自**2007年1月4日**起，本文档的联系信息已更新为以下内容：

有关本指南使用或解释的问题，请联系Loren Zarmeba，电话：240-276-3666或电子邮件：loren.zaremba@fda.hhs.gov

本指南编写于1997年2月27日实施FDA的良好指导规范（GGP）之前。其不会为任何人创造或赋予任何权利，也不对FDA或公众具有约束力。如果替代方法满足适用的法律、法规或其两者的要求，可以使用替代方法。本指南将在下一版本中更新，以纳入GGP的标准部分。

**草案** 1994年9月26日

**用以支持体外器械的510 (R)文档提交的数据收集所需注意的事项**

由于1990年实施的“安全医疗器械法案”（SMDA 90）、FDA临床审查委员会的报告，器械评估办公室（ODE）倡议以注重实际的方式提高对上市前提交进行的科学审查的质量，对有关体外诊断器械的FDA法规进行了变更。IVD的安全性与导致患者误诊的概率成反比。SMDA 90扩大了FDA审查IVD的范围，并授予FDA在必要时请求临床数据的权利，以确保新型IVD与具有相同预期用途且合法出售的IVD有相同的安全性和有效性。

以下几点代表了关于支持IVD提交510 (k)认可的数据收集的当前FDA政策。本文档旨在补充但不取代现有的用于上市前提交的FDA指南文档。

研究方案：

在数据收集过程之前、之中必须为所有临床中心确定统一方案（51）。应清楚记录、说明研究设计中的任何变化并将其体现在数据解释中。应将方案副本作为510 (k)提交文档的一部分。临床研究方案应该清楚地定义研究人群以及纳入和排除标准。

抽样方法

1. 应在方案中明确说明用于选择患者样本的抽样方法的性质，并在提交文档中报告。抽样可具有目的性（由研究者定义的分布范围），允许进行分析方法评估（其中需要大范围的正常和异常样本来表征表现）。设计用于说明预期临床表现的研究可能需要随机和盲（隐形）抽样。
2. 定量试验：

与合法销售的类比器械相比，使用定量测量技术对试验进行的评估应至少包括随机和系统误差的评估（1-8，12，15，20，27，29，32，36-38，40，41，43，44，46- 48，51，53，56，62-64）。与参考方法或“黄金标准”相比，可以在两个器械之间直接和／或在新器械和旧器械之间间接对比。这些对比研究应包括回归分析的关键参数，包括预期斜率和截距以及其95％的置信区间。对于所有试验，（在新试验和类似试验之间发现统计差异时的要求），应确定器械用途的决策点，并且应当对每一处决策点执行误差分析（52，54）。

考虑所有已知或预期的混合因素的同时，应该评估分析特异性。基于可报告分析范围的较低端结果进行预期临床决策时，还需要确定分析敏感性。

线性回归分析的统计理论要求在用于计算斜率、截距和其95％的置信区间的各种样本对中的数据具有独立性（即每个患者只有一个样本对）。如果使用的多个样本对来自于相同的患者，则需要适当说明和统计程序来说明患者间和患者内的变异性。

1. 定性试验：

使用定性测量技术对试验进行的评估应包括使用临床／诊断特性的试验的特性描述（9，13，14，16，18，19，21-23，24，30，33-35，33-35，39，42，45，49，55，56，59-62，66）以及／或使用鉴别区和临界点的分析特性的说明（17，23，26）。

应选择在临床相关临界值附近的阳性对照（校准物），但不应限于高滴度阳性对照。

1. 评估器械的统计方法

用于评估510 (k)提交的统计方法应适用于本研究方案、收集的数据类型和IVD的预期用途。该方法应该是一个公认的标准方法（有关拟定方案，请参见随附参考文献）或应该清楚解释和说明。

置信区间应尽可能包括在统计分析中（11，28，31）。ANOVA模型是首选方法，因为详细的ANOVA表可提供相关信息，例如指定的统计模型，效应类型（随机、固定或混合）、变异源、自由度、平方和和均方、F统计以及p值。对于ANOVA模型，应呈现来自ANOVA而非置信区间的ANOVA。（32）

1. III级510 (k)审查的应用

如果以下条件适用：分析物已确定，但存有新的方法或新的矩阵，提高根据510（k）要求提供临床数据的能力可以允许FDA将一些IVD作为III级上市前通告510（k）申请而不是上市前审批申请（PMA）来处理。这些提交文档应包括临床或诊断的说明和实际结果，以及新器械对于其预期用途的分析性能特性（9，13，14，16，18，19，21-23，24，30，33-35，39，42，45，49，55，56，59-62，66）。这需要将新试验与既定实验室或临床标准进行比较，并根据诊断敏感性和诊断特异性报告新试验结果。

