行业和食品药品监督

管理局工作人员指南

II类特殊控制指导性文件：
卵巢附件肿块评估评分测试系统

**发布时间：2011年3月23日**

关于本文件的问题，请联系Donna Roscoe，电话301-796-6183

（Donna.Roscoe@fda.hhs.gov），或Marina Kondratovich，电话301-796-6036

（Marina.Kondratovich@fda.hhs.gov）



**美国卫生及公众服务部**

**食品药品监督管理局**

**器械与放射健康中心**

**体外诊断器械评估和安全办公室**

**免疫学和血液学器械部**

**前言**

**公众评论**

贵公司可随时将供机构考虑的书面意见提交至美国食品药品监督管理局文档管理部，5630 Fishers Lane，rm.1061，（HFA-305）Rockville, MD,，20852，供部门审议。电子意见提交至http://www.regulations.gov.请使用在《联邦公报》发表的生效通知中所列出的文档号标识所有意见。可能直到文件下次修订或更新时，评论才会被机构受理。

**其他副本**

其他副本可从网上获取，网址http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ ucm237299.htm.贵公司也可以向Dsmica@fda.hhs.gov发送邮件，获取指南的电子版，或向301-827-8149发送传真请求获取打印件。请使用文件编号（*1707*）标识贵公司所需的指南。

**目录**

**[I.](#_Toc479613593)****[简介](#_Toc479613593)** [5](#_Toc479613593)

**[II.背景](#_Toc479613594)** [6](#_Toc479613594)

**[III.范围](#_Toc479613595)** [7](#_Toc479613595)

**[IV.健康风险](#_Toc479613596)** [7](#_Toc479613596)

**[V. 器械描述](#_Toc479613597)** [8](#_Toc479613597)

[A. 背景 8](#_Toc479613598)

[B. 质量体系法规（QS Reg） 8](#_Toc479613599)

[C. 预期用途/适应症 9](#_Toc479613600)

[D. 测试的理论依据 9](#_Toc479613601)

[E. 测试组件和方法学 9](#_Toc479613602)

[F. 测试结果 11](#_Toc479613603)

**[VI.分析性能验证](#_Toc479613604)** [11](#_Toc479613604)

[A. 标本 12](#_Toc479613605)

[B. 可重复性/再现性 12](#_Toc479613606)

[C. 单独分析物的线性度 14](#_Toc479613607)

[D. 低水平时的性能 14](#_Toc479613608)

[E. 干扰 15](#_Toc479613609)

[F. 交叉反应/非特异性结合 15](#_Toc479613610)

[G. 单独分析物的钩状效应 15](#_Toc479613611)

[H. 遗留污染 16](#_Toc479613612)

[I. 矩阵比较 16](#_Toc479613613)

[J. 稳定性 16](#_Toc479613614)

[K. 校准和对照 16](#_Toc479613615)

**[VII.软件](#_Toc479613616)** [16](#_Toc479613616)

**[VIII. 临床性能评估](#_Toc479613617)** [18](#_Toc479613617)

[A. 研究人群/样本 18](#_Toc479613618)

[B. 临界/临床决策点 19](#_Toc479613619)

[C. 临床参考标准（“金标准”） 19](#_Toc479613620)

[D. 研究设计 20](#_Toc479613621)

[E. 其他良性和恶性疾病下的预期值 25](#_Toc479613622)

[F. 参考间期 25](#_Toc479613623)

[G. 评分中所含单独分析物的相关性 25](#_Toc479613624)

**[IX.标签](#_Toc479613625)** [26](#_Toc479613625)

**[X. 参考文件](#_Toc479613626)** [28](#_Toc479613626)

**行业和食品药品监督管理局工作人员指南**

**II类特殊控制指导性文件：卵巢附件肿块评估评分测试系统**



***本指南代表食品药品监督管理局（食品药品监督管理局）当前对本主题的见解。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。如果贵公司希望讨论一种替代方法，请联系负责实施本指南的FDA工作人员。如果贵公司无法确定适当的FDA工作人员，请拨打本指南标题页上列出的合适的电话号码。***

1. **简介**

开发本文件的目的是作为一份特殊控制指南，支持卵巢附件肿块评估评分测试系统对II类器械进行分类（特殊控制）。卵巢附件肿块评估评分测试系统测量血浆中一种或多种分析物，将这些值与单独评分结合，随后用于确定尚未转诊肿瘤学家的女性中术前附件肿块为恶性的可能性。本测试用于经医师将其与患者的临床和放射学评估相结合，以确定患者是否应转诊给妇科肿瘤学家进行手术。

