时可以有些灵活性, 但是也需要具体案例具体讨论。

目前, JRA 病人常不适于入选到有效性研究中, 除非他们对至少一种标准的二线药物 (例如甲氨蝶呤, 剂量每周 10 mg/m² 体表面积) 的治疗是失败的。对此也有例外, 例如有证据显示在疾病较早期使用受试药物可以获得更好的疗效, 在晚期病人中使用发生毒性反应的潜在风险更大, 或者根据目前研究情况, 在相似人群中有良好的安全性和有效性, 并且受试药物的作用是很容易逆转的。单关节 JRA 病人更适合被入选到针对葡萄膜炎治疗的药物研究中去, 或者入选到可以替代目前治疗方法的安全性更好, 需要血象监测更少, 和/或更有效的药物研究中。

一些接受新药治疗的 JRA 患者需要在更长的时间内定期监测。监测期间需要关注骨骼生长、发育、行为、性成熟、生殖能力和继发恶性肿瘤等情况。登记或队列随访研究对提供长期的安全性信息是非常有帮助的。

E. 合并使用的抗风湿药物

在 JRA 试验中合并使用抗风湿药治疗的原则与成人 RA 类似: 尽可能合理地减少合并用药的随意性, 从而避免对有效性和安全性的评价带来影响。但是, 这种对合并用药的限制不能违背伦理学的要求, 也不能设计一个对家长、病人和医生没有吸引力的方案, 继而影响到患者的入组。

如果病人同时接受慢作用药物或者强的松治疗, 剂量需要和研究开始前一致, 并且在实验过程中保持一致。合并使用的药物通常对预后有很大的影响, 所以需要分层设计。如果可能, 尽量不要入选在试验开始前 1 个月内接受过关节内类固醇注射治疗的患者; 否则, 该相应关节在治疗效果评价时不得计入。

F. 多中心试验和中心效应

尽管 JRA 是儿童最常见的风湿免疫病，但与成人 RA 相比，患病率较低。因此，需要大量病人的 JRA 试验需要在多个研究中心来完成。多个研究中心的试验要求采用统一的标准化方案和数据采集表。在试验之前所有研究者和其他参与人员的会议是非常有必要的，这样可以保证对方案解释、病人评估和数据录入的一致性。研究显示，在多个中心的协作过程中，每个中心的表现是基于其入选病人的数量 (Sylvester 1981)。所以，研究中心少但每个中心入选的患者数量多，相对比研究中心多但每个中心入选患者少的研究要更好。每个中心至少需要入选 10 到 12 名病人，来提供更好的质量控制。在所有的多中心试验中，要检验中心效应。治疗效果应当在一个以上的中心显示出来。当严格的入选标准限制了病人入组，每个中心甚至无法做到入选 10 个病人。在这种情况下，每个中心中采取随机化封锁，而不是在所有的中心中采取这样的措施，可能会有助于减少中心效应的潜在影响。

参考文献

American Academy of Pediatrics (AAP), Committee on Drugs, 1982, General Considerations for the Clinical Evaluation of Drugs in Infants and Children, Report to the U. S. Food and Drug Administration, available through the Freedom of Information Office of the FDA, FDA Contract Nos. 223—79—3003 & 223—82—3009.

_____, Committee on Drugs, 1995a, “Guidelines for the Ethical Conduct of Studies to Evaluate Drugs in Pediatric Populations,” *Pediatrics*, 95: 286—294.

_____, Committee on Bioethics, 1995b, “Informed Consent, Parental Permission, and Assent in Pediatric Practice,” *Pediatrics*, 95: 314—317.

Cassidy, J. T. , J. E. Levinson, J. C. Bass, et al. , 1986, “A Study of Classification Criteria for a Diagnosis of Juvenile Rheumatoid Arthritis,” *Arthritis and Rheumatism*, 29: 274—281.

Feldman, B. M. and J. P. , Szalai, 1995, “Towards a More Ethical Clinical Trial Design: The Randomized Placebo Phase Design (RPPD),” abstract, *Arthritis Rheum*, 38 (supplement): s177.

