**医疗器械关键临床研究的设计考虑因素**

**行业、临床研究者、机构审查委员会和美国食品药品管理局工作人员指南**

**文件发布日期：2013年11月7日**

**本文件草案发布日期：2011年8月15日**

对于有关本文件与CDRH监管器械相关的问题，请致电(301) 796-5750 或发邮件给以下电子邮件地址 greg.campbell@fda.hhs.gov（如必要）联系Gregory Campbell博士。

对于有关本文件与CBER监管器械相关的问题，请致电301-827-6210联系Stephen Ripley。

|  |  |
| --- | --- |
|  | **美国健康与人类服务部****美国食品药品管理局****器械与放射卫生中心****生物制品评价与研究中心** |

**前言**

**公众意见**

如有问题，可以在任何时候联系美国食品药品管理局，文件管理部（地址：5630 Fishers Lane, Room 1061, (HFA-305)， Rockville, MD, 20852）提交书面意见和建议供机构考虑。

**其他副本**

可从互联网获得其他副本。您也可以向dsmica@fda.hhs.gov 发送电子邮件请求获得本指南的电子副本或发送传真至301-847-8419请求获得硬拷贝副本文件。请使用文号1776确定您所请求的指南文件。也可以从以下机构获得本文件的其他副本：

**生物制品评价与研究中心**(CBER)

对外信息、外联与发展办公室 (HFM-40)

地址：1401 Rockville Pike, Suite 200N, Rockville, MD 20852-1448

或致电1-800-835-4709 或301-827-1800， 或发送电子邮件ocod@fda.hhs.gov，或从以下网址

http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformatio n/Guidances/default.htm获得其他副本。

**目录**

**1 引言 4**

**2 使用范围 5**

*2.1 本指南涵盖的研究类型 5*

*2.2 本指南未涵盖的研究类型 6*

**3 证据等级和研究设计的监管概要 7**

*3.1 PMA批准的法规标准：安全性和有效性的合理保证 8*

*3.2 有效的科学证据 9*

*3.3 风险-受益评估 10*

*3.4 临床研究证据等级和监管要求*  *11*

*3.5 简明概念和研究设计原则 12*

**4 医疗器械类型 13**

*4.1 基于预期用途的医疗器械类型 14*

*4.2 器械临床研究需要考虑的特殊因素 15*

**5 关键研究设计中探索性研究的重要性 16**

**6 临床研究设计选择的一些原则 17**

*6.1 研究类型 18*

*6.2 一般考虑：器械性能的偏倚和变化 19*

*6.3 研究目的 19*

*6.4 受试者选择 20*

*6.6 试验机构选择 22*

*6.7 比较研究设计 23*

**7 临床有效性研究 24**

*7.1 临床研究终点 24*

*7.2 临床有效性研究的介入分配（随机） 26*

*7.3 盲法 28*

*7.4 对比临床有效性研究的对照品 29*

*7.5 安慰剂疗效和其他现象 31*

*7.6 非对比临床有效性研究 32*

*7.7 诊断临床有效性研究 34*

*7.8 某些临床有效性研究的优势和不足 35*

*7.9 一些监管考虑因素 38*

**8 诊断临床性能研究 38**

*8.1 预期用途考虑因素 39*

*8.2 目标条件的临床参考标准 40*

*8.3 诊断性能评价的研究人群 41*

*8.4 研究计划，受试者悬着和样本采集 42*

*8.5 诊断临床性能对比研究 43*

*8.6 诊断性能研究中的设盲 44*

*8.7 与器械交互人员的技能和行为（从测试概念） 44*

*8.8 诊断临床性能研究中的常见偏倚类型 45*

**9 保持临床研究的质量 46**

*9.1 临床数据处理 46*

*9.2 研究实施 47*

*9.3 研究分析 49*

*9.4 关键研究的预期变化 50*

**10 临床研究计划或方案 50**

**11 术语表 51**

**医疗器械关键临床研究的设计考虑因素**

**行业、临床研究者、机构审查委员会和美国食品药品管理工作人员指南**

|  |
| --- |
| ***本指南代表了美国食品药品管理局（FDA）目前对这一主题的看法。本指南并非产生或授予任何人权利，其作用也并非约束FDA或公众****。****如果方法符合适用法规和条例的要求，您可以使用替代方法。如果您想讨论替代方法，请联系负责实施本指南的FDA工作人员。如果您无法确定适当的FDA工作人员，请致电本指南标题页上列出的相应电话号码。*** |

**1 引言**

本文件旨在为支持医疗器械上市前提交的临床研究参与者和审查注册提交的FDA工作人员提供指南。虽然本机构整合了支持特定器械类型研究设计相关的政策，并调整满足监管要求的一般政策，本机构未进行描述可能适用于支持器械上市前提交的不同的临床研究设计，或规定申办者应该如何决定哪些关键临床研究设计应该用于支持特定器械的注册提交。本指南文件描述了可能用于满足上市前临床数据要求的医疗器械临床研究开展相关的不同研究设计原则。本指南并不是为研究医疗器械提供最佳临床和统计实践的综合教程。

医疗器械可能经历三个普遍的临床开发阶段。这些阶段可能是相互依赖，且在某个阶段进行全面的评估可以使下一阶段更加简单。首先，医疗器械可能经历一个探索性临床研究阶段。在这个阶段，会对医疗器械的局限性和优势进行评价。这一阶段包括首次人类使用研究和可行性研究。下一阶段关键阶段是用来获取评价器械对于已确定预期用途安全性和有效性的必要信息。通常包括一项或多项关键研究。最后，器械需要经过上市后阶段，其中可能包括一项或多项额外研究，以便更好地了解器械安全性，如罕见的不良事件和长期有效性。本指南探讨了与关键临床研究相关的设计问题，但没有详细探讨其他阶段。

医疗器械的关键研究是一项决定性研究，在该研究中采集支持医疗器械对于其预期用途的安全性和有效性评价的证据。根据一项或多项关键临床研究获得的证据通常作为确定医疗器械上市前批准申请(PMA)安全性和有效性的合理保证和 FDA 给出总体风险-受益决策的主要依据。在某些情况下，PMA 可能包括多项旨在回答不同科学问题的研究。

本指南的重点就如何设计临床研究支持PMA向申办者提供建议。然而，进行临床研究支持市场前公告(510(k))和新注册提交的申办者也可以采用本指南文件中所述原则。

FDA的指南文件，包括本指南，不具有法律强制性责任。相反地，这些指南描述了FDA关于某个主题的当前想法，并应仅被视为建议，除非引用了具体的条例或法规要求。FDA指南中使用的“应该”这个词时，表示建议或推荐的但不是必需的要求。

**2 适用范围**

本指南描述了上市前临床研究[[1]](#footnote-0) 的设计原则，这些研究在确定医疗器械的安全性和有效性方面很关键。

在本指南中，讨论了关键临床研究设计的实际问题和容易犯的错误，以及它们对可能从有关安全性和有效性的研究中得出的结论的影响。本指南中所讨论原则旨在帮助申办者的营销申请和帮助研究者设计足以提供有关器械安全性和有效性合理保证的研究。

***2.1 本指南涵盖的研究类型***

由于预期使用的范围和风险与医疗器械相关的风险以及执行临床研究的约束，本指南将按照一般方式处理关键临床研究。本指南概括了FDA有关两大类医疗器械的建议：

* 治疗和整容器械
* 诊断器械

根据本指南，器械开发人员可以了解每一个类别的器械的重要关键研究设计问题。与此同时，与FDA审查人员沟通（如通过提交前交流）在达成实用和合适的关键临床研究设计方面通常很有价值。

本指南还包括了适用于21 CFR第3部分定义的组合产品的器械特定问题的原则（如药械组合产品；器械-生物制品产品）。然而，对于本指南中未介绍的组合产品，特定药物或生物学特定问题也可能是相关的。

本指南旨在补充其他现有指南，并不是为了取代其他指南文件中描述的各项政策。在存在问题的情况下，根据负责器械审查的中心不同，请直接咨询相应的FDA审查部门或器械与放射卫生中心(CDRH)小型制造商，国际和消费者援助中心或生物制品评价与研究中心(CBER)对外信息、外联与发展办公室(OCOD)。

***2.2 本指南未涵盖的研究类型***

尽管本指南未探讨以下研究类型，本指南讨论的一些原则适用于其中许多研究类型：

* 非临床研究 （如实验室研究、动物研究或测量研究，以及对于体外诊断器械进行的分析验证研究）；
* 旨指支持人道主义器械豁免(HDE) 申请的研究；[[2]](#footnote-1)
* 上市前的可行性临床研究，或不属于关键阶段的其他上市前临床研究；
* 用于确立同类诊断器械临床有效性的研究（如提供使用相应治疗产品的安全性和有效性基本信息的体外诊断测试）。同类诊断器械的临床开发项目，通常是相应治疗产品临床开发项目的一部分；
* 上市后临床研究。虽然上市后临床研究的要求可能源基于上市前临床结果，上市后研究不能初步确定产品的安全性和有效性，且本指南并未探讨其设计问题。但是，本指南中讨论的原则在设计此类研究中可能是有用的；
* 受CBER 监管产品的研究可能需要试验性新药注册申请和生物制剂许可证注册申请，因此本指南的范围不包括捐赠者筛查。

尽管这份指南主要是为了用于支持PMA的临床研究制定的，本指南的建议也可以用于支持某些510(k)和新注册提交（在适用的情况下可能包括临床资料）临床研究的设计。

**3 证据等级和研究设计的监管概要**

医疗器械的临床研究必须符合一定的法规要求。本节将介绍以下内容：

* 适用于支持医疗器械上市前注册提交的临床研究设计的监管设计；
* PMA批准的法规标准；
* 适用于满足PMA批准的法规标准的临床和其他数据的监管要求；
* FDA如何评价数据以评估器械的风险和受益；
* 关于研究器械豁免(IDE)申请基本信息；以及
* FDA有关简明定义确定的良好法规实践的当前意见。

本指南反映了机构有关涉及人类受试者参与研究的设计、实施和记录、报告的标准。有关国际文件包括“赫尔辛基宣言”，并在*国际标准化组织(ISO)14155:2011，医疗器械人类受试者临床研究 - 良好临床实践和通过国际协调会议(ICH)*人类使用药品注册技术要求(ICH)*E6良好临床实践：统一指南*中进行了进一步解释。21 CFR第50、54、56和812部分下的FDA法规清晰地阐明了适用于医疗器械临床研究的药品临床试验管理规范(GCP)要求。此外，FDA指南文件介绍了FDA当前有关GCP 和临床研究实施的意见。[[3]](#footnote-2)

符合GCP（适用于医疗器械）能保护人类受试者的权利，安全和福祉，确保合理科学地实施临床研究以及结果的可信度，规定了申办者和临床研究者的职责，并协助申办者、研究者、机构审查委员会(IRB)，以及其他伦理委员会，监管当局和参与医疗器械开发、评审的其他机构。

如果器械的临床研究是在美国地区以内进行的，则该临床研究必须符合21 CFR 第56部分(IRB)，第50部分（知情同意），第812部分（研究器械豁免(IDE)）中适用的法规。

如果器械的临床研究是在美国以外进行的，提交旨在为了支持PMA，按照IDE进行的，则该研究应当符合21 CFR 第812部分。21 CFR 814.15(a)。如果器械临床研究完全是在美国以外地区进行，且不是按照IDE进行的，则根据21 CFR 814.15，如果数据是有效的且研究者是按照“赫尔辛基宣言”或研究进行国家的法律法规进行的，则FDA将接受为支持PMA而提交的研究，无论是否给人类受试者提供了更多的保护。如果外国数据适用于美国人群和医疗实践，如果研究是由具有被公认能力的临床研究者进行的，如果需要进行检查，FDA可以通过现场检查或其他适当的途径来验证数据，则可能批准完全基于国外临床数据的PMA，否则申请应符合批准的标准21 CFR 814.15(d)。根据适用标准，申办者必须能够证明外国临床数据足以支持器械在美国的批准。联邦食品、药品和化妆品法案（FD&C法案或法案）第569B部分。如果您预期寻求根据国外数据，或根据外国和美国数据组合获得批准，我们希望您在预提交会议中会见FDA工作人员。

***3.1 PMA批准的法规标准：安全性和有效性的合理保证***

正如FD&C法案第513(a)(1)(C)部分所示，PMA必须提供器械安全性和有效性的合理保证。FD&C法案第513(a)(2)部分：

*[T]器械的安全性和有效性待确定*

*(A)对于器械预期代表或适用的人员，*

*(B)对于器械标签中规定，推荐或建议的使用条件，以及*

*(C)考虑器械使用的任何可能健康受益相对于这种使用带来的任何可能的损伤或疾病的风险。*

此外，FDA已经通过法规按照如下方式解释了PMA批准的法规标准的解释：

*21 CFR 860.7(d)(1)。当根据有效的科学证据可以确定，对于其预定用途和使用条件，采用足够多的不安全使用说明和警告时，器械使用带来的可能健康受益超过任何可能的风险时，应有合理保证确保器械是安全的。用于确定器械安全性的有效科学证据应充分证明不存在与对于器械预期用途和使用条件的器械使用相关的不合理疾病或损伤风险。*

*21 CFR 860.7(e)(1)。应合理保证器械的有效性，当根据有效的科学证据可以确定，在很大一部分目标人群中，对于其预定用途和使用条件，采用足够多的不安全使用说明和警告时，将提供临床上显著的结果。*

这些法律和法规规定，安全性和有效性的合理保证的结果必须得到与目标人群相关的数据支持，并根据器械标签进行评价。此外，是否满足PMA批准的标准的确定是根据平衡可能的健康受益和可能的风险做出的。

***3.2 有效的科学证据***

法规指出，器械的安全性和有效性应根据有效的科学证据确定。21 CFR 860.7(c)(1)有效的科学证据是通过如下方式确定的：

*21 CFR 860.7(c)(2)的有效科学证据是根据严谨对照研究，部分对照研究，没有匹配对照的研究和客观试验，由合格的专家保留的记录病历，以及已上市器械的重要人类经验报告得到的证据，根据这些证据，可以由合格的专家公平地和负责任地得出对于其使用条件有合理的保证确保器械的安全性和有效性。根据器械的特征，其使用条件，警告和其他限制条件的存在和充分性，及其使用经验的程度，所需的证据可能会有所不同。孤立的病例报告，随机试验经验，缺乏足够细节以便进行科学评价的报告，以及未经证实的观点不被视为是证明安全性或有效性的有效科学证据。然而，在确定安全性和有效性是有问题的器械时可以考虑这些信息。*

*FDA法规同时考虑了哪种类型的证据支持安全性和有效性的合理保证：*

*21 CFR 860.7(d)(2)在确定有合理的保证确保器械安全而可能需要的证据类型中（如适用），包括采用实验动物进行的研究，涉及人类受试者的研究，和包括体外研究在内的非临床研究。*