卵巢附件肿块评估评分测试系统未指定为卵巢癌的筛选测试或用于诊断。（这些测试相关贴标所需的其他规定请参见第IX章。）预期用于计划进行手术的患者，不得用于决定患者是否应接受手术。本指南不适用于基因表达试验或基于组织的试验。

本指南为制造商提供了关于计划上市前通告和卵巢附件肿块评估评分测试系统特表的建议。本文件中的建议适用于对全血制品中所得的单独一种或多种蛋白进行测量的测试。医师使用结果或评分作为辅助测试，以补充，而非取代其他诊断和临床程序。计划进行手术干预的女性应在医师独立术前评估、或卵巢附件肿块评估评分、或二者均提示有恶性肿瘤可能性时，转诊给妇科肿瘤学家。

本指南与宣称卵巢附件肿块评估评分测试系统分类的《联邦公报》通知一同发布。任何为卵巢附件肿块评估评分测试系统提交510（k）上市前通告的厂商均应处理本特殊控制指南中包含的问题。厂商必须显示其器械可解决本指南中识别的安全性和有效性问题，可通过符合本指南的建议，或通过一些可提供等效安全性和有效性保证的其他方式解决。

指定本文件作为特殊控制意味着，任何当前上市或预期上市卵巢附件肿块评估评分测试系统的厂商均需处理本特殊控制指南中包含的问题。厂商需显示其器械可解决本指南中识别的安全性和有效性问题，可通过符合本指南的建议，或通过一些可提供

等效安全性和有效性保证的其他方式解决。

**II.背景**

医师可在所有年龄段女性中常规发现盆腔附件肿块，可在标准妇科评估过程中意外发现，或在由于女性表现出症状而进行的检查后发现。约5至10%的女性将对可疑的卵巢恶性肿瘤进行手术，这些肿块中有13至21%将被诊断为卵巢癌（参考1）。对有附件肿块的患者进行鉴别诊断和管理的指南已经确立，其中包括将有可疑卵巢癌的患者转诊给妇科肿瘤学家。研究显示，卵巢癌患者在由妇科肿瘤学家进行手术时，与一般妇科病学家和外科医师相比，无进展生存期和整体生存期有改善。这些已发表的观察结果和指南支持通过恶性肿瘤补充评估，对可增加患者向妇科肿瘤学家转诊的测试进行认证。

预期上市这种通用型器械的制造商必须：

• 符合联邦食品、药品和化妆品法案（FD&C法案），包括21 CFR 807 E分部分中所述，上市前通告要求。

• 通过处理本指南中识别的卵巢附件肿块评估评分测试系统相关具体健康风险，符合针对本器械开发的特殊控制，以及

• 在器械上市前获取食品药品监督管理局的实质等同性测定。（21 CFR 807.81、21 CFR 807.87和21 CFR 807.100）。

食品药品监督管理局认为，当与法规的一般控制相结合时，特殊控制足以提供这些器械安全性和有效性的合理保证。

本特殊控制指导性文件确定了卵巢附件肿块评估评分测试系统的分类条例和产品代码（请参见第III章范围）。本指导性文件的其他章节向制造商提供了

关于处理这些器械相关风险的建议。

**III.范围**

本文件的范围限制于21 CFR 866.6050中所述的以下器械（产品代码ONX）：

21 CFR 866.6050 卵巢附件肿块评估测试系统是一种测量血清中一种或多种蛋白的器械。针对计划进行手术的女性中附件盆腔肿块为恶性的可能性，其可产生单一结果。在主要临床和放射学评估为阴性的背景下，试验仅供用于辅助，以增加对需要肿瘤学专家和资源进行妇科手术的患者的识别。

**IV.健康风险**

卵巢附件肿块评估评分测试系统未指定用作卵巢癌的筛选或诊断测试。如文中所示，未能进行试验可导致对卵巢恶性肿瘤患者的评估不当和管理错误。特别是，虚假的低卵巢附件肿块评分可导致测定患者可能没有卵巢恶性肿瘤，造成手术专家和资源达不到最佳。虚假的高卵巢附件肿块评分可导致测定患者可能有卵巢恶性肿瘤，造成手术决定不当和患者不必要的焦虑。试验的标签外使用（如，在已确定为需要手术治疗盆腔肿块，或无患者独立临床/放射学评估参考的患者中），可能导致虚假阳性结果造成的较高频率非必需进一步测试和手术，或导致虚假阴性结果造成的肿瘤诊断延迟。