用于生成提交用数据的器械

理想情况下，研究应该使用生产模型器械（既定读头、光学器件等）的最终版本来进行，而非使用原型。仅基于原型器械的数据可能引起对器械在非专业环境下发挥性能的能力的担忧。

机构审查委员会审查

所有研究必须在机构审查委员会的审查下进行。

提交文档中所有研究的质量控制

应使用CLIA'88法规概述的公认质量控制方法进行研究。在任何运行期间（其中，试验似乎不受控）收集的数据不用于支持提交。

研究场所要求：

用于收集数据以及确立上述性能特征所需的研究场所的类型和数量将取决于提交类型。如果IVD的分析物、方法和预期用途已经确定，则除IVD认可外，无需进行极为严格的试验，其中，IVD的这些组件未完全确定。

* 1. 对于具有定义不明确的参考标准或方法的IVD，且预期将出现基于人群和／或性能的场所差异，制造商应提供至少三个研究场所。所有场所应遵循统一方案。
	2. 对于已标准化的器械，其中，不会出现预期基于人群和／或性能的场所差异，则单个研究场所可能足够。应清楚说明和记录使用的样品来源，特别是研究场所是非临床中心的时候。
	3. 对于在医生办公室实验室推广的器械，必须至少包括三个研究场所，并包括代表可能进行试验的个人类型的器械操作者，例如护士、医生、医疗办公室助理等。
	4. 对于非处方（OTC）或家用器械，必须至少包括三个研究场所。一般来说，几百个未经训练的外行用户或一般消费者应参与在这些场所的试验。对于所有OTC器械，即使这些器械仅为采样器械，也要必须对包装说明书标签和说明书的可读性和理解性以及外行人员对试验信息的使用进行评估。如果外行人员应实际执行OTC IVD的分析程序，则必须使用配对样品将外行人员操作时的试验性能与专业实验室人员操作时的试验性能进行比较（50，53）。

试验应基于IVD的设计和预期用途。器械是否限于收集和邮寄样本？试验是否由外行人员执行？谁来解释结果？制造商向普通用户提供的支持和帮助是什么，例如由医疗保健专业人员配备的免费电话号码。

临床研究主要研究者的责任：

在制造商自己的机构之外的研究场所进行研究时，责任（主要）研究者必须在研究上签字，表明研究方案已经落实到位，并在整个研究过程中得到遵循，且研究者已经审查和验证了向FDA提交的提交文档中的数据和制造商的数据分析说明。这与同行审查的医学期刊的要求相似。

产品说明书（包装说明书）有关研究数据的声明：

7. 所有数据的性质应在510（k）和包装说明书中清楚记录。制造商应指明用于确定510（k）和包装说明书（PI）中描述性能特性的研究场所的数量和类型。制造商对IVD的声明将限于FDA认可的实际试验数据。