*21 CFR 860.7(e)(2)用于确定器械有效性的科学证据应主要包括严谨对照研究，如(21 CFR 860.7)(f)所规定的，除非(FDA)授权根据(FDA)已经确定足以确定器械有效性的证据的其他有效的科学证据，即使进行没有严格研究。当(21 CFR 860.7)(f)所述的严谨对照研究不能合理地适用该器械，(FDA)可能会提出这种决定。*

因此，21 CFR 860.7中明显的一个关键原则是医疗器械的有效性证据通常必须根据严谨对照的研究获得（如21 CFR 860.7(f)所述）。然而，这些法规为FDA在确定被认为能证明医疗器械安全性的有效科学证据的证据类型方面提供了一定的灵活性。

FDA认为，在大多数情况下，为了证明原始PMA中审查器械的有效性，临床资料是必须的。本指南的第6、7和8部分提供一些指导原则，以帮助申办者确定适当的研究设计。研究结果必须为FDA 提供足够的证据确定合理的安全性和有效性保证，如以上法规所规定的。根据注册提交前与申办者的讨论，FDA可能决定替代研究设计可以产生机构可以就安全性和有效性做出决定的数据。

即使是计划良好的设计，该研究可能不能产生预期结果或证明安全性和有效性必须的结果。申办者可能需要重新评估其相对医疗器械的目标并进行更多的研究，以获得证明安全性和有效性所必须的证据。

***3.3 风险-受益评估***

在上市前评审过程中，在作受益-风险决定时FDA会考虑很多因素。FDA会考虑器械预期用途范围内的各种因素，包括目标人群。[[4]](#footnote-3)

根据FD&C法案第513条(a)，FDA会通过在其他相关因素中“权衡器械使用带来的任何可能的健康受益和这种使用带来的任何可能的伤害或疾病风险” 确定PMA申请是否能提供 “安全性和有效性合理保证”。21 CFR 860.7(b)(3)指出，在确定器械的安全性和有效性时，FDA 必须权衡“权衡器械使用带来的任何可能的健康受益和这种使用带来的任何可能的伤害或疾病风险。”

FDA会根据有效的科学证据来确定是否有合理保证确保器械是安全和有效的。21 CFR 860.7(c)(1)。当根据有效的科学证据可以确定可能的受益大于任何可能的风险时，则有合理的保证确保器械是安全的。21 CFR 860.7(d)(1)。当根据有效的科学证据可以确定对于器械的预期使用人群将提供临床显著的结果时，则有合理的保证确保器械是安全的。21 CFR 860.7(e)(1)。

在权衡医疗器械可能的受益和风险时，FDA认为要考虑很多因素。FDA发布了指南“行业与美国食品药品管理局工作人员指南；医疗器械上市前审批和新分类中做出受益-风险决定时要考虑的因素(2012)”中声称FDA在上市前的审批流程中做出受益-风险分析决定时会考虑的因素。[[5]](#footnote-4)

对数据进行彻底评审之后，FDA有关器械可能的受益和风险以及其他因素的评价目的在于确保有合理的安全性和有效性保证。在考虑可能的受益程度时FDA会考虑的因素包括：

* 受益类型
* 受益程度
* 患者获得一种或多种受益的概率，及
* 影响的持续时间

FDA在考虑可能的风险或伤害程度时可能考虑的因素包括：

* 与器械使用相关的损害程度、类型、数量和发生率
* 损害事件可能性
* 损害事件的持续时间
* 诊断器械的假阳性或假阴性的结果风险

在临床试验设计和实施过程中应该考虑这些因素，以确保在研究过程中采集相应的信息，以便在上市申请时做出这些受益-风险决定。鼓励制造商受益-风险框架中列出其临床研究方案。

***3.4 临床研究证据等级和监管要求***

根据21 CFR第812部分法规描述，在批准IDE申请时，在启动临床研究之前需要提出申请。申办者或申办者IRB（请求或不请求FDA）必须首先确定研究器械是否是重大风险器械或非重大风险，尽管最终FDA是这个决定的最终仲裁者。参见21 CFR 812.2(b)。同样，如果研究涉及重大风险器械，申办者必须向FDA提交一份IDE以便在这项研究开始之前获得批准。同样，如果研究涉及非重大风险器械，则当申办者在提出IRB审查时简要说明为什么该器械不是重大风险器械后获得IRB研究批准后，该研究被认为已经获得了IDE，并保留这类批准文件和符合812.2(b).*Id*下的缩略IDE法规。同样，在这些情况下，除非FDA通知申办者根据812.20(a)申请必须获得批准，否则不需要向FDA提出IDE。

在任何情况下，临床研究所需的科学严谨性和需要采集的证据的可靠性不依赖于是否需要获得IDE以启动该项研究。此外，严谨性和可靠性不应受临床研究分类为重大风险器械，非重大风险器械，或豁免IDE监管器械的影响。FDA鼓励申办者就相应的研究设计向FDA寻求指导，以支持潜在的上市前注册提交，以便通过方案敲定前的预提交流程就方案的关键内容与FDA进行沟通。即使是对于不需要IDE批准的研究，FDA鼓励在预提交中提交方案草案以便在研究开始之前获得FDA的反馈。

***3.5 简明概念和研究设计原则***

在考虑合适的临床研究设计时，FDA也会遵循证据负担必须符合相应的监管和科学要求的原则。这一原则反映在两个适用于PMA和510(k)注册提交的临床和非临床数据要求的法定条款中。以下两项条款被称为“简明条款。”

法规第513 (a)(3)(D) 条的相关部分指出：

*(ii)伦理秘书书面规定的，用于证明器械有效性合理保证的任何临床数据（包括一项或多项完全的对照研究）应由伦理秘书决定做出，即这些数据确立器械有效性所必需的。伦理秘书应当考虑，同时与申请人协商，评价器械有效性的简明合适评估方法应能获得批准的合理可能性。*

 *(iii)出于条款(ii) 之目的，“必要”一词是指伦理秘书做出申请能提供器械有效性合理保证的最低要求信息。*

 *(iv)本款之任何规定均不得改变评价器械申请上市前批准的标准。*

同样，法案第513(i)(1)(D)条规定：

1. *每当伦理部长要求提供信息证明技术特征不同的器械是实质等同的，伦理部长应只请求提出实质等同决定所必须的信息。在做出这样的要求时，伦理部长应考虑证明实质等同的方法并相应请求信息。*

*(ii)出于条款(i)的目的，“必要”一词是指支持做出新器械和实质等同器械之间实质等同所需的最低要求信息。*

*(iii)本款之任何规定均不会改变确定实质等同的标准。*

FDA已经发布了指南，在 “1997年美国食品药品管理局现代化法案的简明规定：概念和原则，FDA和行业最终指南” （简明指南） (2002)中解释了他们打算如何应用简明规定。该简明指南解释称，“最小负担”指相关上市前问题解决的最有效方法，该问题涉及对行业部分和FDA的最佳时间，精力和资源投入。该指南规定简明规定不影响器械的法定上市前审查标准，该标准是一个由美国食品药品管理局安全和创新法案（2012）以修正案的方式增加至简明法规的原则。[[6]](#footnote-5)

该指南进一步规定，出于临床研究设计之目的，当与替代对照相关的潜在偏倚可以最小化时，FDA和行业应该考虑随机临床研究的替代方法。本指南中讨论的研究设计原则与简明指南中讨论的原则是一致的，并通过讨论可能影响满足上市前批准或许可标准所必须的证据的因素对它们进行了扩展。

***3.6 研究器械豁免的批准***

本指南中所讨论的原则旨在帮助申办者进行上市申请和帮助研究者设计足以提供证明器械安全性和有效性合理保证数据的研究。遵循这些原则和修饰关键临床研究的设计以便符合这些器械的上市前审查标准可以通过避免FDA评审专家提出其他信息，数据再次分析，或者进行全新关键研究请求而简化审批流程。然而，FDA将不会反对IDE获批，因为研究不足以支持上市申请。FD&C法案第520(g)(4)(C)部分，根据FDASIA 第601部分修订认为FDA不会反对(IDE)获批，因为：

(i) 该研究可能不支持器械的实质等同或新分类或批准；

(ii) 该研究可能无法满足与器械批准或许可有关的要求，包括数据要求；或

(iii) 可能有必要进行额外的或不同研究以便支持器械的批准或者许可。

因此，IDE的批准并不一定表明FDA认为这项研究足以支持上市申请的批准。

本指南的重点是整组临床研究设计原则将有助于提供器械安全性和有效性的合理保证，而不是IDE批准的要求。

**4 医疗器械的类型**

本文件适用于基于预期用途的两大类医疗器械：(1)治疗和整容器械和(2)诊断器械。器械是否用于治疗，整容或诊断取决于器械标签的适应症声明。在本节中将描述这些类型的器械，

以及在设计关键临床研究时应该考虑的每种类型的器械独有的特征。本指南不包括每种环境中的每种类型的器械。

***4.1 基于预期用途的医疗器械类型***

**治疗和整容器械**

治疗器械通常用于治疗特定的病情或疾病。整容器械是通过身体结构的物理改变提供所需的视觉外观。

**诊断器械**

在本指南中，诊断器械大致描述为提供单独或与其他信息一同使用，用于帮助评估受试者健康状况，或目标状态的器械。目标状态可以受试者过去、现在或未来的健康状况，一种特定的疾病，疾病阶段或任何其他可识别状态，或可能促使采取注入启动，改变或终止治疗的临床措施。对于本指南之目的，诊断器械包括器械预期用于采集、准备和检查取自人体的样本（体外诊断(IVD)器械），诊断成像系统（如数字乳房x光检查），体内诊断器械（非成像）[[7]](#footnote-6)，提供解剖结构测量（如骨质密度，大脑体积，视网膜厚度）的器械，以及提供受试者功能测量的器械（如心脏射血分数，受试者反应时间）和组合受试者数据以产生特定输出的算法（如分类、评分或索引）。需要注意的坏死，虽然有些人能理解术语诊断是指评估疾病的存在与否，例如，诊断测试包括用于孕检，评估特定疾病的免疫力，提供关于疾病进展信息，提供关于致病物质信息，或提供将用于匹配患者和治疗产品的基因信息，以及可以通过自动化各种功能，如病理系统染色功能帮助结果读取者的器械。

**具有一种以上用途的器械**

虽然许多器械可以简单地归类为治疗，美容或诊断器械，有些器械可以属于多个类别，例如：诊断病情且同时为该病情提供治疗的器械，且该病情是通过该器械确定存在的。也有些器械可能归类为其中一个类别，但具有多个用途类，例如：在两个截然不同的患者人群中治疗两种截然不同疾病的治疗器械，或提出最初诊断且同时监测该病情进展的诊断器械。这两种情况下都可能会导致需要进行一项以上的临床研究且可能需要进行更多支持性研究，例如：实验室研究和分析研究。

***4.2 器械临床研究需要考虑的特殊因素***

在设计器械的临床研究时，应考虑医疗器械某些独特的考虑因素。虽然器械类型可能影响研究设计决策，这些考虑适用于所有器械，它们可以是治疗器械，整容器械或诊断器械。医疗器械的以下独特的特点和功能可能会影响FDA如何评价该器械，且会影响在临床研究设计中应该如何解决：

* 器械工作原理：在所有器械中，了解器械功能内在的科学原理和相关的作用机制在评估所提出的研究设计性能和充分性方面可能是相关的。作为预提交的一部分，这个信息尤其重要的，其中申办者要求FDA就制定他们的临床研究提出建议。
* 用户技术水平和培训：有些器械需要进行大量培训和技能才能安全有效使用该器械。这显然将适用于要求用户训练有素的外科专家的植入式器械，特别是当所涉及的手术很复杂时。有时，器械的正确使用需要多个人员和多种技能。例如，对于IVD，具备某些技能的人员可能采集样本，而另一个具有不同技能的人可能处理样本，而另一个具有第三种技能的人员可能会解释测试结果。在设计器械研究时，应该考虑安全有效使用器械所必须的技能。研究人员和人员的技能组合应该能反映在上市批准后在预期用途环境下可能使用器械的人员的技能范围。针对器械正确使用，向研究者和人员提供的培训应能指导器械上市后给用户提供的培训。如果就上市器械进行培训，则研究人员不应该专门接受器械使用方面的培训，以确保这项研究反映了预期使用条件。
* 学习曲线：有些器械是如此之新，以至于器械的使用存在一定的学习曲线。对于各种新技术，在临床研究中使用各种器械之前，需要时间去掌握各种使用步骤。对于植入物，这通常包括植入手术的特定手术技术。对于某些器械，学习曲线的确定可以在探索性阶段包括试点研究中解决。如果申办者在上市前研究中能为器械操作者提供实习培训，则预计将在上市后使用过程中提供学习培训。学习曲线陡峭的器械可能不适合某些环境（如家用器械），因为在这种环境下它们可能是不安全且是无效的。当在关键的研究中学习曲线明显时，重要的是要考虑在这段学习时间内采集信息如何应用于方案（如通过在研究中清晰地定义哪些受试者是学习曲线时间段的一部分）以及将在统计分析计划(SAP)中如何报告这些结果。如果学习曲线陡峭，这可能影响岁标签和用于培训需求产生一定的后果。
* 要考虑的人为因素：在医疗器械开发过程中，人为因素可以起到至关重要的作用[[8]](#footnote-7)。在器械开发的任何过程中，与器械使用相关的人为因素的研究可能会使得有必要对器械设计或使用说明书进行变更以便使其更安全更，更有效，或让受试者或医学专业人事更容易使用。集成软件或提供用户界面的器械应设计为最大程度地减少用户错误。需要更多人工介入或对用户部分做出主观判断的器械可能需要更多用户技能。当前社会情况下的器械-用户界面的清晰文件记录应该是人为因素评估的一部分。

**5 关键研究设计中探索性研究的重要性**

在开发过程中，医疗器械经常需要进行设计改进，生命周期的改进是从早期研究开始的，延伸至临床使用和批准或许可产品的初期销售，并继续延伸至延续批准或许可的商业器械版本。

对于各种新的医疗器械，以及已上市器械的重大变更，在许多情况下，临床发展以以下三个阶段为标记：探索阶段（首次人类使用，可行性），关键阶段（确定器械的安全性和有效性），和上市后阶段（设计改进，更好地了解器械的安全性和有效性以及开发新的预期用途）。虽然可以区分这些阶段，重要的是要指出，器械开发可以是一个持续，迭代的过程，因为当获得了新的信息和开发了新的预期用途时，需要进行额外的探索研究和关键研究。在开发后期获得的见解（如根据关键研究获得的）可能导致需要进行额外的研究，包括临床和非临床研究。

本节关注的是探索工作（非临床和临床研究）对于制定关键研究设计计划的重要性。非临床测试（如实验室，尸体、建模、或动物）能让我们更好的理解作用机制且可以为可能给受试者带来风险的这些器械提供基本安全信息。临床器械开发的探索阶段，包括首次人类使用和可行性研究，旨在对器械设计进行任何迭代，以便促使了解器械是如何工作的及其安全性，并设定关键研究阶段。