下表中，食品药品监督管理局已识别到使用通常与本器械相关的健康风险。本指导性文件中描述了推荐用于减轻已识别风险的措施，如下表所示。贵公司应在提交贵公司的上市前通告之前进行风险分析，以识别针对于贵公司的器械的其他任何风险。风险可极度依赖于所使用的检测和测量方法。上市前通告应描述风险分析方法。如果贵公司选择使用替代方法处理本文件中的已识别风险，或识别到本文件中所述风险以外的风险，贵公司应提供充分的详细信息以支持贵公司所使用的处理风险的方法。

|  |  |
| --- | --- |
| **已识别风险** | **推荐****缓解措施** |
| 假阴性结果 | 第VI-IX章 |
| 假阳性结果 | 第IX章 |
| 作为筛选测试、独立诊断性测试或作为确定是否进行手术的测试时的标签外使用。 | 第IX章黑框警告 |

**V. 器械描述**

A. 背景

我们推荐贵公司使用上文第III章所述的条例的产品代码标识贵公司的器械。贵公司必须标识一个合法的上市等效比较产品（21 CFR 807.87（f））。贵公司应在表格中列出等效比较产品和贵公司的器械之间的相似性和差异性，从而，食品药品监督管理局能够有效确定贵公司的器械和等效比较产品之间是否可进行比较，并识别出贵公司的器械可能需要或可能不需要其他性能研究的方面。

贵公司的测试中从多种分析物获取的值源自每个分析物的单独试验。如果贵公司的测试使用了从其他制造商出售的单独试验（需要材料但未提供）中获取的值，贵公司应指明测试的分类，试验是否已由食品药品监督管理局认证或批准（即，在510（k）或PMA过程中分别作为II类或III类器械进行正式审查），以及其已被认证或批准的预期用途。

其他制造商出售的单独试验（即，贵公司已作为获取试验结果所需材料列出，但由其他制造商分别出售的物品）可视为贵公司器械的组件。对于已认证或批准的试验，贵公司应提供一份食品药品监督管理局审查后贴标的副本，并汇总每次试验的分析性能特征。如果其他制造商出售的试验尚未经认证或批准（如，贴标为I类器械豁免），贵公司也应提供申办者的标签，汇总本章要求的信息。然而，贵公司将需要分别进行并提交针对这些试验的分析性能研究（更多信息请参见第VI章，分析性能验证），并将其作为贵公司II类器械的组件进行审查。

贵公司的提交物应充分描述贵公司的卵巢附件肿块评估评分测试系统的以下特征：

B. 质量体系法规（QS Reg）

体外诊断测试的所有制造商均必须遵守21 CFR第820部分，质量体系法规的要求。如果贵公司已设计了算法用于从其他制造商处购买的测试，贵公司应对这些试验负责，

这些实验为贵公司器械的组件，无论食品药品监督管理局此前是否已认证了这些试验。贵公司的责任包括但不限于：器械设计、器械、组件购买和操作、生产和过程控制、包装和贴标控制、器械评估、安装、投诉处理、维修服务和记录。贵公司应为贵公司的器械（包括所有组件）提供一份符合质量体系法规的声明，以及一份贵公司已如何或将如何在贵公司的测试上市前完成本要求的简述（即，通过与其他组件制造商建立关系）。由于建立与质量体系法规的依从性并非510（k）上市前通告审查的一部分，食品药品监督管理局从这些试验类型的制造商处寻求保证，保证其在认证前充分承担起本测试所有方面的责任。

C. 预期用途/适应症

贵公司的提交必须在使用声明中包括预期用途/适应症，汇总贵公司（制造商）预期产品的使用方法，以及测试的临床目的（21 CFR 807.87（e）和21 CFR 807.92（a）（5））。卵巢附件评估评分测试系统的预期用途应包括测试名称、测试结果输出的特性描述、标本类型、测试方法、目标人群（可能受研究人群的人口统计学限制，如，年龄、人种和月经状态及疾病分期等临床状况）。卵巢附件评估评分测试系统旨在用于由卵巢/附件肿块、计划进行手术但尚未转诊至妇科肿瘤学家的女性，在医师独立临床和放射学评估不能指明恶性肿瘤时，作为一种进一步评估存在恶性肿瘤可能性的辅助方法。

D. 测试的理论依据

在试验中提供一份测试分析物和其后理论依据的汇总，包括每种分析物的测量值。可提交已公布的信息支持用于测试适应症的单独分析物。纳入任何导致做出纳入单独测试分析物结论的任何未发表研究汇总。贵公司的描述应包括关于不同RNA剪切或转录后修饰造成的不同蛋白状态的相关信息，以及在所调查的疾病人群中，与无疾病人群相比，这些状态如何不同。