Felson, D. T. , et al. , 1993, “The American College of Rheumatology Preliminary Core Set of Disease Activity Measures for Rheumatoid Arthritis Clinical Trials,” *Arthritis and Rheumatism*, 36 (6): 729—740.

Felson, D. T. , et al. , 1995, “American College of Rheumatology Preliminary Definition of Improvement in Rheumatoid Arthritis,” *Arthritis and Rheumatism*, 38 (40): 1—9.

Food and Drug Administration, 1978, General Considerations for the Clinical Evaluation of Drugs, FDA 77—3040.

_____, 1988, Guidelines for the Clinical Evaluation of Anti—Inflammatory and Antirheumatic Drugs (Adults and Children) . April 1988.

_____, 1997, *Guidance for Industry: Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro*, April 1987.

_____, 1993, "Guideline for the Study and Evaluation of Gender Differences in the Clinical Evaluation of Drugs," *Federal Register*, July 22, 1993, 58 (139): 39406—39416.

Fries, J. F. , P. W. Spitz, and D. Y. Young, 1982, "The Dimensions of Health Outcomes: the Health Assessment Questionnaire, Disability, and Pain Scales," *Journal of Rheumatology*, 9: 789—793.

Giannini, E. H. , N. Ruperto, A. Ravelli, et al. , 1997, "Preliminary Definition of Improvement in Juvenile Arthritis," *Arthritis and Rheumatism*, 40: 1202—1209.

Hare, L. , L. Wagner—Weider, A. Poznanski, et al. , 1993, "Effects of Methotrexate on Radiologic Progression in Juvenile Rheumatoid Arthritis," *Arthritis and Rheumatism*, 36: 1370—1374.

Hashimoto, Y. , and L. B. Sheiner, 1991, "Designs for Population Pharmacodynamics: Value of Pharmacokinetic Data and Population Analysis," *Journal of Pharmacokinetics and Biopharmaceuticals*, June 1991, 19 (3): 333—53, NLM CIT. ID: 91341632.

ICH, "S5A Guideline on Detection of Toxicity to Reproduction for Medicinal Products," *Federal Register*, Thursday, September 22, 1994, 59 (183): 48746—48752.

_____, "S6 Preclinical Safety Evaluation of Biotechnology—Derived Pharmaceuticals," *Federal Register*, April 4, 1997, 62 (65): 16437—16442.

_____, "EIA Extent of Population Exposure Required to Assess Clinical Safety for Drugs Intended for Long—Term Treatment of Non—Life—Threatening Conditions," *Federal Register*, March 1, 1995, 60 (40): 11269—11271.

Larsen, A. , K. Dale, and M. Eek, 1977, "Radiographic Evaluation of Rheumatoid Arthritis and Related Conditions by Standard Reference Films," *Acta Radiol. Diag.* 18: 481—491.

Lindsley, C. , 1996, "Juvenile Rheumatoid Arthritis Workshop," *Proceedings*, pp. 34—43.

Little, R. , and D. Rubin, 1987, *Statistical Analysis with Missing Data*, John Wiley & Sons.

Meenan, R. F. , P. M. Gertman, J. H. Mason, et al. , 1982, "The Arthritis Impact Measurement Scales: Further Investigations of Health Status Measure," *Arthritis and Rheumatology*, 25: 1048—1053.

Murray, K. J. , and M. H. Passo, 1995, "Functional Measures in Children with Rheumatic Diseases," *Pediatric Clinics of North America*, 42: 1127—1154.

Nepom, B. , 1991, "The Immunogenetics of Juvenile Rheumatoid Arthritis," *Rheumatic Disease Clinics of North America*, 17: 825—842.

Paulus, H. E. , M. J. Egger, J. R. Ward, et al. , 1990, "Analysis of Improvement in Individual Rheumatoid Arthritis Patients Treated with Disease—Modifying Antirheumatic Drugs Based on Findings in Patients Treated with Placebo," *Arthritis and Rheumatism*, 33: 477—484.