探索性阶段全面和完整的器械评估能更好地了解器械及其预期性能。这种理解可以帮助确认器械的用途将符合申办者的预期。还可以帮助选择合适的关键研究设计。合理的探索阶段也应该尽可能地让器械进入将在关键试验和商业市场中使用的形式。[[9]](#footnote-8)

这会减少因意外结果而需要修改关键研究的可能性。这是一个重要的考虑因素，因为改变正在进行的关键研究可能会增加成本，时间，和患者资源。这也可能使研究无效或导致将其放弃。一般来说，为了做出科学有效的证实推论，在探索阶段没有进行先进的规划的情况下，探索阶段可行性研究不应关键研究数据组合在一起。

对于诊断器械，确立等性能特征如分析特异性，精密（重复性/再现性），和检测限的器械分析验证常常是探索阶段的一部分。此外，对于这些器械，可以采用探索阶段开发算法，确定临床决策的阈值，或开发用于关键临床研究的器械版本。对于体内和体外诊断器械，根据早期临床研究得到的结果可能会促使器械改进，因此有必要在人体或利用人类样本进行额外的小型研究。在这些研究开始之前，应该咨询FDA。

即使是开始了临床器械开发的关键阶段，可能继续进行探索性研究。例如，FDA可能要求在植入后6个月，2年和3年继续进行植入器械动物试验。尽管可能允许在获得数据后6个月开始关键研究，还可能需要采集其他数据。作为另外一个实例，如果预期用于儿童，可能需要进行额外的动物实验。对体外诊断器械，在进行关键研究的过程中（或之后），不太可能继续进行器械稳定实验（如货架寿命）。

尽管关键阶段通常是一个明确阶段，在此期间会采集有效的科学证据支持评价医疗器械对于其预期用途的主要安全性和有效性，应该采用探索阶段完成器械最终设计和关键阶段的相应终点。这样做是为了确保研究器械就像21 CFR 860.7(f)(2)中所描述的那样是标准化的，其中指出：

*为了确保研究结果的可靠性，严谨对照研究使用组成或设计和性能标准化的测试器械。*

**6 临床研究设计选择的一些原则**

FDA会作为上市申请的一部分提交的医疗器械关键临床研究评审，以确定它们是否能提供器械安全性和有效性的合理保证。FDA认识到可能有几个研究类型能满足这种预期。因此，FDA鼓励申请者在研究开始之前与相应的FDA审查部门开会讨论研究设计选择以便证明器械安全性和有效性的合理保证。

在本文件中，将区分两大类型的临床研究：临床有效性研究和诊断临床性能研究。以下讨论是基于需要通过采用有临床意义和统计研究的终点研究回答的相应问题的选择。

这一节将探讨一些适用于所有医疗器械关键临床研究的考虑因素。在设计任何医疗器械临床研究，包括偏倚，可变性，和有效性的一般考虑因素，以及与研究目的，受试者选择、层级、试验机构选择和比较研究设计相关的特定考虑因素。下文定义和讨论其中每个因素。

***6.1 研究类型***

**临床有效性研究**

在临床结果研究中，受试者被分配到介入治疗组，然后在计划的时间间隔采用经确认的评估工具评估临床结果参数（或它们的经确认的替代指标），以确定介入的安全性和有效性。第7部分更详细地描述这些研究。可能的情况是，对器械临床表现（如器械在临床环境是否具有预期有效性）也进行研究，但研究的重点是一个或多个临床结果。对于本文的目的，术语“介入”是指使用研究器械或对照器械。研究器械可以是治疗器械或整容器械。诊断器械，术语“介入”与基于诊断器械得到结果的受试者管理策略相关。当目标是评价器械结果如何改变受试者的医疗护理提供者后续疗程或管理的影响时，采用临床结果研究评价诊断器械。

**诊断临床性能研究**

对于大多数诊断器械，关键临床评价不是临床结果研究，而是诊断临床性能研究。可能的情况是，性能研究也可能有临床结果，但这些结果并不是研究的重点。第8部分中更详细地描述了这些研究。在诊断临床性能研究中，诊断测试结果是从受试者获得的，但并不用于受试者管理。相反，测试的诊断临床性能是通过性能衡量指标表征的，这些指标对每位受试者的诊断器械输出结果与临床参考标准的一致性如何进行量化，其中临床参考标准用于评估受试者的目标条件，如第8部分更详细的描述。

**一般考虑**

同时具有诊断和治疗功能的器械，例如：用于检测病情，然后用于进行治疗的器械，可能采用诊断临床性能研究和/或临床研究结果（包括诊断性能元素）进行评估。

在本文件中分别讨论了临床有效性研究和诊断临床性能研究。有关临床有效性研究的更多信息，请参阅第7部分，而有关诊断器械临床性能研究的更多信息，请参阅第8部分。对于同时具有诊断和治疗或整容组成产品，请阅读第7部分和第8部分。第9部分提供了有关支撑临床研究证据水平的信息和技术，同时适用于临床有效性研究和诊断临床性能研究。

***6.2 一般考虑：器械性能的偏倚和差异***

设计研究采集正确的数据比只是为了采集更多的数据而单纯设计研究更重要。研究设计应考虑偏倚和可变性。在评价一项研究设计来确定设计是否支持PMA的批准，一个重要的考虑因素是统计偏倚的概念。偏倚是真理系统误差的简称。可能在受试者选择，研究设计、研究实施，和数据分析过程中引入偏倚。在临床研究中，偏倚可能会导致对安全性和有效性做出不正确的决定。引入很少或没有偏倚的研究设计优于不控制偏倚的设计，可能由于几个原因在临床研究中引入偏倚。下文通过可以帮助消除或减少设计阶段偏倚的策略对其中一些进行了回顾（同时参见第7部分至第9部分）。

偏倚可能扭曲研究结果的解释。当器械的性能很好时，中等偏倚的存在可能不会扭曲研究支持有关总体有效性结论的能力；

当性能是已知（或被认为是）处于边际范围时，在一些研究设计中性能可能会被偏倚掩盖。特别是当在探索阶段没有进行充分研究时且对器械有效性可能不是很好理解，可能很难选择合适的研究设计，在该研究设计下，器械的有效性将不会被偏倚掩盖。考虑可能的研究偏倚是设计研究以减少偏倚可能使最终研究结果无效偏倚风险的一个关键因素。

在评价研究设计的证据等级时的第二个一般考虑是抽样的差异，这是受研究的样本量控制的。一方面，较大的样本容量能提供了更多的数据，这样的化，性能估计的抽样变异性小，因此变得更加精确。另一方面，更大的样本量还可能导致需要对临床上无意义的结果进行分析，证明在统计学上是有意义的。应该设计研究证明临床和统计学意义。同样重要的是要注意，增加样本量不一定能解决偏倚问题，不合适的结果评估，未能显示临床意义结果的边际改善，或其他研究设计问题。

***6.3 研究目的***

研究目的提供了为什么进行研究的科学依据。这些目标应该提供支持器械的预期用途，包括任何期望的标签声明。

可以通过正式的假设检验或通过相应置信区间的点估计，在统计学上支持这些声明。对于设计用于测试科学假设的关键研究，研究目的应包括零假设和对应于任何预期声明的替代假设的声明。对于估计目标的研究（如某些诊断性能研究），而不是假设测试，可以通过点估计和描述器械性能的置信区间支持声明。

***6.4 受试者选择***

21 CFR 860.7(f)(1)(ii)指出，研究计划或方案必须包括：

*选择受试者的方法：*

*（a）提供足够的保证确保受试者适合研究之目的，提供带治疗或诊断病情的诊断标准，提供经确认的实验室测试（如适用），对于预防疾病或病情的情况，提供需要预防病情易感性和保良局的证据；*

根据特定的入组标准和确认实验室或其他测试，选择进行任何临床研究受试者应充分反映器械的目标人群（如器械适用人群）。如果研究入组了那些并不能代表目标人群的受试者，则研究结果有可能存在受试者选择偏倚。

确保临床研究的受试者反映预期目标人群，方案应该明确规定匹配预期目标人群关键特征的合格标准。进行试验时，申办者应该确保只入组那些满足这些标准的个体。根据纳入/排除标准对这些患者入组到研究中。

在考虑目标人群时，FDA鼓励申办者入组那些能反映受影响人群年龄、性别、种族和民族的人口统计学。[[10]](#footnote-9)，[[11]](#footnote-10) 一些人群段参与不足会导致对于重要的人群亚组的器械安全性和有效性相关的信息不足。我们建议纳入器械适应疾病类型的发病率，诊断和治疗模式的背景讨论。这项讨论应包括：包括目标适应症既往试验中性别和种族特定患病率；并讨论解决任何经确定或提示的因素的计划，这或许可以解释可能的女性和少数民族代表（如适用）。

我们建议在方案中包括这一信息的总结和研究者培训材料。应该考虑研究机构的入组，其中可能更容易促进研究所需人群的招募。在研究设计中应包括在患者人群(21 CFR 812.25(c))和使用外国数据(21 CFR 814.15(d)(1))方面的描述，以及各自如何适用于美国人群和美国医疗实践的考虑。

当临床研究涉及弱势群体，如儿童，囚犯，孕妇、身体残疾或精神残疾人，或者经济上或教育上的弱势群体时，申办者应该提前准备与FDA讨论可能的问题，以便能遵守21 CFR 56.111(b)和21 CFR 第50部分。

在研究之前可能有事先已知的信息可以改善研究的实施，提高成功的机会。在规划研究时很重要的要考虑可能与结果相关的因素，如用户/外科医生的技能，受试者的疾病严重程度，性别或年龄。对于充分代表对当前研究尤为重要的所有重要亚组，如、性别、年龄、种族和人群时，需要谨慎行事。当感兴趣的情况很少见时，诊断研究的受试者选择可能是很有挑战性的且可以考虑替代方法。

方案可能包括多个可能的受试者选择方法：随机选择，连续选择、系统选择、和/或方便选择。

* 在**随机选择**中，所有受试者被选则参与临床研究的机会是已知的（通常是相同的）。当可实施时，受试者随机选择进入临床研究有可能提供所选择人群的无偏估计。由于逻辑方面的困难，随机选择往往是不现实的。
* 临床研究通常使用**连续选择**（即按顺序选择机构就诊的每一位受试者）或在那些满足入组/排除标准的患者中进行系统选择（如每10位受试者选择一位）。只要研究期间不与可能影响结果的，与受试者相关的其他变量混淆，这些选择方法可能会提供无偏估计。例如，如果研究持续一个早上，则目标受试者可能不能代表目标让那群，因为早上就诊的受试者可能不能代表就诊的所有受试者。
* **便捷选择**是一种受试者因为方便研究者选择而被选择的方法。这种方法可以提供的是不能推广到目标人群的结果。

***6.5 受试者选择分层***

当研究在多个机构入组受试者，有必要从能充分代表目标人群的机构选择受试者。有时，不能通过简单地选择有代表性的机构做到这一点。需要在预计会存在性能或使用差异的重要亚组中对器械性能进行充分表征。

例如，对于适用于男性和女性使用的器械不应该入组大多数男性，适用于所有成年人的器械应该完全代表所有成年年龄组。

受试者选择有两大类型技术：分层选择，和只是根据入组/排除标准选择。

分层是将目标人群划分为预先确定的非重叠受试者亚组或层级。分层选择受试者是指从每个亚组（层级）分别选择受试者。例如，您可以按照性别（男、女）和年龄组（低于或高于给定年龄）对受试者人群进行分层，从而得到四个层级，每个层级均通过性别和年龄独特组合定义。当受试者入组进入研究时，记录这些特征，而不是介入的结果。分层受试者选择不仅能确保足够代表重要的亚组，还可以提供统计学上更精确的器械性能估计。当有理由认为不同亚组的器械表现是不同时，与FDA咨询以确定可接受的设计可能是有益的。

通常，是不考虑特定基线特征或分层进行招募的，而只是根据预先确定纳入/排除标准进行招募。当预计器械在所有亚组中表现都相同时，这种类型的选择可能是充分的。

在这类临床有效性研究中，当决定研究重要的亚组或层级，例如实施研究的多个中心或被认为能高度预测受试者预后的协变量，如是否存在并行症（如糖尿病），考虑分层随机化也往往是明智的，其中每个预先确定的层级中是分别进行随机的。

***6.6 试验机构选择***

申办者应该选择适合器械预期用途的受试者招募机构（中心）。对诊断器械，测试机构通常不同于受试者招募机构。

在评价新器械的初始可行性时，单中心研究者研究可能是一个有用的起点，因为它们在逻辑上更容易协调，不是那么资源密集型，且通常关注的是一个混杂变量更少的，更均匀的人群。他们还有助于规划更大的多中心研究。一项单中心研究很少适合关键临床研究。

相反，研究器械安全性和有效性的评价通常是依赖于证明更大的多中心研究不同研究机构之间的一致结果。多中心研究的一个优势是更容易招募受试者且通常能更快达到所需的样本量。

一项多中心研究可以保证目标人群更具代表性的抽样且使得更容易推广研究结果。各中心之间结果的差异在医疗器械研究结果评价中是非常重要的，因为他们可能会反映受试者选择，手术技术和临床技能，以及任何学习曲线的差异，所有这些因素均可能对研究结果的理解产生偏倚。

同样，在诊断临床性能研究中，受试者或样本可以从其他研究中心转借且执行测试人员的技能，以及解释结果的人员可以不同。因为不同中心的研究结果和诊断临床性能研究可能会相差很大，可能需要采用特殊的统计技术组合几个中心的研究结果（如进行支持汇总的分析）。

在适用的情况下，应特别注意确保研究机构将纳入反映所治疗疾病的流行病学分布的受试者，如性别、年龄、种族、种族、社会经济地位、共存疾病等变量。此外，患有各种不同于预期使用人群的受试者（如用于筛查受试者的器械纳入罕见疾病更多受试者）可能会导致器械性能的偏倚估计（转介偏倚）。对于一些诊断研究，不同的机构可能反映受试者的特征，如高风险等或平均风险疾病，且可能需要单独考虑这些结果。（参见第6.5部分）

同样，根据器械不同，可能重要的是要考虑不同机构研究者或操作者经验的多样化。例如，对于临床结果研究，三级护理机构的外科医生可能比社区医院的外科医生具有更专业的经验。因此，这类临床研究只有选择三级护理机构才可能导致器械性能评估偏倚。

在美国进行的支持上市前提交的研究应与了解器械用于美国患者时的安全性和有效性，评价的指标博阿扩人口统计，护理标准和医学实践。这对于在美国国内外进行研究都是很重要的。不符合这些标准的研究可能不足以支持器械批准。