E. 测试组件和方法学

贵公司应详细描述贵公司在器械中所用的、试验格式/方法学、仪器和软件。

1.*测试试剂*

贵公司应为提供货推荐使用的所有试剂和组件提供一份说明（包括校准器、控制装置、仪器）。包括对每种试剂来源的描述（如，小鼠、细胞系）、其纯度方法和测试中使用的验证过程。如果试剂从外部供应商处获取，贵公司应提供分析证书。如果在贵公司的测试中，试剂盒组件包含其他制造商供应的单独测试试剂盒，则我们认为贵公司的测试中这些测试试剂盒试剂，以及与贵公司的测试一同使用的具体制造商测试试剂盒和仪器，均应在标签中说明。本信息要求有一份单独试验的表格格式。

*2. 测试方法学*

贵公司应提供一份对贵公司器械中所用试验方法学的描述。应包括试验平台和测量方法。要求有一份汇总试验中所包含所有测定中本信息的简表。对于包含未曾审查并经食品药品监督管理局认证或批准的新分析物或免疫分析的器械，要求有额外信息：

• 贵公司器械所需的仪器仪表和仪器仪表软件，包括系统内的组件及其功能。

• 解释贵公司如何筛选、选择和确定最佳抗体-分析物-抗体组合以进行测试中目标分析物的捕获和检测。

• 用于将捕获抗体附到基质上的方法。包括关于怎样对本方法进行最佳化的描述（如，最大抗体和抗原浓度）

• 二抗共轭物的描述包括标签和共轭步骤。

• 反应组分和条件、洗脱步骤和信号检测组分及方法，以及如何对这些条件进行最优化的描述，包括抗体浓度。

• 测试中如何尽量对交叉反应和非特异性结合进行最小化的描述。提供免疫印迹作为证据。

• 测试中从采集至使用期间的样本采集要求和样本处理，包括用于确保分析物稳定性的任何要求（如，防腐剂）。

• 所有控制和校准器，以及其如何在系统中起作用的说明。

• 提供一份如何对背景信号进行最小化或标准化的描述（如，信噪比）。

***3.评分算法***

贵公司应提供

• 一份培训数据集的描述（入选/排除标准、临床中心描述、受试者数、恶性肿瘤患病率等）；

• 一份分类法/算法开发的描述，包括样本类型和使用的统计模型及技术。

• 一份性能测量描述（内部验证）。

如果评分临界点在培训数据集中选择，贵公司应提供关于如何确定临界点的信息。

贵公司应提供一份最终计算方法的简介，说明测量间期外的单独分析物值在评分值计算中的应用，以及用于获取评分结果的软件。

如果用于评分测试结果计算的算法使用贵公司测试中单独分析物的临床决策点，贵公司应提供关于这些临界点的信息。例如，贵公司的试验可能整合了一种分析物，其通常用作诊断炎症的辅助方法。在有良性和恶性盆腔肿块的女性中，本分析物的推荐临界点可能不同。贵公司应解释是否使用了已知参考间期，或是否从文件中获得了数值，或从贵公司的调查研究中获得了结果。

F. 测试结果

贵公司应说明对测试结果输出（如，应用临界点的患者分类或连续数值）和评分结果解释的描述。贵公司应提供临床医师生成的试验报告范例（如，打印输出）。

**VI.分析性能验证**

在贵公司的510（k）中，贵公司应详细描述贵公司用于评估以下所列每种性能特征的研究设计。应使用贵公司卵巢附件评估评分测试系统器械的最终版本进行所有分析性能研究。

对于以下描述的每项分析性能研究，贵公司均应说明贵公司为每种分析物单独预先确定的验收标准，以及对总体结果的影响。

如果贵公司的测试使用了从其他制造商出售的单独定量试验（需要材料但未提供）中获得的值，且这些试验经食品药品监督管理局认证或审查，贵公司可能无需为这些试验单独提交除精度以外的额外性能研究，

如下文所述，前提是卵巢附件肿块评估评分测试系统基于这些测试的食品药品监督管理局审查贴标中所述性能特征确定。例如，如果贵公司的测试未整合除单独试验已声明测量间期以外的单独定量值，贵公司将无需进行额外的线性度研究。食品药品监督管理局已审查的测试（如I类豁免）应证明其作为贵公司II类器械组件的分析性能。重要的是应注意这些单独I类豁免试验在作为II类试验时不接受单独认证，也不得视为已审查其I类豁免预期用途。这些试验的分析性能研究证明，这些试验适合在卵巢附件肿块评估评分测试系统中使用的目的。