示例：

制造商将试验仅限于制造场所，受制于PI中的声明，例如：

**“此性能数据仅在x公司的场所生成”**或一些等同性语言。

然而，制造商在制造场所和两个临床中心获得性能数据，可在PI中做出声明，例如：

**“性能数据于德克萨斯州的x公司、的大型**（大学，一般等）**（高复杂性）医院实验室以及新泽西的小型医院实验室（中等复杂性）中生成”**或一些等同性语言。

一些IVDS，特别是测量与低发病率条件相关的分析物的IVDS，可能需要超过3个试验场所。制造商可以在PI中声明FDA已批准的所有试验数据的结果。

PI的目标是标记中提供真实信息。PI标签应向用户说明有关产品试验性能已知信息的实际范围。公司可以根据其产品的实际试验使用其数据来支持声明。

参考文献

1. Barnett，R.N.，一种比较定量方法的方案。美国临床病理学杂志。43，562，1965。
2. Barnett，R.N. 和Youden，W.J。比较定量方法的修订方案。美国临床病理学杂志。54，454，1970。
3. Bauer，S.，Kennedy，J.W. ，临床实验室应用统计：111。随时间变化。临床实验室杂志。Autom. 1：35-40，1982。
4. Bauer，S.，Kennedy，J.W.，临床实验室应用统计：VII。评估准确性的初步概念。临床实验室杂志。Autom. 2：354，1982。
5. Bauer，S.，Kennedy，J.W. 临床实验室应用统计：VIII。不准确措施。临床实验室杂志。Autom. 3:46，1983。
6. Bauer，S.，Kennedy，J.W.，临床实验室应用统计：IX。分布和正常性。临床实验室杂志。Autom. 3：183，1983。
7. Bauer，S.，Kennedy，J.W.，临床实验室应用统计：X。评估不准确。临床实验室杂志。Autom. 3:340，1983。
8. Bauer，S.，Kennedy，J.W. ，临床实验室应用统计：XI。学生的t试验用于评估不准确。临床实验室杂志。Autom. 4:56-63，1984。
9. Beck，J.R.，Shultz，K：在试验性能评估中使用相对工作特性曲线（ROC）。病理学与检验医学文献。110:13，1986。
10. Begg，C.，诊断试验评估中的偏见。医学统计。6:411，1987。
11. Braitman，L.E.，置信区间从数据中提取临床上有用的信息。内科学年鉴 108:296，1988。
12. Broughton，P.M.G.，Gownlock，A.H.，Mccormack，J.J.，Neill，D.W.，用于评估临床化学中使用的自动仪器的修订方案。临床生物化学 11:207-208，1974。
13. Brown，G.W.，比较两个试验。联合医学。pp.28- 31，1985。
14. Buck，A.A.，Grant，J.J.，流行病学研究中筛查试验与参考试验的比较 - 1.协议指标及其与患病率的关系。美国流行病学杂志。83:586，1966。
15. 美国病理学家学院，实验室仪器评估，验证和维护手册，1989。
16. Coughlin，S.S.，Trock，B.，Criqui，M.H.等人，诊断试验的灵敏度，特异性和预测价值的逻辑建模。45：l-7，1992。
17. Delger，D.M.，Rea，M.R.，Kessler，G.等人，评估非定量测定的标准：应用于血清绒毛膜促性腺激素。临床化学。33:150，1987。
18. Delong，E.R.，Vernon，W.B.，Bollinger，R.R.，监试验验的灵敏度和特异性。生物识别。41:947-958，1985。
19. 在制造和临床评估中考虑的要点，以检测针对人类免疫缺陷病毒的抗体，1型（1989），生物制品评估和研究中心。
20. Emancipator，K.，Kroll，M.H ，非线性的定量测量，临床化学。39:766，1993。
21. Gail，M.H.，评估患有复发性疾病风险的患者的连续癌症标志物研究。生物特征 37:67-78，1981。
22. Galen，R.S.，Gambino,S.R. 超越正常，医学诊断的预测价值和有效性。New York: Wiley和Sons，1975。
23. Garrett，P.E.，对我来说看起来不错，还是应该评估定性免疫测定？临床免疫杂志 10:147，1987.
24. Goodman，S.N.，Berline，J.A.，在规划实验时使用预测的置信区间，并在解释结果时使用电力遗漏。内科学年鉴 121:200-206，1994。
25. Greenberg，R.A.，Jekel，J.F.，确定结核菌素试验的假阳性和假阴性率的一些问题。美国呼吸疾病综述 100：645，1969。
26. 诊断试剂盒评估指南 - 第2部分：定性试验试剂盒评估的一般原则和概要程序。欧洲临床实验室标准委员会，1987。
27. Hackney，J.R.，Cembrowsky，G.S.，需要改进仪器和套件评估。A.J.C.P. 86:391，1986。
28. Harris，E.K.，在P值和置信区间（为什么我们不能有更多的置信度？）临床化学39:927，1993。
29. Harris，E.K，回归，最小二乘法和相关性。在CRC临床化学手册第111卷。Boca Raton，Florida，CRC Press，pp.193-213，1982。
30. Henderson，A.R.，评估试验准确性及其临床后果：受试者工作特征曲线分析的引物。临床生物化学 30:521，1993。