申办者负责选择具有研究器械所必须的培训和经验的研究者。参见21 CFR 812.43(a)。注意，对于某些新型器械，作为临床研究计划的一部分，可能需要进行额外的培训，以促进器械的安全使用。申办者应该确保研究器械的用户具有这类使用所必须的任何培训和经验。预计研究者还应了解适用的法规和指导临床研究实施的指南。

***6.7 比较研究设计***

研究比较两种或多种治疗方式，两种或两种以上诊断测试性能的研究称为比较研究设计。有几种不同类型的比较设计。

* **平行组设计**：每个受试者或样品指定接受一种可能的治疗或测试比较。因为是将不同组的受试者（或样品）分配给每个治疗（或诊断测试），在不同受试者之间进行比较。对于这种类型的研究，建议将随机到介入组的受试者确保进行公平的比较，所以在基线或测试前各组是具有可比性。
* **配对设计**：每位受试者或样品同时接受所有治疗或测试。因此，可以在相同受试者上比较不同的治疗或测试。一个实例是左右面设计的其中两个面采用不同的器械进行治疗。一般来说，配对设计往往比平行组设计更精确，因为对于平行组设计，是在受试者组间进行比较的，包括受试者之间的差异。配对设计不太常用于一些治疗或美容治疗，但在诊断研究中是很常见的,因为不同诊断器械可以用于同一受试者或样本。这种设计的一个缺点是，可能很难识别非局部不良事件是否与一种或两种介入措施相关。
* **交叉设计**：每位受试者或样品在不同的时间接收两种或两种以上的治疗（或诊断测试），但是在一个预先确定的序列中。通常会研究多个序列的介入措施，每位受试者均按指定的顺序排列接受不同的治疗（或诊断测试）。当所有治疗（或诊断测试）不能同时分配给一个受试者时（即配对设计不可行时）且当一种治疗（或诊断测试）的影响不延续到下一种治疗时，可能适合采用交叉设计。对于这种类型的设计，除非另有合理说明，否则顺序应该是随机的。交叉设计可以用于治疗、美容和诊断研究。

**7 临床有效性研究**

设计临床结果研究时，需要考虑各种重要因素。本节讨论的就是这些因素，包括：

* 受试者终点，介入分配（随机选择），盲法，安慰剂疗效，对照和非对比研究的特定考虑因素。
* 偏倚来源，一般考虑因素以便实现偏倚最小化。

***7.1* 临床研究的终点**

在任何一项临床研究中，都要选择将证明器械性能和临床结果的关键研究变量。对于临床有效性研究，这些变量是主要和次要的临床终点。很重要的一点是，所有主要终点和次要终点都是在关键临床研究的设计阶段预先确定的。

应该用最小偏倚客观地衡量器械性能和临床结果。有些考虑因素包括：

* 终点、结果或指标应该提供足够的证据来表征器械对于预期用途的临床有效性（包括安全性和有效性）。
* 方案应该规定测量那些终点或结果，如何以及何时测量，以及如何进行统计分析。单独的统计分析计划中可能包括统计分析的更详细描述。
* 终点、结果或测量指标应该在临床上是有意义的且与所规定的研究目的和预期用途是相关的。应该设计关键研究来证明对于指定受试者人群的临床受益，而不是简单证明器械的功能。
* 只要有可能，终点应该是客观的、内部和外部有效的，且可以通过最小偏倚确定。例如，当治疗终点涉及诊断评估如是否存在中风或心肌梗塞时，诊断临床参考标准优于主观评估。当有更客观的评估方法时，依靠主观临床评估以确定临床研究结果的终点通常是不充分的。在这种情况下，通过临床专家委员会裁决终点可能解决一些问题。
* 方案应该规定评估终点，结果或与研究受试者和/或研究人员相关的测量指标，例如，不清楚介入分配的评价者和研究组。
* 对于一些研究，可能有必要成立独立的裁判委员会裁决终点，例如，当没有客观评估方法且采用的是主观评价时，如x光照片的解释。按照什么样的规则裁决终点，结果或测量指标应事先在关键研究方案中进行规定。
* 当从受试者的角度来看（如减少，疼痛），最好测量感兴趣的结果和预期用途时，最好采用受试者报告结果。在这种情况下，重要的是要选择一种针对受试者人群和所治疗的病情进行过确认的并符合预期用途的评分评估方法。由于这个原因，在研究设计阶段和FDA进行早期讨论是很重要的。这些更主观的衡量指标通常配合更客观的评估方法一起使用，作为一个复合终点。当在跨国试验中使用主观等级评分量表时，申办者应该确保评分量表在不同文化和语言之间是可解释和有效的。欲了解更多有关受试者报告结果使用及其验证的信息，请参阅FDA指南。[[12]](#footnote-11)
* 复合终点是指多个终点预先确定的组合终点。复合终点的使用可能是具有挑战性的，且需要与FDA进行仔细和早期讨论以便制定合适的终点和分析计划。如果使用复合终点，除了分析器械对整个复合终点的影响外，FDA还将评价器械对每个终点组成的影响，可以在以通过其中任何一个组件或单个组件缺乏一致性的结果对符合终点进行评估。

* 对于多个主要有效性（或安全性）终点，鼓励提供每个终点的作用和相对重要性的解释。一份编写良好的方案应该能仔细规定多个主要终点中每个终点的研究通过/未通过标准，并预先规定合适的统计方法来处理多重性问题和控制总体 I 型错误率。在考虑可能有其他预期用途或声明而采用多个次要终点时，该方案应预先规定合适的统计方法来分析数据和解释结果。
* 在经过确认且与临床受益直接相关时，使用替代终点可能是合适的。在研究设计阶段与FDA进行早期讨论确定替代方法是否建立了有效的，或者如果没有，确定一种可接受的确认替代的方式，是至关重要的。

在为临床结果研究选择主要终点时应该考虑以下问题：

* 应该精心挑选终点，结果或测量以避免规定不清楚或大部分受试者无法获得这些终点的情况。
* 在设计研究的总体持续时间（包括完成入组和随访总时间），安全性和有效性终点的评价时间点，以及过知情同意过程受试者要多久同意随访。其中值得考虑的方面包括应该评价安全性和有效性以便进行风险/受益分析的最早时间点，以及批准后可能需要进行额外多年随访的可能性（就像PMA的批准条件）。主要终点的时间点选择应该考虑产品活动的时间长度，考虑既往研究获得的证据。申办者应该留心的是，如果他们在方案和通过知情同意过程承诺进行一定时间的随访，通常预计申办者需要遵循方案并在整个时间段内对受试者进行随访。
* 在特定器械研究中理解科学或医学变化，特定终点的相关性，结果或测量值可能会改变。在实验过程中改变研究终点可能使实验难以向FDA评审小组和申办者解释，特别是对于关键实验。对于严谨对照试验设计，应该预先确定所有主要元素，和任何关键要素的变化，包括研究终点，可能严重影响试验解释和数据分析。在这种情况下，建议申办者联系相应的FDA审查部门讨论最佳行动方针。

***7.2 临床有效性研究的介入方法分配（随机）***

21 CFR 860.7(f)(1)指出，研究计划或方案必须包括：

*(ii)选择受试者的方法：*

*(a)提提供足够的保证确保受试者适合研究之目的，提供带治疗或诊断病情的诊断标准，提供经确认的实验室测试（如适用），对于预防疾病或病情的情况，提供需要预防病情易感性和保良局的证据；*

*(b)应以最大程度减小任何可能偏倚的方式分配受试者（如果使用）；*

*(c)保证试验组和对照组的相关变量，如性别、疾病严重程度或持续时间，和使用治疗而不是测试器械之间的可比性；*

通常建议受试者随机分配介入组，确保进行合适的比较，这样的话在介入或测试前基线时具有可比性。随机化倾向于保证介入组的相关变量，如性别和其他人群统计学变量、疾病的严重程度或持续时间，既往治疗，专业用户偏倚和/或偏好，和研究器械以外其他介入措施使用之间的平衡。同样也很重要的是，随机同样能平衡未测量或未知的协变量。在某些情况下，可能会考虑不在个体层面而是不同层面进行随机化，例如，在临床机构层面，在这种情况下，将临床机构随机分配接受不同的介入治疗，然后特定临床机构的所有受试者被分配到同一组；这就是所谓的集群随机化。鼓励申办者在通过集群随机化分配介入治疗前咨询FDA。

在平行组临床结果设计中，通常使用随机化以公正的方式将每位受试者分配到介入治疗。在临床结果配对设计中，每位受试者作为他或她自己的对照，根据随机化分配两种介入治疗或位置的顺序（如左脸和右脸，左膝和右膝)不同，对于每个特定受试者应用每种介入治疗有助于减少偏倚。在交叉设计中，每例受试者的介入治疗的顺序通常是随机决定的。在平行研究，随机配对研究或交叉设计研究中未能随机会因允许存在扭曲结果的偏倚而使研究面临失败的风险。

在器械设计或预定受试者人群使得不可能对介入治疗分配进行随机时，这项研究可能会面临未知大小和方向的偏倚，且这种偏倚可能会负面影响研究提供的证据和所依赖数据是否有效的能力。

机构承认，在有些器械研究情况下，随机化是不可能的，很困难的，或可能不合适的。例如，研究者在向受试者推荐一项随机研究时，当他们认为研究的不同介入治疗并不是同样安全和有效时（如它们缺乏临床均势），可能面临伦理困境。在这种情况下，建议申办者向提交上市前批准申请或通知联系FDA，讨论有关随机化的问题并确定在这种情况下能提供充足证据水平的合适研究设计。

***7.3 盲法***

限制相关人员知晓介入分配情况，同时不危及受试者护理或研究目的的情况，称为设盲，也称为掩蔽。（掩蔽这个词更适合眼科产品）。在临床结果研究背景下，了解介入分配可能会影响受试者，临床医生、研究者、照顾者和第三方评价者的行为和决定，无论是有意识的或者无意识的。

如果不对受试者进行介入治疗分配设盲（遮蔽），则受试者的行为可能会因了解介入治疗方法而受到影响，因此，可能引入偏倚，特别是当临床测量或终点是主观的时候。

如果不对研究者或第三方评价进行干治疗分配设盲（遮蔽），则研究者或评价者偏倚会通过影响临床结果的解释，器械手术植入的性能，以及随后的临床决策而负面地影响研究。

在受试者和/或研究者不可设盲的情况下，仍然是可能的并强烈建议对临床测量结果和/或终点的独立的第三方评价者进行介入治疗分配进行设盲。最好是采用不知道研究目的，而是被要求根据客观标准（如临床、影像学）进行评价的评价者。另外，在了解介入治疗分配会影响评价的情况下，可以聘请采用前瞻性定义关键的定义和标准操作程序的独立核心实验室和读片中心，和/或临床事件委员会以使偏倚最小化。

在某些临床结果器械研究，尤其是那些具有高度有创或器械治疗与药物治疗或手术介入治疗进行比较，或不可能对受试者或研究者进行介入治疗分配设盲。然而，即使不方便或很困难，FDA建议如果可能的话考虑或尝试设盲。当研究设盲后，对于研究设计纳入设盲完整性、有效性评价通常是非常有用的，评价的房室是通过询问在研究结束时不清楚所在的介入分配的参与者(如受试者和/或研究者)说明他所在的是哪个介入治疗组及为什么会这么认为。这可能有助于更好地理解这项研究结果，特别是当了解分配组可能会影响总体结果时。

在在整个研究过程中不可能对研究参与者的介入治疗分类设盲时，尽可能考虑以下方法以便最大程度减小偏倚：

* 在完成潜在受试者筛查和完成入组前，受试者和研究人员应不了解介入治疗分配。
* 强烈建议受试者在手术不了解介入治疗分配以便可能与了解介入分配相关的差异性脱落。
* 如果没有设盲，更客观的终点通常优于受试者报告结果。在采用受试者报告结果的情况下，应该考虑由与研究不相关的任何临床人员测量这些指标。
* 为临床人员起草一个脚本以便使询问研究参与者的随访问题标准化。

***7.4 对比临床有效性研究的对照品***

21 CFR 860.7(f)(1)(iv)确定了四种类型的对照。其中指出研究计划或方案必须包括：

*以允许进行定量评价的房室将治疗或诊断结果与对照进行比较。必须规定对照的确切性质并提供所采用方法减少观察者和数据分析人员任何可能偏倚的解释。应记录“设盲”的水平和方法（如适用）。一般来说，有四种公认的比较类型；*

*(a)不治疗。在客观测量有效性且安慰剂效应可以忽略不计时，比较治疗组和非治疗患者比较组的客观结果；*

*(b)安慰剂对照。对于器械使用可能存在安慰剂效应的情况，比较所使用器械结果与设计用于尽可能代表研究使用条件的无效器械的结果；*

*(c)积极治疗对照。在可以采用有效治疗方案进行比较的情况下，如所治疗的病情为使用安慰剂或终止治疗是不当的或违背患者利益时；*

*(d)历史对照。在某些情况下，如涉及具有高的和可预测死亡率的疾病或可预测的持续时间或严重程度的症状和体征，或发病率是可预测的预防治疗中，使用器械的结果可能与来自于该充具有可比性患者或人群的疾病或病情没有接受治疗或者接受已经确立的有效疗法（治疗、诊断、预防）治疗的历史经验进行定量比较。*

除了CFR中确定的四种类型的对照外，本指南还考虑了第五种对照，“受试者作为自身对照。”在本指南中，在描述上述(a)和(c)中的对照时，采用术语“介入治疗”而不是“治疗”，因为这个术语适用于诊断介入以及治疗和整容介入治疗的临床有效性研究。