食品药品监督管理局预期将试验认证的范围限制于经评估的具体试验试剂盒和仪器。

1. 标本

*分析前因素：*如果贵公司的测试包含新型分析物（如，此前从未在体外实验室测试中涉及的分析物），贵公司应指明是否需要血液采集（如，采血期间患者的姿势）和血液采集管处理的具体说明（由于分析物的性质不稳定）。

*稳定性：*贵公司应证明贵公司测试中所用这些参数极端值（如，温度、冷冻时间、冻融和运输）之间的标本稳定性。贵公司应描述贵公司如何为每种分析物选择验收标准。贵公司试验的标本稳定性要求限制于贵公司评分试验系统中稳定性最低的分析物。

B. 可重复性/再现性

贵公司应根据可变性对测试评分的整体可接受影响，提供一份贵公司如何测试和确定单独分析物可接受可变性的描述。

我们推荐贵公司根据扩大评分测试结果范围的样本提供一份贵公司评分测试系统的精密度评估以及样本。CLSI文件EP5-A2 “定量测量方法的精密度性能评估；已批准指南”和EP12-A2“定性测试性能评估的用户方案；已批准指南”包含的指南可能有助于开发精密度研究中数据的设计和计算。

精密度研究中的样本应扩大评分数值的范围；贵公司应纳入一些由于与分析物的不同结合导致评分值接近评分测试临界点的样本（3-5个样本）。

理想情况下，贵公司应识别评分试验可变性的所有来源，并在精度研究中将其纳入。贵公司应提供一份3个实验室中贵公司评分测试的精确度证明，并提供一份可重复性（运行内精确度）、运行之间、不同日之间、操作者之间和研究中心之间组分不精确度的评估。还需纳入一份研究中评估的天数、操作者数、试验、仪器、批次和校准周期的详细描述。

贵公司应提供验收标准，并使用跨度测试评分结果范围的样本证明测试评分的精确度。贵公司也应纳入每种分析物的精密度研究数据，以证明每种分析物的精密度均符合预定义的验收标准。贵公司应指明贵公司如何得出结论称容许的分析物不精密度不会减弱所报告指数值的准确性。贵公司还应使用3个不同测试批次提供一份批次对批次精确度的评估。这包括组成贵公司试验系统的每种单独测定、校准器和对照的多个批次。

除了上述对测试评分系统的精密度研究，贵公司还应根据每种单独分析物的精密度特征，提交一份测试评分系统精密度的可能结果模拟。对于在上述精密度研究样本中所展示的分析物，通常的精密度研究会提供关于一些特定单独分析物用量组合的精密度。然而，有很多单独分析物用量的可能组合可给出相同的测试评分值，但精密度不同。承认不可能评估出所有可能分析物组合的精密度，其他模拟可提供关于测试评分系统可能精密度特征的单独分析物值不同组合的信息。如果模拟预测到在临床决策点有不可接受的精确度水平，评估反映该特定情景下的人为样本可能很重要。

在此展示了一种贵公司可能选择使用的此类模拟方法范例。应使用此前进行的精密度研究中的精密度数据（对于经认证/经批准的分析物和提供用于新型分析物的数据）构建每种单独分析物的精密度特征。应使用来自实际样本的可重复性和实验室内精密度研究的已知精密度数据，通过进行线性插入，构建单独分析物的可重复性精密度特征以及实验室内精密度的精密度特征。多于单独分析物值的每个可能组合，均应评估对应于本分析物值组合的评分值，以及基于相应精密度特征的评分可重复性可实验室内精密度。由于评分基于样本中单独分析物的分别测量，每种分析物的随机测量误差可视为不相关。额外统计模拟的基本步骤如下（为简单起见，考虑2种单独分析物X1和X2，评分= F（X1，X2）以及可重复性精确度数据）：

1） 提供既往所进行精密度研究中的可重复性精密度数据（平均值、标准差（SD）和百分变异系数（%CV））。

2） 考虑2种分析物的组合，值为X1=U和X2=V。使用可重复性精密度特征，获取X1=U的SD1（U）和X2=V的SD2（V）。

3） 使用U的平均值和标准差SD1（U）的正态分布生成X1\*，使用V的平均值和标准差SD2（V）的正态分布生成X2\*。计算评分\*=F（X1\*，X2\*）。执行本步骤K次（例如，100）后，计算K测量值的评分平均值Score\*mean（评分\*平均值）（对应于X1=U和X2=V的评分平均值），以及K评分测量值的标准差SD和%CV。