31. Henderson，A.R.，化学置信度：临床化学需要分析和其他数据的置信区间？ 临床化学。39:929，1993。
32. Henry，J.B.，Beeler，M.F.，Copeland，B.E.和Wert，E.B. 用于描述临床病理学方法的形式。美国临床病理学杂志。52:296，1969。
33. Hsu，S.L.，Walter，S.D.，估计诊断试验的错误率。生物识别。36:167，1980。
34. Ilstrup，D.M.，微生物统计学方法。临床微生物修订版 3:219，1990。
35. Kent，D.L.，诊断试验的临床影响的决策分析和评估。实验室医学 22:718，1991。
36. Krouwer，J.S.，Steward，W.N.，Schlain，B.，用于评估随机访问分析器的多因素实验设计。临床化学。33:1894-6，1988。
37. Krouwer，J.S.，多因素设计IV。多因素设计如何通过考虑方案特定的偏差来改进总误差的估计。临床化学。37: 26-29，1991。
38. Krouwer，J.S.，估算总分析误差及其来源 - 改进方法评估的技术。病理学与检验医学文献116:726，1992。
39. Lachs，M.S.，Hachamkin，I.，Edelstein，P.H.等人，诊断试验评估中的频谱偏差：尿路感染快速试验仪试验中的经验教训。内科学年鉴 118: 230，1993。
40. Lasky，F.D.，通过使用患者标本相关性来分配校准器值，实现常规临床化学方法的准确性。病理学与检验医学文献117:412，1993。
41. Lawton，W.H. Sylvestre，E.A.，Young-Ferraro，B.J.，多重分析程序的统计比较：临床化学应用。技术计量学。21:397，1979。
42. Metz，C.E.，ROC分析的基本原则。核医学杂志8:283，1978。
43. 国家临床实验室标准委员会。评价定量分析方法的线性度：提议的指南，1986。
44. Lasky，F.D.，通过使用患者标本相关性来分配校准值来实现常规临床化学方法的准确性 - 一种用于管理基质效应的方法。病理学与检验医学文献117:412-419，1993。
45. 国家临床实验室标准委员会。使用受试者操作特性（ROC）绘图评估实验室试验的临床精度：暂行指南，1993。
46. 国家临床实验室标准委员会。使用患者样本的方法比较和偏差估计；暂行指南，1993。
47. 国家临床实验室标准委员会。临床化学中的干扰测试；提议的指南，1986。
48. 国家临床实验室标准委员会。临床化学装置精度性能评估；暂行指南，1992。
49. 国家临床实验室标准委员会。传染病免疫试验规范；1991。
50. 国家临床实验室标准委员会。家用体外试验产品的标签，拟定的指南。
51. Mills，J.L.，Data torturing,，N.E.J.M.，329:1196-1199，1993。
52. Peters，T.，Westgard，J.O.，方法评估，第7章：Tietz，N.W.ed，临床化学基础，第三版：Philadelphia，Saunders，pp.225-237，1987.
53. 关于家用体外诊断器械的标签和上市前提交的要点，1987年，器械与放射健康中心。
54. Powers，D.M. 确定和维护性能声明 - 制造商的观点。病理学与检验医学文献， 116:718，1992。
55. Poynard，T.，Chaput，J.，Etienne，J.，诊断测试的有效性，疾病流行率和不可解释结果百分比之间的关系。医学决策. 2:25，1982。
56. Ransohoff，D.F.，Feinstein，A.R.，评估诊断试验功效的频谱和偏倚问题。N.E.J.M. 299:926，1978。
57. Ross，J.W.，精度评估。CRC临床化学手册 - 第一卷。Boca Raton，Florida，CRC Press，pp.391-422，1982。
58. Schatzkin，A. Connor，R.J.，Taylor，P.R.，Bunnag，B.，比较新老筛查试验时，不能对所有筛选进行参考程序。美国流行病学杂志. 125:672，1987。
59. Staquet，M.，Rozencweig，Lee，Y.J.，Muggia，F.M.，评估新的二分法诊断性试验方法。慢性疾病杂志 34:599，1981。
60. Thibodeau，L.A.，评估诊断测试。生物识别。37:801，1981。
61. Valenstein，P.V.，以不完善的标准评估诊断试验。美国临床病理学杂志 93:252，1990。
62. Walter，S.D.，Irwig，L.M.，评估错误率，患病率和错误分类数据的相对风险：综述。临床流行病学杂志。41：923-937，1988。
63. Westgard，J.O.，精度和准确性：方法评估测试的概念和评估。临床实验室科学评论. 13:283，1981。
64. Westgard，J.O.，Carey，R.N.和Wold，S.，判断方法开发和评估的精度和准确性的标准。临床化学。20:825，1974。
65. Westgard，J.O.，Hunt，M.R.，在方法比较研究中使用和解释常见的统计检验。临床化学。19:49-57，1973。
66. Zweig,M.H.，Campbell，G.，受试者操作特征（ROC）图：临床医学中的基础评估工具。临床化学。29:561，1993。