每种对照用于临床研究都有一定的优势和局限性。一般来说，采用并行对照的研究设计的偏倚小于那些使用非并行对照。

表1概括了每种类型的对照与研究偏倚相关的一些考虑因素和由此得到的证据水平。

表1：临床有效性研究的不同对照类型

|  |  |  |  |
| --- | --- | --- | --- |
| **对照类型** | **亚类** | **描述** | **考虑因素** |
| 并行对照 | 积极介入治疗对照（ “阳性”） | 对照组介入提供了另一种能提供已知有效性的介入治疗（通常是另一种器械或手术，但可能是药品或生物制品）。 | * 证明相对阳性对照的有效或非劣效。
* 合适对照的选择是基于预期受试者人群的现行护理标准
* 有关阳性对照有效性的了解
 |
| 安慰剂对照(“Sham”) | 对照组可能是另一种器械，模拟手术或一种被认为没有治疗或诊断有效性的药物或生物制品。 | * 如果被认为具有安慰剂效应，安慰剂对照是有用的。
* 构造一种似乎能像研究器械那样发挥功能，但不提供治疗的安慰剂对照。
* 在某些情况下，将受试者随机分配服用不能提供任何已知有效性的安慰剂可能是不道德的。
 |
| “无介入治疗”对照 | 对照组不提供介入治疗（或诊断作用）。 | * 选择“无介入治疗”对照在招募不接受介入治疗的受试者或保留被随机分配到“无介入治疗”对照组的受试者方面可能造成挑战。
* 选择“无介入治疗”对照存在固有偏倚，因为对照组受试者预计不能获得任何受益，而实验组受试者预计会获得受益。
* 尽管不同研究中心之间对照可能存在不同，“无介入治疗”对照有时护理/最佳医疗管理标准，能提供关于任何增量受益或风险的证据。
 |
| 受试者作为自身对照 | 受试者作为自身的并行对照（如左右脸，对侧眼等）。 | * 对于治疗和/或整容器械研究，受试者作为他/她自己的并行对照，以便允许有利的使用受试者本身的相关性。
* 这种设计只有当实验器械和对照介入治疗影响是局部的且不重叠才可能。
 |
| 非并行对照 | 受试者作为自身对照 | 受试者的基线结果与终点结果进行比较。 | * 对于大多数治疗研究，使用基线结果作为终点评价结果是不够的，因为受试者可能会因为与临床实验器械不相关的原因（如均值回归，安慰剂效应）而改善。
 |

|  |  |  |  |
| --- | --- | --- | --- |
| **对照类型** | **亚类** | **描述** | **考虑因素** |
| 非并行历史对照 | 平行组的受试者水平数据 | 对照组包括过去接受过治疗的不同受试者组，对于这些受试者，各受试者水平数据可用于相同结果和相同协变量，就像当前研究一样。 | * 一个重大关切是两组之间相对基线协变量的相似性。不同时间进行的比较研究的使用会引入严重的和未知的选择性偏倚。
* 协变量分析（如统计方法或倾向评分分析)可能解决一些问题。
* 历史对照组可能不能反映现有医学实践且可能包括不同于当代研究(时间偏倚)的受试者人群和/或结果。
* 当使用主观终点或者之前没有对所有必要的终点以不同的方式评价或评估，这种对照是具有挑战性。
* 这种对照在节距缺失数据含义方面提出了重大挑战。灵敏度和缺失数据分析可能解决一些与偏倚有关的问题。
 |

***7.5 安慰剂疗效和其他现象***

许多临床有效性研究的问题是，器械可能没有实际有效性，但仍可能具有演示性有效性。为了应对这一问题，会使用安慰剂器械（有时称为“sham”器械）。安慰剂器械故意设计为不提供任何明显的有效性，但可能具有演示性有效性。这种现象被称为安慰剂效应。安慰剂效应经常发生在疼痛，功能和生活质量研究中，且可能是相当大的。

采用客观以及主观终点时均可以观察到安慰剂疗效，并已知会持续数月甚至数年。

对于安慰剂疗效，有几大公认原因。

* **预期受益**— 在一项随机，盲法研究中，预计会观察到受试者受益，因为受试者可能随机分入任何一个组，这与非盲研究相反，其中“无介入治疗”对照组的受试者预计不会获得受益。
* **研究有效性**— 与安慰剂效应相关的概念是当他们知道在一项研究正正在被测量研究时，人们倾向于表现不同。此外，受试者在研究中可能得到更好或更细心的护理。这两种效应均会影响任何研究中的客观和主观报告结果。

安慰剂疗效会在研究器械改进和对照的见到那比较中引入偏倚。出于这个原因，在可能比较研究器械和无效的疗法时，需要纳入安慰剂对照。如果可以证明优于安慰剂，则可以推断，研究器械是有效的。当介入治疗对受试者，研究者和第三方评价者设盲时，这些研究工作的结果是最佳的。

尽管阳性对照的使用不允许直接测量安慰剂效应，在安慰剂效应可以认为在这两个介入组之间具有可比性的情况下，它确实允许对两组之间的相对安全性和有效性进行充分的比较。不幸的是，在采用阳性对照的随机研究中，每个组具有不同程度的安慰剂效应，与测试程序相关的“仪式”近似成正比（如开放手术的安慰剂效应已经超过口服药物）。

在诊断临床有效性研究中，当“仪式”与不同机构的临床医生成正比可能具有不同的护理标准时，在解释研究结果会存在一定的关切。安慰剂对照研究的最接近近似将是临床医生在使用器械之前并不知道哪一组受试者，为的是最大程度减小相对于标准护理的行为变化。

还有其他可能使研究结果解释困难的相关现象。

* 平均回归— 对于具有随机性元素受试者的任何测量，如果这位受试者在进入研究时具有极端的测量结果（如研究的入组标准要求），那么同一位受试者的后续测量往往会更接近总体平均值。作为一个纯粹的统计现象，那些最初“病得更重的”会倾向于改善更多。
* 在临床试验过程中增加对受试者的关注可能会导致他们改善。
* 自发缓解-有些受试者在研究过程中可能自然愈合或不再表现出症状。

***7.6 非对比临床有效性研究***

有一些临床结果研究设计并不是严谨对照研究，因为它们没有使用并行（或历史）对照，因此，没有直接对比对象。

**7.6.1 采用客观性能标准(OPC)的单组研究**

客观性能标准(OPC)是指根据临床研究和/或注册研究获得的历史数据推导的数值目标且FDA可能以二分（通过/未通过）方式用于安全性或有效性终点的评审和比较。重要的是要指出，目前很少有经确认的OPC。当器械技术已经足够成熟且可能是基于公开信息或根据一种特殊器械的所有研究汇总的信息时，通常会制定OPC。需要根据所有相关来源的既往荟萃分析审查精心构建OPC，且优选受试者水平荟萃分析。如果OPC是经医学或科学协会或标准组织委托或被采用或在FDA指南文件中描述，则该OPC往往会具有更大的有效性。OPC通常不能由单个公司只采用他们的数据或者根据相关科学文献回顾制定，OPC通常也不是由FDA单方面制定。同样重要的是要注意，随着技术的成熟和提高，OPC可能随着时间的推移而过时。希望利用OPC进行关键临床研究的申办者应该在研究开始之前与FDA工作人员讨论和合作。

**7.6.2 采用性能目标(PG)的单组研究**

性能目标(PG)提供了劣效于OPC的证据等级。PG是指被FDA认为足以用作安全性和/或有效性比较终点的数值(点估计)。在某些情况下，PG可能是根据有效性和/或安全终点的上（或下）置信限确定的。一般来说，器械技术发展得不够完善和不够成熟以便采用PG作为OPC，且用于生成PG的数据并不认为像用于制定OPC的数据那样可靠。为了挑战患者人群或如果任何对照都没有临床均势，则考虑PG。如果PG经医学或科学协会或标准组织委托或被采用或在FDA指南文件中描述，则该PG往往会具有更大的有效性。通常不推荐由申办者发起或FDA单方面针对特定注册提交制定PG。就像OPC一样，随着技术的成熟和提高，PG可能随着时间的推移而过时。希望在关键临床研究中采用PG的申办者应该在研究开始之前与FDA工作人员讨论和合作。

采用PG时需要非常小心。特别是，要问的一个重要问题是是否有令人信服的证据表明，能实现安全性能目标（或有效性）的任何器械能否在严谨对照研究中实际成功地证明这种安全性（或有效性）。PG的实现（或未能实现）并不一定导致立即接受（或拒绝）研究结果。在某些情况下，如果是混杂的或在结果中发现了不正常的信号，则需要更多定性地对研究结果进行探讨。FDA可能采用PG将PMA交给相关的顾问小组以便获取有关研究结果解释的外部科学顾问。

**7.6.3 观察性研究或注册研究**

检查临床数据库以便比较治疗有效性充满了偏倚。而临床试验随机化能防止根据预后进行治疗分配，在观察性研究中就没有这种偏倚控制保证且注册研究包含临床结果。在文献中有些例子，随机临床研究的结果明显不同于观察性研究报道的结果。差异的一种解释是，观察性研究的介入分配可能依赖于受试者的预测。

用于流行病学研究的其他设计可能要求各个病例根据匹配的重要协变量选择一个或多个对照受试者。匹配可能是有问题的，因为所选病例可能是不成比例从总体目标人群的一个子集选择的，从而导致这些对照也不能代表目标人群。在上市前研究中，无论是否是诊断器械还是治疗器械，不推荐这种类型的观察性研究。然而，在可以确定特定事件和特定器械之间关联性的上市后研究中，这有时是有用的。

**荟萃分析**

试图证明医疗器械的安全性和有效性，而不生成新的临床数据的荟萃分析可能会引入潜在的偏倚，因为结果无关紧要或预后差的研究结果通常是不发表的。在这种研究设计可能有用的罕见实例中，关键是要采用精确的统计方法和预定的严格入组和和拒绝质量以便选择已发表的文献研究使选择偏倚最小化。

一种被荟萃分析广泛接受的方法是在尝试进行任何分析之前，确定纳入荟萃分析的研究选择标准（如随机临床研究）。这种方法可以称为前瞻性荟萃分析。然而，一个重要的缺陷在于，大多数出版物并不包括受试者水平数据或足够的细节，以至于无法对每项研究的数据进行独立分析。其他常见的问题包括不同研究的不一致的纳入/排除标准，终点定义存在显著差异以及受试者的随访时间长度存在差异。重要的是要注意，荟萃分析应该只涉及申办者希望上市销售的器械版本的研究。

**7.6.4 文献总结**

文献总结可能包括由合格专家实施的记载明确的病例历史，和已上市销售器械的重大人类经验报告。

在这些报告中，没有生成新的临床数据，但他们与荟萃分析的区别在于，没有进行新的分析。包括文献总结的PMA可能取决于所选择的已发表文献中进行的分析，且有可能取决于特定研究者的证据确凿的经验。这些报告对于证明有效性很少有用，甚至比使用荟萃分析具有还有更重要的局限性。

***7.7* 诊断临床有效性研究**

对于诊断器械，关键临床研究通常是诊断临床性能研究（参见第8部分）；然而，有时需要进行临床结果研究。在诊断临床结果研究中，诊断器械结果会用于治疗或管理介入中。通过介入治疗对受试者结果的影响，部分评估了器械性能。也可能评价诊断临床表现。例如，如果疾病或病情的诊断和治疗是同时进行的（如某些内镜手术中），或者准确诊断的临床受益（临床结果改善）是不清楚的，则临床有效性研究可能是合适的。对于只需要采集样本，但诊断结果不用于确定研究管理的介入治疗，不是本指南考虑的诊断临床有效性研究。

安全性和有效性是通过合适的临床终点，诊断性能，或两者同时进行测量的。为了全面评价安全性和有效性，有时需要设置对照组，但临床医生不会使用对照组的诊断结果。为了比较临床实验组和对照组，平行组或配对设计可能是合适的。

在一些对照诊断临床有效性研究中，临床医生必须了解受试者是在实验组还是对照组，因为临床医生知道，如果她（他）是否正在使用诊断结果。不过，只要有可能，评价临床终点的临床医生不应该了解受试者在哪一组。

***7.8 一些临床有效性研究的优点和缺点***

对于一种给定的器械和预期用途确定合适的研究设计取决于很多因素，包括器械的特点、使用条件、相同预期用途的替代治疗（或诊断测试），关于器械使用的足够警告，以及有关该器械的经验程度。同样重要的是要考虑最终所需标签要求和使用说明书，因为该研究需要提供足够的证据支持标签声明。考虑到特定器械类型的具体约束和特征，FDA建议可能为研究器械选择研究设计的申办者提供必要的证据证明对于所提出的预期用途的安全性和有效性的合理保证。

有一些研究设计有可能比其他研究设计提供更高等级的证据。选择只能提供低水平证据的研究设计可能需要说明该设计合适的理由，并以支持预期用途的方式充分控制潜在的偏倚。每当申办者认为临床结果研究严谨对照，随机和/或设盲是不合适或没必要时，申办者应该解释为什么可能忽略可能的偏倚。研究设计得越能最大程度的减小偏倚，证据等级将越强（其他任何事物都是相同的）。

以下部分描述的临床有效性研究中常见的研究设计优缺点。

**7.8.1 随机、双盲、对照、平行组临床研究**

考虑并行研究设计时，这种研究设计通常是优先的，因为它通常可以提供最高等级的科学证据和最小的偏倚。双盲表明受试者或研究人员（包括研究者或任何第三方评价者）都不清楚介入治疗的分配。本研究设计提供了最高保证水平，确保实验组和对照组的受试者人群具有可比性，同时避免了已知的和未知的可能影响安全性和有效性结果的基线变量的组间系统差异。然而，有些器械的设计既不可行，也不实用。

为这种研究设计选择的对照可以是积极的治疗或安慰剂（参见第7.5部分）。当有可能存在安慰剂效应的情况下，或者当主观结果测量指标作为研究终点时，偏离这种研究设计尤其成问题。尽管安慰剂对照的使用可能是可取的，因为这样的设计能提供研究器械受益和风险的直接证据，剥夺对照组受试者治疗权利通常是有问题的。因此，积极或安慰剂对照的选择可能取决于伦理和实际考虑因素。在考虑阳性对照时，一个重要的考虑因素设计研究旨在否为了证明优效性，在在与阳性对照比较的情况下，证明非劣效性。

**7.8.2 随机、受试者作为自身对照，配对临床研究**

在这样的研究设计中，受试者可以同时采用研究和对照介入治疗进行治疗。实例包括一半脸接受研究器械治疗，而另一半脸接受对照介入治疗。在这种设计中，介入治疗的分配是随机的（如脸的一侧）。当器械作用只是局部明显时，可以采用这种研究设计。当采用这种研究设计时，不可能评价和区分系统性安全或有效性结果。这种研究设计的优势（如果使用得当）是在同一受试者中测量两种介入治疗的作用且变异性较小，所以可能需要较小的样本量。

这种研究设计的另一种类型是两组交叉设计研究，其中每位受试者均按顺序接受研究和对照介入治疗，并按随机方式分配。同样，这样的设计允许比较每位受试者的研究器械和对照介入治疗的性能。然而，采用这种设计时，需要假设第一种介入治疗的有效性不会延续到第二个介入期。当这种假设不合适时，可能需要在研究的不同介入治疗之间间隔较长的时间（“洗脱期”）。

**7.8.3 采用并行对照的随机、非盲研究（积极，安慰剂或“无介入治疗”）**

采用并行对照的随机，非盲研究和第7.8.1部分和第7.8.2部分所述研究设计之间的主要区别在于设盲不完整或不设盲。不完整设盲是指为对研究者，研究者或第三方评价者设盲。当均不设盲时，研究通常被称为开放性研究。正如上面所讨论的，在比较临床研究中，如果受试者，研究者和第三方评价不清楚介入分配，则偏倚可以最小化。然而，采用积极治疗或安慰剂对照时，可能并不总是可能对受试者或研究者设盲，有时甚至对第三方评价者设盲也可能是一个挑战（如研究器械和作为阳性对照的器械成像时具有完全不同的外观）。