4） 提供评分的可重复性精密度特征：来自所有可能U和V组合且精密度特征可用的既往步骤中的平均评分值Score\*mean和SD及%CV。可重复性精密度特征应以表格形式（EXCEL）提供，并以图表表示（X轴为平均值Score\*mean，Y轴为对应的%CV）。

使用单独分析物的实验室内精密度特征，对实验室内精密度特征进行相似的统计学m模拟，获得Score（评分）。

C. 单独分析物的线性度

每种单独分析物的线性度证明基于整合进算法中的测量范围确定。对于使用单独免疫分析的测试评分系统、延伸至超过已认证/已批准分析物测试范围的测量范围要求，贵公司应提供一份新的线性度证明。未认证/未批准测试分析物也应证明线性度。贵公司应指明，如果贵公司的患者人群中获得的值很可能落于单独分析试验的声明测量范围以下，贵公司如何控制超出范围的结果对贵公司评分的影响。我们推荐贵公司参考CLSI文件EP6-A “定量测量程序的线性度评估：一种统计方法；已批准指南”，以获取关于进行单独分析物线性度研究的更多信息。

***评分值范围***

贵公司应根据单独分析物的测量范围，提供关于测试评分系统数值评分值的信息。如果，除提供基于临界点定性测试评分结果外，贵公司还计划报告测试评分的数值，数据应证明测试评分系统的高数值与恶性肿瘤可能性的逐步升高或逐步降低相关。

D. 低水平时的性能

贵公司应证明任何未认证/未批准分析物测试的检测限和定量限，以及可增加已认证/已批准分析物测试下端测量范围的任何变化的测试限和定量限。贵公司应解释如何将与检测限和/或定量限相关的低水平值整合进算法中，

从而不会输入测量间期外的结果，且不产生测试结果。我们建议贵公司参考CLSI文件EP17-A “检测限和定量限测定方案；已批准指南”获取关于对每种分析物进行检测限和定量限研究的更多信息。

E. 干扰

干扰贵公司测试中任何分析物的物质很可能会干扰测试结果。贵公司应指明（最好以表格形式说明所评估的浓度）是否有任何分析物会受到血红蛋白、胆红素（结合和非结合）、甘油三酯、总蛋白、嗜异性抗体（HAMA）以及类风湿因子的干扰。对于上述干扰物，贵公司应通过比较有干扰物样本和相同的无干扰物样本对总体指数的影响，以及95%置信区间，证明试验结果的百分差异。报告本分析中每种分析物的百分差异也会有帮助。理想情况下，所评估的分析物浓度应接近测试评分系统的临床决策点。

贵公司应指明对于贵公司测试中的分析物，是否有任何已知干扰物来源出现，如有，证明该干扰物对评分的影响。

贵公司应证明常见药物不会干扰测试。

F. 交叉反应/非特异性结合

应对已知交叉反应物及其对测试评分系统的潜在影响进行评估。测试可受一些因素影响，如抗体固定值底物导致的配体结合影响、蛋白的非特异性吸收以及矩阵中其他蛋白的影响。如果贵公司的测试为多通道免疫分析，贵公司应证明分析物之间不会出现交叉反应、非特异性结合和交叉干扰。指明贵公司的测试中是否有任何潜在的分析物交叉反应。贵公司还应提供证明：使用贵公司的抗体进行的分析物检测为特异性。免疫印迹应可提供本证明。

G. 单独分析物的钩状效应

如适用，贵公司应证明过量分析物不会造成钩状（前带）效应。应对贵公司的测试中未认证/批准的单独试验（组件）以及评分测试系统中的每种分析物进行本证明。

H. 遗留污染

针对使用既往经认证仪器仪表的测试系统，贵公司应提供对遗留污染可能性的描述。

1. 矩阵比较

对于相同分析物，当使用可导致测试性能变化的不同抗凝剂测试血浆样本时，可能出现矩阵效应。如果贵公司的测试推荐了一种以上样本类型，贵公司应评估矩阵对测试产生影响的可能性。应展示测试评分中每种单独分析物的矩阵效应影响及其对总体评分结果的影响。

J. 稳定性

贵公司应针对确定试剂盒仪器的实时稳定性，以及开放药瓶和运输稳定性（如适用）方面描述贵公司的研究设计。贵公司的稳定性研究应包括关于贵公司的试验系统和试剂时间、温度和储存情况的信息。对于每项研究，贵公司均应提供贵公司的验收标准和一份贵公司如何选择验收标准值的描述（即，决定不影响结果的验收标准界限）。我们推荐贵公司参考CLSI文件EP25-A “体外诊断方法产品的稳定性评估；已批准指南”获取关于对单独分析物进行稳定性研究的更多信息。