Pinals, R. S. , et al. , 1981, "Preliminary Criteria for Clinical Remission in Rheumatoid Ar-

thritis,” *Arthritis and Rheumatism*, 24 (10): 1308—1315.

Sharp, J. T. , D. Y. Young, G. B. Bluhm, et al. , 1985, “How Many Joints in the Hands and Wrists Should Be Included in a Score of Radiologic Abnormalities Used to Assess Rheumatoid Arthritis?” *Arthritis and Rheumatism*, 28: 1326—1335.

Silverman, E. , G. D. Cawkwell, et al. , 1994, “IVIG in the Treatment of Systemic Onset JRA,” *The Journal of Rheumatology*, 21 (12): 2353—2358.

Smolen, J. S. , F. C. Breedveld, G. Eberl, et al. , 1995, “Validity and Reliability of the Twenty—Eight—Joint Count for the Assessment of Rheumatoid Arthritis Activity,” *Arthritis and Rheumatism*, 38: 38—43.

Sylvester, R. J. , H. M. Pinedo, M. De Pauw, et al. , 1981, “Quality of Institutional Participation in Multi Center Clinical Trials,” *New England Journal of Medicine*, 305: 852—855.

Temple, R. J. , 1994, “Special Study Designs: Early Escape, Enrichment, Studies in Non—Responders.” *Comm. Statistic. Theory Meth.* vol 23 (2): 499—531.

附录 A：回应率比较试验^①

风湿性疾病（CSSRD）的三个协作系统性研究：

- (1) 甲氨喋呤 vs. 安慰剂
- (2) 金, 金诺芬 vs. 安慰剂
- (3) D—青霉胺 高, 低 vs. 安慰剂

临床试验终点时有效率（基于对“改善”的不同定义）

	安慰剂	(金诺芬) (低剂量 D—青霉胺)	高剂量 D—青霉胺 金制剂, 甲氨喋呤
改善的定义	n = 199	n = 18	n = 155
ACR > 20%	10 (8.4%)	30 (25.4%)	64 (40.3%)
ACR > 30%	5 (4.2%)	14 (12.0%)	46 (29.7%)
ACR > 40%	2 (1.7%)	7 (3.4%)	18 (11.6%)
ACR > 50%	0 (0%)	4 (3.4%)	14 (9.0%)
ACR > 60%	0 (0%)	3 (2.5%)	4 (2.6%)
ACR > 70%	0 (0%)	0 (0%)	1 (0.6%)

附录 A（续前）

金诺芬与甲氨喋呤对比的多中心临床试验（试验终点）

临床试验终点时有效率（基于对“改善”的不同定义）

改善的定义	金诺芬 (n = 118)	甲氨喋呤 (n = 119)
ACR > 20%	34 (28.8%)	77 (64.7%)
ACR > 30%	30 (25.4%)	65 (54.6%)
ACR > 40%	22 (18.6%)	51 (42.9%)
ACR > 50%	21 (17.8%)	42 (35.3%)
ACR > 60%	9 (7.6%)	22 (18.5%)
ACR > 70%	7 (5.9%)	11 (9.2%)

^① CSSRD database and Tugwell et al., “Combination Therapy with Cyclosporine and Methotrexate in Severe Rheumatoid Arthritis,” N Engl J Med 333; 137—141, 1995

附录 A (续前)

环孢素 A 与甲氨碟呤联合使用与甲氨碟呤单独使用的对比试验
 临床试验终点时有效率 (基于对“改善”的不同定义)

基于 ACR 标准的 提高率	患者满意标准	
	甲氨碟呤 + 环孢素 A (n = 71)	甲氨碟呤 + 安慰剂 (n = 74)
0%	81.7	50.0
10%	49.3	16.2
20%	45.0	12.2
30%	33.8	8.1
40%	22.5	2.7
50%	22.5	2.7
60%	5.6	2.7
70%	1.4	0.0
80%	0.0	0.0
90%	0.0	0.0