在对照是“最佳医疗管理”或“无介入治疗”对照的情况下，研究通常对受试者和研究者不设盲。因此，对照组的每位受试者都知道他或她不接受研究器械治疗。了解这些信息通常造成未知程度的偏倚。

如果研究参与者不设盲，即使采用了真实目标终点，也很难评估由此导致的偏倚大小，且可能威胁原本严谨的研究的科学性。在对任何或所有参与者（受试者、研究者、评价者）均不可能设盲的情况下，应向FDA提供所提出的用于解决与偏倚相关问题的方法的详细原理和解释。

**7.8.4 采用并行对照（积极或安慰剂或“无介入治疗”）的非随机研究**

在采用并行阳性对照的非随机设计中，受试者和研究者不对介入治疗分配设盲。因此，这种研究设计囊括了随机、非盲研究与并行对照设计的所有缺点。此外，因为没有随机化且每位受试者仅接受一种可能的介入治疗，因此，非常可能的是因介入治疗分配造成已知程度的偏倚。

通常不推荐采用这种设计，因为与随机研究一样是劳动密集型的，但由于不同研究组，机构，研究者，可能存在的差异（包括未测量但是可能混淆差异）会引入了更多的偏倚。即使研究总的来说两个介入组之间似乎是平衡的，每位参与人员可能并不平衡，可能存在研究者-器械相互作用，其中研究器械的优势似乎会因研究者而不同。

**7.8.5 与基线比较的单组研究**

在很多治疗性研究中，一个非常重要的考虑因素是，尽管可能会尝试采用受试者的基线状态作为对照，设置采用积极治疗或安慰剂对照（或者甚至是一个“不介入”控制）的随机组通常是可取的。盲法研究中这样一个随机组将提供更严格的控制并避免安慰剂效应偏倚以及时间偏倚。

**7.8.6 采用历史对照或信息的单组研究**

当器械技术开发成熟且对感兴趣的疾病了解透彻，可以进行采用历史对照或一些历史信息的单组研究。

* 与提供受试者水平数据的历史对照组比较：

如果可以提供受试者水平数据（包括历史和当前研究每位受试者的所有重要变量），至少可能做一些统计比较。挑战在于证明历史对照与当前研究具有可比性。在当前研究完成后，有可能会采用倾向评分模型来评估两组之间的可比性，然而，存在一个重大风险是，最终的数据可能不具有可比性。在受试者入组并采集和分析基线之前，没有办法评估可比性，因此这种方法可能会有风险。

采用历史对照的固有明显偏倚是时间偏倚，因为各组不是并行的。这种时间分离会带来两个介入组可比性的担忧以及医学实践会因目标受试者人群和预期结果造成的变化而很可能已经发生了变化。因此，这种设计的缺点历史对照的受试者结果可能不明显或不适用于当前目标人群。

* 与根据历史信息衍生的OPC和PG比较

如果无法提供历史对照组，则可以通过与数值目标值OPC或PG（与安全性或有效性终点有关）比较来评价器械性能。这样的研究设计涵盖了与历史对照比较的所有挑战和局限性。此外，没有独立的方法来评估当前组可能与衍生OPC或PG的历史组间的可比性如何，且无法量化偏倚的。

由于这些研究不涉及对照组，与OPC和PG比较不能证明优效性或非劣效性。

***7.9 IDE申请考虑因素***

对于临床有效性研究，根据21 CFR 812.20，申办者的IDE申请必须包括所提出研究设计的细节和选择该研究设计的理由。申办者应该解释为什么提出研究设计足够可靠。具有较少偏倚可能性的任何替代研究设计（曾经考虑过但被拒绝）的讨论对于有效评审所提出的研究设计将是有帮助的。

**8 诊断临床性能研究**

对诊断器械，关键临床研究通常是一项诊断临床性能研究。在这类研究中，诊断器械的诊断临床性能可通过量化诊断器械输出与临床相关参考标准之间紧密程度的衡量指标确定[[13]](#footnote-12) （在第8.2部分已规定），其中这些参考标准是用于评估受试者的目标条件的。合适的临床性能测量指标的选择取决于器械的预期用途、诊断器械输出的性质和临床参考标准。诊断临床性能研究的目标是确定器械诊断临床性能并相关支持与器械用于目标人群的临床性能相关的有利风险/受益分析。

诊断临床性能研究往往是受描述与器械测量相关的实验室研究，非关键临床研究和/或研究（如IVD的分析研究）驱动的[[14]](#footnote-13)。例如，在检测高风险人乳头瘤病毒(HPV)DNA预测30岁以上巴氏试验结果正常女性患宫颈癌时，考虑体外诊断器械。

在临床性能研究中评价了人乳头状瘤病毒检测预测宫颈癌（目标条件）的能力，而器械测量高危HPV DNA的菌株（感兴趣的测量）是在分析性能研究（包括但不限于，评估测量偏倚，精度，检测限和定量限，线性，干扰和遗留 ）中进行的。

诊断器械的安全性和有效性往往是不分离的。当诊断器械报告的结果不正确（如结果被错误地归类为假阳性或假阴性）或误解时，受试者可能因为后续管理不善或心理创伤而受到伤害。对于二分（有/没有）目标条件，诊断器械的安全性和有效性通常可通过其正确识别目标条件的存在与否予以确认。除了误诊外，诊断器械有时在样本采集或器械使用过程中可能会引入安全问题。例如，它可能使受试者遭受辐射或其他形式的能量，导致使用又创检查，或导致需要给予治疗性产品治疗。在这些情况下，在评价研究设计的合适性和确定器械是否会对受试者构成重大风险时，也会考虑诊断临床性能研究中受试者的风险。[[15]](#footnote-14)

影响诊断临床性能研究的临床研究设计的关键因素可能包括定义研究设计的器械预期用途，合适研究人群的选择，和降低特定偏倚来源。在接下来各部分进一步讨论了这些因素。

***8.1 预期用途考虑因素***

诊断器械的预期用途相差很大，这些器械提供的结果类型也是如此。因此，诊断临床性能研究的设计有会相应变化。很多诊断器械尝试通过特定目标状态或疾病的存在，缺少、发展阶段对受试者进行分类。提供测量功能（如病毒载量、血糖水平或视网膜厚度）的其他诊断器械作为诊断评价的辅助手段或用于受试者监测。

关键诊断临床性能研究应该支持诊断器械的预期用途。一种诊断器械可能用作一种单纯的诊断器械取代现有的诊断器械或程序，或者，也可能与其他信息结合使用（有时通过使用一种算法）用于评估受试者的目标状态。或者，诊断器械可能提供辅助诊断信息（例如，额外的信息并不毁掉基于现有器械或程序提出的建议）。

在设计诊断性能研究时，应该在预期使用环境中评价器械，包括以下各方面（如适用）：

* 器械测量或测试指标；
* 器械的报告内容；
* 所检查的细胞、组织、器官、局部或系统；
* 样本来源、样本类型和样本矩阵
* 器械如何使用（按照使用说明书）；
* 器械何时使用（使用条件）；
* 器械由谁使用 （操作者或目标用户）；
* 用于什么样的疾病（目标条件）；
* 器械用于什么样额人群（目标人群）。

改变器械已批准或许可的使用方法可能使得有必要与FDA讨论以确定必须进行何种类型的评价。

***8.2 目标条件的临床参考标准***

理想情况下，诊断器械临床性能的表征要求单独知道受试者的真实状态与目标状态（参见第4.1节）。在诊断临床性能研究中，使用临床参考标准评估每个受试者的目标条件[[16]](#footnote-15)。临床参考标准（适用于监管目的）是确立受试者真实状态与目标状态的最佳方法（如用于确定乳腺癌的活检病理结果）。它可以是单一方法或不同方法和技术的组合，包括临床随访，但不应该考虑研究器械输出。

正如在其他FDA指南中所描述的[[17]](#footnote-16)，“最佳方法”是根据医学，实验室和监管协会的现有实践建立的。有时会选择几种可能方法中的一种。在其他情况下，会构建一种复合方法或达成专家小组共识。

在研究开始前，应预先确定临床参考标准。还应该充分详细地描述，包括确定其结果所必须的所有信息。尽可能，应该记录研究中每位受试者的这些必要信息。理想情况下，应该确定每位受试者的参考标准结果。

有时是没有临床参考标准的，无法提供，或者由于其有创性本质不能用于临床研究[[18]](#footnote-17)。在这种情况下，可以规定和使用不是临床参考标准（非引用标准）的参照物，并将这些结果与研究器械的输出进行比较。例如，研究用乙肝检测通常与多个经FDA批准的乙型肝炎病毒标志物检测结果进行比较，而不是与乙型肝炎病毒临床参考标准进行比较。申办者在计划不使用临床参考标准的研究之前应向FDA咨询，以确保该研究将支持器械的预期用途。不使用临床参考标准评估目标条件的诊断临床性能研究被称为一致性研究。在一致性研究中，不能直接估计诊断器械的“正确性”；研究器械可能会与非参考标准参照物一致，但也可能与可用的临床参考标准的结果是不对应的。在比较两种诊断器械结果背景的其他FDA指南中讨论了协议测量指标解释的问题。[[19]](#footnote-18)

因为随着知识的增加和医疗系统的进步，临床参考标准随着时间推移可能会发展，应在使用临床参考标准的背景下报告和解释参考标准诊断临床性能的测量指标。当测试开发人员认为，新器械和临床参考标准之间由于临床参考标准错误而存在不一致，申办者在进行这样的研究之前应向FDA咨询。

***8.3 诊断性能评价的研究人群***

为支持器械的预期用途而进行的研究选择受试者或样品的机构应该能代表器械预期使用的各个机构类型。受试者或样品也应该代表所推荐的目标人群。非代表性机构或受试者总体性能估计可能存在选择偏倚。必须在方案中规定研究选择受试者或样品的实际方法 （参见21 CFR860.7(f)(1)(a)）。之前在第6.3部分已经描述了不同选择方法的优缺点。

入组研究的受试者应该能代表目标条件谱。当入组的受试者与目标条件谱不匹配时，诊断临床性能的估计会面临谱系影响。例如，如果仅对目标条件的特殊情况采样（如健康正常受试者或患有终末期疾病的受试者)，则性能可以似乎比真实的要好。这是因为删除的目标条件谱中间虎贲的受试者往往更难以正确诊断。

有时，目标人群包括罕见病受试者，包括招募难以治疗且治疗费用昂贵的罕见疾病受试者。完全代表受试者人群罕见疾病（与目标人群的比例相比）的设计有时可能是合适的。然而，这种设计的总体性能估计可能存在潜在偏倚，所以应该在统计分析计划中考虑这种潜在的偏倚。

***8.4 研究计划、受试者选择和样本采集***

诊断器械可能直接测试受试者得到受试者特定的数据，或可能测试从受试者采集的样本。可以立即采集和测试样本，或在某些情况下，可以在测试前采集并存储。在采用预先确定方案时，样本或受试者是指前瞻性获得的数据，且只获得满足方案标准的受试者的数据。根据没有预先规定预期用途混合的或不同研究（如生物库）预先确定方案部分采集的样本并不认为是前瞻性获得的，但可以获得用于回顾性研究。同样，从直接测试受试者的器械（如ECG，EEG，图像集）采集的受试者数据可以存储下来供以后选择和分析；这是另一种类型的回顾性研究。

在前瞻性研究计划中，采用的是预先规定的方案。这样的方案会预先规定研究设计，包括入组/排除标准，受试者招募和选择的方法，测试方案和拟采用的分析方法。在整个研究期内选择符合入组/排除标准的受试者。严谨执行的前瞻性计划可能帮助确保研究人群充分代表目标人群，以使该研究能提供证据支持预期用途。

在某些情况下，用生物库样本或以前采集的受试者数据（例如，当目标条件非常罕见且很难以前瞻性方式获得足够数量的目标条件患者）补充前瞻性研究或只使用生物库样本或受试者数据评价器械性能是可以接受的，前提是潜在偏倚和本指南中讨论的其他问题可以得到充分解决。当生物库样本或以前采集的数据或图像添加到研究时，进行测试或解释测试结果的人员不能识别所增加样本或前瞻性获得样本的数据。

之前储存样本或数据的回顾性选择可能引入其他问题。在有一些回顾性研究设计中，研究者用器械所使用的现有数据，样本、图片或其他存储媒体或信息寻找受试者。回顾性选择的示例包括到三级护理中心获得样本或采用从涉及长期随访的既往研究获得的注册数据。

一般来说，对于以前瞻性方式选择的样本或受试者，选择过程由研究者控制。相比之下，回顾性受试者或样本选择可能仅限于诸如有存储样本和有临床参考标准结果的受试者。回顾性选择样本或受试者数据的一个问题可能是不能代表目标人群（回顾样本或数据可能只代表目标条件的极端情况）。另一个问题是回顾性样本的便利样本可能混淆诊断器械的临床性能和组织存储或处理，或目标条件的协变量预测。因此，没有前瞻性计划的回顾性获得样本和受试者数据的使用会引发很多可能的问题，包括档案中样本或受试者数据是因何采集的（有关当前目标人群的代表性），样本可能降解或用于采集样本和存储数据的技术随着时间推移可能已经改变，以及非随机消除存档样本。欲了解更多信息，请参见FDA关于使用剩下人体样本的知情同意[[20]](#footnote-19)。申办者应该咨询FDA以便确定可用样本或受试者数据是否适用于支持诊断器械的预期用途。

在设计任何类型的诊断临床性能研究时，采集样本或受试者数据的方案是很重要的。IVD，受试者准备、样本采集、贮存和处理程序是按研究方案中应充分描述的关键组件。对于除了IVD之外的诊断器械，测量或数据采集过程是一个关键组成部分。研究方案应该描述如何进行受试者测量或如何采集结果，包括特定说明（如特定刺激过程，特定电极植入，数据采集过程中的特定受试者条件)。

***8.5 诊断临床性能对比研究***

诊断临床性能研究的目标是确定研究器械的性能。比较研究器械诊断临床性能和已确定器械或方法的诊断临床断性能的比较研究只有在使用临床参考标准时才可能。建议设计此类研究的申办者在设计阶段咨询相应的FDA评审部门。

当无法提供临床参考标准时，研究器械有时与一致性研究中的另一种器械进行比较。非常高水平一致性可能表明研究器械非劣效于已确定的器械。然而，高水平一致只有在已知已确定器械具有可接受水平的性能时才有意义。

***8.6 诊断性能研究中设盲***

诊断器械的临床研究可能涉及多次评估和用户/阅片者。例如，诊断性能的临床研究可能包括研究器械的用户/阅片者，获得临床参考标准结果的人员，有时包括用于对比研究的已确立器械的用户/阅片者。研究诊断器械的用户不应该知道（所以应该设盲）临床参考标准或根据其他诊断评估得到的结果，反之亦然。