K. 校准和对照

对于组件中包括已认证/已批准单独免疫分析的卵巢附件肿块评估评分系统，应基于在试验中的使用对每种试验的校准器和控制进行描述。

对于所有其他测试，仅应描述贵公司对照和校准材料的以下方面：

• 贵公司在贵公司的系统中纳入、或推荐使用的各种对照的

性质和功能。

• 控制和校准材料的值分配及验证方法。对于贵公司的测试系统所整合的任何试剂，如果供应商提供了分析证书，应纳入其中。

• 可用于测试仪器仪表故障以符合所要求规范的控制参数。

**VII.软件**

贵公司应按照关注程度，提供贵公司器械中所用软件的详细信息。额外信息请参见食品药品监督管理局文件“医疗器械中所含软件的上市前提交内容

指南”1。贵公司应在危险减轻前确定关注程度。本类型的体外诊断器械通常被视为中等水平顾虑，因为软件缺陷可导致将假结果报告给临床医师和患者，从而对患者造成伤害。

贵公司制备用于食品药品监督管理局审查的软件文件时，应包含以下几点，如适用：

* 软件设计的完整描述。贵公司的软件所包含的实用程序不得特别设计用于支持超出贵公司预期用途。在贵公司的设计中，贵公司还应考虑保密性与安全性问题。关于这些问题中某些问题的信息可参见健康保险携带和责任法案（HIPAA）相关的以下网址http://www.hhs.gov/ocr/privacy/hipaa/understanding/index.html.
* 基于器械设计和子系统组件任何故障相关关键考虑的危险分析，如患者报告错误、仪器故障及操作者安全性相关的信号检测和分析、数据储存、系统通信以及网络安全。
* 将提交以证明实质等同性的软件版本完整验证与确认（V&V）活动文件。贵公司还应提交关于试验软件与任何仪器仪表软件兼容性验证的信息。

如果贵公司在510（k）中纳入的信息基于所发布版本以外的版本确定，请标识510（k）版本的所有差异以及这些差异（包括任何未解决异常）如何影响器械安全性和有效性的详细信息。

以下为帮助贵公司在软件生命周期质量管理规范和食品药品监督管理局条例下开发和维护贵公司器械的其他参考文件。

软件验证的一般原则；行业和食品药品监督管理局工作人员的指南终稿；可在以下食品药品监督管理局网站获取：

http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/Guid anceDocuments/ucm085371http://www.fda.gov/downloads/MedicalDevices/DeviceR egulationandGuidance/GuidanceDocuments/ucm085371.pdf.

医疗器械中现用软件使用指南；终稿；可在食品药品监督管理局网站获取。2

21 CFR 820.30 C分部分 – 质量系统条例的设计控制。

1 http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm073779.pdf

http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm073779.pdf

ISO 14971-1；医疗器械 - 风险管理 - 第1部分：风险分析的应用。

AAMI SW68:2001；医疗器械软件 - 软件生命周期过程。

**VIII. 临床性能评估**

来自贵公司临床研究的数据应支持贵公司器械的适应症和声明。临床验证研究应使用预期用途人群中获取且与培训集（贵公司用于开发算法的标本（如，评分））不同的患者样本。贵公司应描述每项临床研究的方案，包括入选和排除标准、研究设计、统计分析方法和样本量的统计学理由。贵公司应提交单独分析物值的数据，以及来自贵公司临床验证研究的评分测试结果。

A. 研究人群/样本

卵巢附件肿块评估评分测试系统的预期用途人群包括有已知需要进行手术的盆腔肿块且在基层护理背景（即，妇科医师、内科医师、家庭医生但非妇科肿瘤学家）下已进行评估的患者。卵巢附件肿块评估评分测试系统与术前临床病理学信息向结合，可增加对需要肿瘤学专家和资源进行妇科手术的患者的识别。贵公司应提供贵公司的入选和排除标准。患者应为预期使用人群的代表。食品药品监督管理局推荐贵公司入组美国人群中来自几个独特地理位置的患者。培训数据集中使用的样本不得作为贵公司验证数据集的一部分纳入。

贵公司应计划在绝经前和绝经后女性中分别以及合并评估贵公司的结果。贵公司应指明将如何识别月经状态。如果自我识别，贵公司应提供有定义的患者，并对未提供答案的病例提供计划。最好是对定义可已知应用于患者人群的月经状态有目标方法。例如，年龄、末次月经时间或促卵泡激素（FSH）的血清水平。