***8.7 与器械交互人员的技能和行为（从测试概念）***

诊断器械的使用通常需要接受过不同水平的培训或具有不同技能的人员进行多项活动，例如：外行，抽血者，实验室技术员，病理学家，放射科医生。这些活动可能包括采集和准备样品，将器械定位到受试者身上，和解释视觉输出。当任务要求具备通过培训获得的技能、学科知识、阅读图像和/或波形的能力，和经验时，人类表现的差异并不少见且可能会影响器械的性能。因此，在评价诊断器械的临床性能时，临床研究方案应该会导致与器械交互人员的表现差异。

有时，有必要进行由不同与器械交互的人员进行的更多研究以检验特定器械的性能。申办者需要记录为这项研究的参与人员提供的培训，并提供培训材料标记为用于FDA评审。在用于同一位受试者的两种诊断评估的临床性能对比研究中，当由同一个经过培训的人员阅读或解释评估结果时，可能引入读片顺序偏倚。在这类研究中，由于不能在同一时间完成不同输出（如图片、幻灯片），它们是按照某些预先确定顺序进行的。当按顺序阅读将两种不同评估方法应用于同一受试者或样本时，了解其中一种评估的结果可能会影响另一种方法的评估结果。对第二次评估的影响也可能是因简单地增加评估时间而混淆。减小阅读顺序偏倚的一种方法是在评估间隔很长一段时间 （ “洗脱”期），以便消除阅片者第一次评估的记忆。其他措施也是可能的，但我们建议申办者咨询FDA评审部门以便获取进一步的信息。

如果进行诊断临床性能研究的背景不能反映医疗实践，临床性能估计可能存在偏倚。例如，根据给定环境的不同，目标条件的流行病学可能会有所不同，并可能因此影响诊断器械性能的估计结果。在较高的患病率的情况下，阅片者/解释者可能会更经常地将研究器械结果认为是阳性的，从而也影响诊断器械性能的估计结果。申办者应该考虑这些类型的偏倚会如何影响他们器械性能的估计记过，并尝试确保尽可能地控制它们。

***8.8 诊断临床性能研究的常见偏倚类型***

在第6.2部分中，讨论了最大程度减小偏倚的重要性且某些类型的偏倚会影响在前面几节介绍的诊断临床性能研究。这些和其他一些类型的偏倚（不能提供详尽的列表）进一步描述如下。我们应该认识到这些偏倚并减轻或消除这些偏倚。

* 选择偏倚：选择研究人群时的系统误差，因此，并不能代表预期用途的人群。

o谱偏倚：一种选择偏倚是诊断临床性能研究不会在不同人群亚组中引起测试性能的变化或异质性。当预期用途人群的不同亚组之间的性能有差异且研究不能完全代表所有亚组时，可能发生偏倚。

* **验证偏倚：**通常情况下，将单个临床参考标准应用于研究的所有受试者。当临床参考标准只应用于研究受试者亚组时，则必须相应调整性能估计，否则它们将有可能引起验证偏倚。例如，在应用临床参考标准的频率取决于研究器械结果时，会产生验证偏倚（如，因为当研究器械结果为阴性时，不会经常应用临床参考标准）。
* **疾病进展/回归偏倚：**当 (a)未能同时收集到同一个受试者的研究器械结果和临床参考标准，和(b) 在此期间，疾病自发恢复或发展为一个更晚期的阶段时，会发生疾病进展/回归偏倚。对于确定当前健康状况的研究器械，用于研究的研究器械和临床参考标准或其他诊断器械应该在几乎同一时间应用于受试者。
* **交付时间偏倚**： 采用诊断器械筛选的受试者似乎会因为被称为交付时间偏倚的偏倚而错误被认为受益。当知道测试结果的时间晚于本应该知道测试结果的时间时，从测试开始计算的受试者生存时间似乎不会那么好，但相对于较晚时候检测，早期检测增加了存活时间，结果似乎会更好。这可能是当临床实践的不同区域的筛查间隔不同时，这可能是一个特定的问题。
* **病程长短选择偏倚（幸存者偏倚）：**与那些目标疾病病程很短的受试者相比，目标疾病病程很长的受试者更有可能被纳入临床研究。这些受试者通常具有更好的预后且可能不能代表目标人群的受试者。因此，研究受试者的生存时间估计可能长于目标人群的生存时间。
* **外推偏差**：研究人群的条件或特征不同于测试将应用的那些人群。
* **阅片顺序偏差：**在比较两种或以上的诊断评估结果时，读片者的说明受到竞争评估结果记忆影响。
* **由于缺乏独立评价造成的偏倚：** 如果是在用于开发或改进器械的一个或多个方面的研究中进行评价的，则诊断器械的性能可能会被夸大。一旦诊断器械的所有性能方面已经完成，应该在一项不同于用于开发器械的任何初步研究的新研究中进行评价。例如，器械的诊断临床性能包括了采用不同的独立数据评价（测试）（以便避免可能的偏倚（“过度拟合”））的数据集开发的算法（ “培训”）。

**9 保持临床研究的质量**

本部分讨论是保持临床研究证据水平的计划和技术。适用于临床有效性研究以及诊断临床性能研究。

临床研究生成的证据允许科学有效的评价医疗器械的安全性和有效性。有助于这些证据生成的一个关键因素是研究设计的选择，也将有望减少偏倚的来源。应用声学科学方法谨慎地进行研究和分析数据应能使研究信息最大化。执行不佳或分析不当的研究将降低根据所生成的证据评价器械安全性和有效性的能力。质量研究实施有助于保持证据有问题的临床研究的证据水平。

调查的申办者负责确保正确监测研究实施情况 (21 CFR 812.40)。在设计开发阶段应制定计划和技术以便优化临床研究中产生的数据和信息的可靠性和有用性。这些计划和技术应能解决临床研究的各个方面，如临床数据处理，实施临床研究，计划分析战略，和前瞻性引起在研究过程中可能发生的变化。下面将进一步讨论这些方面。

***9.1 临床数据处理***

标题21 CFR 812.150(a)要求申办者准备并提交某些信息的完整、准确报告且21 CFR 812.140(a)要求准确、完整和当前记录某些信息，如每位受试者的病史。

FDA强烈建议研究申办者在临床研究开始时建立遵循良好临床数据管理实践原则(GCDMP)的数据管理计划和临床数据处理的全面的培训[[21]](#footnote-20)。在电子环境下实施，报告，记录，和分析的临床研究越多，对于建立证据水平，减少偏倚，并生成可靠，有用的数据，数据管理计划是至关重要的。虽然FDA法规没有要求向FDA提交数据管理计划以便进行评审，FDA鼓励申办者在临床研究方案中包括数据管理总结部分。要求召开包括生物学研究监测部门的工作人员讨论临床数据处理最佳实践的预提交会议对申办者也是有利的。

应该按照一致的格式和结构收集研究数据，以便可以很容易解释、理解和评估这些数据。在不同研究，机构和研究者之间保持有效的数据采集标准分方法可以帮助确保获得高质量的数据。此外，通过比较相关的元数据, 可能有助于不同方案设计的解释。鼓励利用标准词汇表和要求采集数据，因为这样能优化数据采集和提高数据质量和可预测性。[[22]](#footnote-21)

***9.2 研究实施***

FDA会仔细审查按照IDE进行的临床研究的进展报告，且如果他们不能按照GCP的方式进行研究，有权取消任何研究者进一步参与临床研究。请参阅21 CFR 812.119。FDA有关临床研究实施的法规，包括可以提供GCP指南和适当人类受试者保护等信息[[23]](#footnote-22)。

FDA有关数据监测委员会(DMC)的指南提供了协助临床研究申办者确定DMC对于研究是否有帮助以及关于DMC应该如何运作的信息[[24]](#footnote-23)。

按照这些计划和管理建议，临床研究将提高研究数据的质量并增加研究将足以支持器械批准的可能性：

* 应该精心保留随机代码和程序。如果采用自适应随机化，应该保留用于创建概率分配的算法和数据。
* 应该严格保持研究盲法且应该评价盲法的完整性。我们建议申办者保留所认为的非盲事件的记录。
* 应严格遵守研究方案并使所有类型的方案偏倚最小化，包括那些被认为是轻微的偏倚。方案应该规定被认为是严重或轻微的偏倚。应详细报告所有方案偏倚。不可接受的严重方案偏倚可能会使得无法推广研究结果。
* 应该根据研究方案滴受试者进行一致和完整的随访。在研究设计和实施阶段应努力减少由于受试者失访引起的数据缺失的发生和影响。例如，方案可能确定了失访或脱落联系程序，包括拒绝接受进一步治疗或有效性评价的继续安全性随访。虽然统计技术可用于解决失访和数据缺失的问题，这些技术通常采用的主要假设是对于特定研究不能完全确认。因此，解决由于失访引起的数据缺失问题的最佳方法是在临床研究计划和管理过程中最大程度降低其发生。然而，除了灵敏度分析外，研究方案应预先规定合适统计数据分析方法以便解决缺失数据的问题。
* 应实施警惕数据监测以便确保获得可靠、准确的数据并最大程度降低数据缺失。应该根据培训和经验选择临床研究助理或学习监察员对临床研究进行监查，以确保研究的质量、可靠性和完整性，包括人类研究受试者的权利，安全，和幸福。研究监查员不得参与研究的实际实施（如受试者评估或记录数据）。应该实施临床质量保证计划以确保研究按照设计和预期进行。
* 应保持一致的依从性和/或承诺最佳临床护理（如药物治疗策略，使用接受过器械或对照品使用培训和具有专业技能的操作者，一指的随访过程和策略）。
* 除非在统计分析计划中预先计划好，应小心保护研究数据以防止因看起来老旧而产生偏倚。这一点也适用于开放标签研究。
* 除非研究加护中预先规定了计划中期分析或停止规则或中止决定是基于安全考虑做出的，否则应准备好测量以避免研究过早中断。当然可以在无需特意进行FDA提交的情况下提前停止研究，但因缺乏资金研究过早停止将导致主要分析失败。然而，申办者继续有义务对这类中断研究的受试者进行安全相关的监测。
* 申办者必须选择研究者资格的培训和经验调查器械21 CFR 812.43(a)。机构所有的研究人员（临床医生、研究协调员等）应充分训练。研究者的培训应该正确记录。

临床研究设计和方案应包括足够的程序来解决，优化和降低上述所有方面的考虑因素。如果计划已从任何临床研究标准（如ISO 14155:2011），则应该在研究方案中标记各项标准。应该针对如果使用这种标准的情况下正确实施对研究者和研究监查员进行培训。

关于方案偏倚，FDA已经发现，有些参与的临床研究者不遵循已获得批准的方案，因为他们不同意研究设计某些方面的内容。FDA鼓励研究申办者让前瞻性临床研究者参与研究方案制定的讨论，以便方案可能出现的问题和潜在的偏倚在确定最终方案之前得到解决。这些讨论可能改进研究设计，否则可能导致方案偏倚，这对于研究分析和汇集数据将是个问题。此外，申办者必须从每位参与的研究者处获得经签署的方案，其中包括研究者承诺按照方案，临床实验计划，21 CFR 第812部分以及审查IRB或FDA批准强加的其他条件。21 CFR 812.43(c)(4)(i)。

***9.3 研究分析***

执行不到位，不合适，和/或事后分析可能对支持器械安全性和有效性证据的有用性产生不良影响。因此，研究方案应该包括详细的、预先确定的统计分析计划，包括评价计划，尽可能的程度，研究设计过程中所做的关键假设（如交叉研究设计中遗留影响的评估，生存分析中的危害比例，或不同临床基地或地理区域之间的汇集分析）。研究完成进行数据分析时应遵从这份预定义的SAP，以支持研究所生成证据的有效性。

一般应避免无计划的事后分析和偏离方案中规定的分析人群。事后分析的实例包括在没有合适理由的情况下采用不同的统计分析，主要终点变化，或使用不是预先确定的亚组进行分析。这些事后分析可能放大I型错误的概率和危及精心设计和实施良好的研究的科学有效性。方案应该预先规定灵敏度分析以便证明推论对于任何已知的偏倚来源都是可靠的。同样重要的是，批判性分析缺失数据对根据研究得出结论的影响。

在某些情况下，事后分析可能补充预先确定的分析，只要用这种类型分析的适当怀疑程度清楚地描述和解释。

***9.4 关键研究的预期变化***

在某些情况下，中期分析结果或不良安全事件的发生可能导致有必要更改器械设计以提高器械在关键临床研究过程中的安全性和/或有效性。例如，更改器械设计可以重要到足以要求采用不同版本的器械治疗的受试者按照单独层级并单独分析，从而导致质疑这些数据是否可以跨层级汇集。应与FDA讨论不同介入治疗组考虑因素的建议。为了减少器械开发后期进行器械设计变更的发生率，申办者应该在投入更多资源密集型关键研究之前进行可靠的探索性研究阶段。

相比之下，研究设计中流的变化可能是计划的，这样的话所研究的受试者人群可以进行汇集。可以提前计划一些适应性研究并集成到研究设计中。具体来说，研究设计的间隔修改（如样本量变化，随机方法修改）可以前瞻性纳入方案以便通过贝叶斯方法[[25]](#footnote-24)或不同频率中期分析方法维持研究统计分析的完整性。可以提前计划适应性设计，以便根据研究结果对研究进行具体修改。这样的设计应该解决由谁决定一旦获得了中期结果是否实施修改（如DMC独立统计学家）。如果申办者正在考虑贝叶斯或适应性试验设计，他们应该尽早寻求FDA输入。没有预先计划的适应性试验会严重削弱关键研究的科学性。

在合适的情况下，FDA也可能在分阶段或有条件批准所计划的受试者队列的亚组试验，在入组的同时平行地回答某些突出的问题。一旦包含必要额外信息的IDE补充资料已经提交给FDA并允许继续进行，则将允许申办者扩大招募。在这种情况下，研究设计可能需要提供适应性研究以便在允许招募扩大之前，研究中断时间是最短的。

**10 临床研究计划或方案**

研究计划或研究方案是提供了详细设计、实施和临床研究分析详细计划的书面文件（参见21 CFR 812.25和21 CFR 860.7(f)(1) ）。方案应包括以下内容：

* 科学的研究依据；
* 待评价的受试人群的定义（包括纳入/ 排除标准）；
* 确定器械计划的预期用途；
* 研究终点列表；
* 将应用于受试者的程序声明；
* 分析方法总结，基于研究数据的评价，包括采用的任何合适的统计方法。

同时建议，方案应包括（对于治疗和整容器械）足够的关于主要终点统计分析的统计详情，以便提供样本量计算的理由。方案应该包括清楚地描述分析数据的精确策略的完整统计分析计划。完整的统计分析计划既可以包含在方案，也可以作为一份单独的文件附上，但应该参考相关的方案。在任何情况下，在获得任何结果数据之前，应该完成完整的统计分析计划，以保护研究的科学完整性。

记录就研究方案所做出决定的里有，特别是关于所选择的临床研究设计且临床终点将通过提供支持所提出研究设计和终点的解释而有助于FDA评审临床研究。为什么不选择减少偏倚的其他终点或替代研究设计的理由可能有助于研究的评审。

美国食品药品管理局欢迎有机会在关键研究设计开发过程中通过预提交程序提供非正式的建议和反馈。同时建议在研究设计阶段寻求研究者输入。临床数据管理在提供输入研究设计输入和基于过去运营类似临床研究的病例报告表设计方面发挥着重要作用。

**11 术语表**

在这个术语表中，各术语将根据其用于本特殊指南的特定解释进行定义。

**阳性对照研究（积极治疗对照研究）**

一项使用有效性先前已经得到确立的介入治疗的研究。在器械研究中，阳性对照品可以是已经批准用于该手术的器械（药物或生物制品）。

**整容器械**

预期通过身体物理结构的改变引起受试者视觉外观的预期变化。

**一致性研究**

诊断器械结果与不是来自临床参考标准的结果进行比较的诊断临床性能研究。

**风险-受益评估**

使用器械带来的可能健康受益相对器械使用造成的任何可能的伤害或疾病。

**偏倚**

相对真理的系统误差简称。在临床研究过程中，偏倚会导致就研究结果做出错误的结论。

**设盲（遮蔽）**

对个人或个体组施加一定的条件，让他们不知道受试者或受试者样本的介入治疗（或试验）的分配情况。对于眼科器械研究，不合适用“设盲”这个词来形容这种情况。

**临床研究**

参见临床研究部分。

**临床结果研究**

受试者被分配接受其中一种介入治疗，然后在计划的时间间隔采用经确认的评估工具进行研究以便评估临床结果参数或其经确认的替代参数以确定介入治疗安全性和有效性的研究。

**临床参考标准(CRS)**

确定受试者真实目标状态的现有最佳方法；它可以是单个方法或各种方法和技术的组合，包括临床随访，但不应该考虑研究器械输出。

**临床研究**

使用人体或样本，实施用于评价治疗性，美容性或诊断器械安全性和有效性的系统性研究（参见临床研究）。

**对比参照**

用于评估研究器械性能水平的手术或其他医疗产品。通常对比参照是另一种医疗器械。

**并行对照**

**复合终点**

预先规定的几个终点的组合。

**感兴趣病情**

参见目标病情术语。

**并行对照**

根据研究器械同一时间段内所收集数据的对照。

**对照**

用于比较亚久器械的器械、药品、生物制品、医疗程序，或“无介入治疗”。

**对照组**

在临床结果研究中，接受对照治疗的受试者或样本组。

**对照临床研究**

临床研究通过与对照组，比较器械或临床参考标准的结果比较评价研究器械的性能。

**交叉设计（交叉研究）**

受试者接受一系列不同介入治疗（或诊断测试）的研究。在最简单交叉设计研究情况下，每位参与者首先接受研究器械或对照器械治疗，而在接下来的时间段内接受其他器械治疗。必要的时候，在两个时期之间设置适当时长的“洗脱期”来减轻早期介入治疗的有效性。每位受试者的研究器械或对照应用的顺序通常是随机的。

**数据监测委员会(DMC)**

一群具有相关专业知识，定期评审一项或多项正在进行的临床研究获得的累积数据人。如果存在安全隐患或者研究目的已经实现，DMC可能建议终止研究。有时也称为数据安全监测委员会(DSMB)。

**诊断临床性能研究**

通过性能衡量指标表征器械性能，这些指标对每位受试者的诊断器械输出结果与临床参考标准的一致性如何进行量化，其中临床参考标准用于评估受试者的目标条件。

**诊断器械**

一种提供预期单独使用或在其他信息背景下使用结果，以帮助评估受试者目标状态的器械。

**探索阶段**

包括初始开发、评价、首次人类使用和其他可行性研究的医疗器械临床开发阶段。

**可行性研究**

初步临床研究，例如，考察关键研究是否可行或提炼关键研究的研究方案。可行性研究有时也被称为试点研究。

**药品临床试验管理规范(GCP)**

设计、实施，性能、监测、审计、记录、分析和报告临床研究的标准，能确保数据和所报告的结果是可信和准确的，并保护研究受试者的权利、安全、健康、完整性和隐私。

**良好临床数据管理实践或GCDMP**

现行临床数据管理的行业标准，包括最佳商业实践和可以接受的监管标准。

**历史对照组**

关键研究前观察的受试者对照组。从这个对照组收集的数据用于比较研究器械的性能。

**介入治疗分配机制**

将研究受试者分配研究组或对照组的方法。

**研究器械**

1)在涉及一位或多位受试者的临床研究中研究一种未经批准的新器械或目前正在销售的器械的未经批准的适应症，以便确定器械的安全性或有效性。2)研究目的-器械（包括过渡器械），其中过渡器械是指FD&C 法案520(l)条所规定的器械，也就是说，一种在1976年5月28日被FDA认为是一种新的药物或抗生素药物的器械（参见21 CFR 812.3(g)和(h)）。

**体外诊断(IVD)器械**

一种预期用于采集、准备和检查采自人体样本的诊断器械。

**学习曲线**

学习使用用于外科植入，用于器械植入手术的医疗器械速率的图形表示。可以用达到想要结果或者直到保证实现成功结果的手术次数衡量。

**证据等级**

对于任何给定介入治疗或诊断测试，受益和威海估计有效性的总体置信水平。

**医疗器械**

仪器、设备、器具、补充物，机器，植入物，体外诊断试剂，或者其他类似或相关物品，包括任何组件、部分或附件，(1)得到美国国家处方集，或美国药典，或任何补充版本承认

(2)其使用旨在达到对人类或其他动物疾病的预防、诊断、治疗、监护、缓解；或(3)预期影响人类或其他动物的结构或任何身体功能，且不能通过人体或动物体内的化学作用达到其主要预期目的且不依赖于被代谢实现其主要目。（参见FD&C第201 (h) 部分）。

**荟萃分析**

根据单独但类似（即具有可比性）研究得到的数据进行统计分析，从而对汇集结果进行定量总结。

**非劣效研究**

通过大于规定的阈值，目的在于证明研究器械的安全性或有效性并不比低于对比器械的研究。

**非盲研究**

没有设盲的研究，也称为开放标签研究（同时参见开放标签研究）。

 **“无介入治疗”对照**

受试者没有接受介入治疗（包括安慰剂）的对照。在治疗研究中，这也可以称为时“无治疗”对照。

**客观性能标准(OPC)**

根据临床研究和/或注册研究获得的历史数据推导的数值目标且FDA可能以二分（通过/未通过）方式用于安全性或有效性终点的比较。

**观察性研究**

得出介入治疗对受试者可能产生的影响的研究，但研究者并没有将受试者分配到介入治疗组。

**开放标签研究**

参与者，医疗护理专业人士，和其他人都知道在该研究中给予的是哪种介入治疗或诊断测试的临床研究（参见非盲研究）。

**配对设计**

在同一时刻对同一受试者或受试者样本应用两种或以上的介入治疗或诊断测试。如果介入治疗或测试相互干扰，则这种设计可能不合适。

**平行组设计**

每位研究受试者或受试者样本仅分配其中一种介入治疗或诊断测试的（非配对）设计。

**性能目标**

被认为FDA足以用于关键研究结果和安全终点，有效性终点比较的数值，或者，在诊断临床性能研究，一种诊断性能的衡量指标。

**试点研究**

参见可行性研究。

**关键阶段**

医疗器械的临床发展阶段，期间收集证据用于支持医疗器械安全性和有效性的评价。该阶段由一项或多项关键研究组成。

**关键研究**

采集证据支持医疗器械对于其预期用途的安全性和有效性评价的明确研究阶段。

**安慰剂(Sham)**

一种被认为是无效的器械。在临床研究中，实验介入往往与安慰剂比较，以便评估介入的有效性(参见安慰剂对照研究)。

**安慰剂对照研究**

将特定研究器械使用的结果与在相似条件下使用的无效器械结果进行比较的对比研究。

**安慰剂效应**

无效器械使用后发生的生理或心理的变化，但这不是器械任何特殊属性造成的。变化可能是有益的，反映了参与者的期望，通常，是器械使用人的期望。

**方案(研究方案)**

临床研究所基于的研究计划。方案会描述诸如何种类型的人可能参与该研究，测试时间安排，程序，药物，和剂量；研究的时间长度。

**随机化**

参与者分组的过程，使得每位参与者都具有已知，通常时平等的机会被分配到给定组。

**随机研究**

参与者被随机（即偶然）分配到临床研究的两种或两种以上介入治疗（或诊断测试）其中一种的研究。

**选择偏倚**

由于选择进入研究人群和选择进入预期用途人群之间的系统差异造成的偏移。例如，在临床结果，平行组研究中，当选择研究组和对照组使其影响研究结果的不同于彼此，则研究结果可能面临选择偏倚。

**样本**

采集用于检查，研究，或一个或多个数量或特征分析的体液或组织分离部分。

**样本矩阵**

其中可能含有感兴趣的分析物（例如、脑脊液、血清、血液、其他组织或病毒媒介）的介质或环境。采集用于检查，研究，或一个或多个数量或特征分析的体液或组织分离部分。

**分层**

将人群划分成相互排斥和详尽亚类（称为层级），就所研究的特征而言，被认为比总人群更均匀。

**分层（亚组）设计**

将目标人群分为受试者子集（或层级）并从每个子集（或层）单独选择受试者的设计。

**研究终点**

用于判断研究有效性的主要或次要结果。

**优效研究**

旨在证明研究器械的安全性或有效性优于比较器械的研究。

**目标状态**

器械预期使用的状态。对于诊断器械是指健康、疾病、疾病阶段的过去，现在或未来状态，或受试者体内任何其他可识别状态；或应提示进行临床操作如开始，修改或终止治疗的健康状况。

**测试器械**

参见研究器械。

**治疗性器械**

用于治疗特定病情或疾病的器械。



1. 鉴于本指南的目的，“studies（研究）”这个词与“investigations（研究）”是相同的。 [↑](#footnote-ref-0)
2. 参见美国食品药品管理局得HDE持有人、机构审查委员会（IRB）、临床研究者和FDA工作人员指南-人道主义器械豁免（HDE）法规

详见问答登陆网址：http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm110194.htm [↑](#footnote-ref-1)
3. FDA的GCP 指南文件列表，请参考http://www.fda.gov/RegulatoryInformation/Guidances/ucm122046.htm. [↑](#footnote-ref-2)
4. 参见行业与美国食品药品管理局工作人员指南；医疗器械上市前审批和新分类中做出受益-风险决定时要考虑的因素，网址 http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM296379. pdf. [↑](#footnote-ref-3)
5. 参见行业与美国食品药品管理局工作人员指南；医疗器械上市前审批和新分类中做出受益-风险决定时要考虑的因素，网址http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM296379. pdf. [↑](#footnote-ref-4)
6. 参见美国食品药品管理局安全和创新法案第602条（FDASIA，公法，112-144）。 [↑](#footnote-ref-5)
7. 不符合“器械”定义的诊断产品，例如：因为它们主要是通过化学作用或在体内实现目的的器械，本指南是没有涵盖的。 [↑](#footnote-ref-6)
8. 参见 医疗器械使用安全性：在风险管理中纳入人为因素工程 （2000年7月18日)，http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm094460.htm [↑](#footnote-ref-7)
9. 在某些情况下，对于用作治疗产品的比较诊断器械的体外诊断器械，非最终器械应该用作疗效产品的临床试验。如存在这种情况，需做出提前全面方案和执行“桥梁”研究以便确定用于销售的体外诊断器械的临床有效性。 [↑](#footnote-ref-8)
10. 临床试验人群和人种数据收集指南，2005年9月，网址 http://www.fda.gov/RegulatoryInformation/Guidances/ucm126340.htm. [↑](#footnote-ref-9)
11. 为了按性别提供有关临床试验结果评价的信息，FDA已经发布的草案指南，“行业与美国食品药品管理局工作人员指南-医疗器械临床研究中性别差异的评估”

网址http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM283707. pdf. 在定稿后，本指南将代表机构有关这个主题的意见。 [↑](#footnote-ref-10)
12. 行业指南：患者报告结果测量指标：用于医疗产品开发支持标签声明，有关本指南的更多信息，请参考 http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. [↑](#footnote-ref-11)
13. 受试者目标条件的独立评估有时被称为“黄金标准”、“地面真理”或“真理标准”；然而，这些术语也有其他含义，所以我们在本指南中不使用这些术语。 [↑](#footnote-ref-12)
14. 评价器械测量质量的研究的详细讨论超出了本文件的范围。CDRH认可很多评价器械测量质量的标准和指南。CDRH认可的一列标准可在FDA网站上找到，网址http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Standards/default.htm 访问日期：2012年2月。 [↑](#footnote-ref-13)
15. 参阅 “IRB，临床研究者和申办者重大风险和无风险医疗器械研究的信息表指南”，网址http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126418.pdf 访问日期：2011年3月 [↑](#footnote-ref-14)
16. 这个定义并不限制目标条件为二分性质（有/无）；否则，这个定义与FDA “评价诊断医疗器械试验结果报告的统计学指南”是相同的，2007年3月13日，网址

http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm和 Bossuyt 等人与诊断准确性标准（参见 CLSI 统一术语数据库，网址：http://www.clsi.org/Content/NavigationMenu/Resources/HarmonizedTerminologyDatabase/Harmonized\_Terminolo .htm; 访问日期：2011年2月）。 [↑](#footnote-ref-15)
17. 行业和FDA员工指南：评价诊断测试结果报告的统计学指南(2007年3月13日)，参见http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm，和评价非诊断医疗器械研究结果报告的统计学指南，参见http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm106757.htm [↑](#footnote-ref-16)
18. 一个实例是难以使用腹腔镜/活检确定在明显健康受试者人群中是否存在卵巢癌。

评价诊断医疗器械试验结果报告的统计学指南(2007年3月13日)，参见http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm~~.~~ [↑](#footnote-ref-17)
19. 使用不能单独识别的剩余人体样本进行研究的体外诊断器械的知情同意，2006年4月，网址http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm078384.htm [↑](#footnote-ref-18)
20. 使用不能单独识别的剩余人体样本进行研究的体外诊断器械的知情同意，2006年4月，网址http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm078384.htm. [↑](#footnote-ref-19)
21. 参见 http://www.scdm.org/gcdmp/. [↑](#footnote-ref-20)
22. CDISC (http://www.cdisc.org) 和HL7 (http://www.hl7.org)标准组具有更多关于数据标准的信息。 [↑](#footnote-ref-21)
23. 参见http://www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/default.htm. [↑](#footnote-ref-22)
24. 临床试验申办者指南：临床试验数据监测委员会的建立和运行

网址http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM127073.pdf [↑](#footnote-ref-23)
25. 行业和FDA员工指南： 贝叶斯统计学用于医疗器械临床试验的指南（2010年2月5日），网址http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm [↑](#footnote-ref-24)