对于所有样本，贵公司均应提供总体和按入组中心划分的汇总患者信息（年龄、人种/种族、月经状态、当前医学状况）。

贵公司应提供研究中所有样本数的理由。提供已排除样本的详细叙述和排除的具体原因。不得根据术后结果排除样本。

由于前瞻性临床研究中采集的样本优先，标本库中表征明确的回顾性样本可用于贵公司的临床验证研究，前提是除以上列出的条件外，还应符合下列条件：

• 贵公司研究中的标本来自贵公司预期用途人群的代表患者。

• 无选择方法导致的偏倚（即，样本采购应包含时间和研究中心的考虑）。

• 无人为分析影响造成的偏倚（即，由于储存条件、多次冻融）。

• 库中检索的标本符合样本采集方案的预定义标准。

• 样本注释中含以下信息：患者人口统计学（年龄、月经状态）、非符合肿瘤学家进行的术前评估（恶性相比于良性）和手术病理学（组织学诊断，以及如为恶性肿瘤，肿瘤分期）。

我们推荐贵公司在使用库存样本进行验证研究前咨询食品药品监督管理局。

1. 临界/临床决策点

贵公司的提交材料中，贵公司应解释如何确定临界点（用于区分存在恶性肿瘤和无恶性肿瘤的值）。根据使用临床样本的培训和/或试点研究的接受者操作曲线（ROC）分析，可通过敏感度和特异性的相关水平进行调整，选择适当的临床临界点。所选临床临界点时的临床性能可使用关键临床研究（验证数据集）轻易进行估计。一些情况下，临床临界点可使用无偏倚的程序和适当样本量，在关键临床研究期间确定。如果临床可接受的灵敏度（或特异性）水平为预先设定，则可使用关键研究确立对应于预先设定敏感度水平的临床临界点，并获取评分测试及其所选临界点的临床性能的无偏倚评估（参考2）。如果测试的结果范围推荐再次测试，或不能得出“阳性或阴性”结果的测定（即，模糊区），贵公司应解释贵公司如何确定模糊区界限。对于其样本得出模糊结果的患者，贵公司还应判断临床意义。

如果贵公司的测试对所有患者使用一个临界点，贵公司应证明对于绝经前和绝经后女性，该临界点均适当。我们推荐贵公司在贵公司的试点（或培训）研究中调查针对月经状态选择临界点的需要。如果月经状态与测试结果的解释相关，贵公司应计划充分代表不同月经状态组，并根据月经状态验证临界点。

1. 临床参考标准（“金标准”）

为评估卵巢附件肿块评估评分测试系统在区别良性或恶性附件肿块方面的性能，应将测试结果与以下手术中获得的组织病理学信息进行比较。对于每例患者/样本，贵公司均应指明肿块为良性或恶性，为恶性时指明病理学诊断（即，上皮性卵巢癌、其他原发性卵巢恶性肿瘤、低恶性潜能（LMP）的卵巢恶性肿瘤、累及卵巢的非原发性卵巢恶性肿瘤、或未累及卵巢的非原发性卵巢恶性肿瘤）、肿瘤分期和组织学（浆液性、粘液性、子宫内膜样等）。病理学结果的分类有2种类别（即，恶性和非恶性），应综合考虑并预先设定。贵公司应说明最终执行手术的人员（妇科肿瘤学家（GO）或符合肿瘤学家以外的医师（非GO））。

D. 研究设计

因为卵巢附件质量评估测试系统与选择用于手术的卵巢肿块患者的临床评估结合使用，但尚未提及GO，因此必须对术前临床进行组织良好的完整记录非GO的评估。对于临床研究中的每例患者，应采集术前和转诊前临床病理学信息（如，患者症状、体检结果、图像、CA 125值），并整合进将肿块识别为“良性”或“恶性”的二进制术前评估中。评分测试的评估作为除术前信息外的患者评估，应在考虑由非GO进行的术前临床评估和贵公司测试的结果后完成。这称为“OR”（或）决策规则，即，如果非GO的术前评估为阳性，*或*如果测试结果发现其为阳性，则将病例视为阳性。使用本规则，如果患者的评分测试结果为阴性，术前信息为阳性，则患者应被视为有较高卵巢恶性肿瘤可能性。类似地，如果一例女性的评分测试结果为阳性，术前评估为阴性，则该女性应被视为有较高卵巢恶性肿瘤可能性。使用本方式旨在改善患有恶性盆腔肿块的患者向妇科肿瘤学家的转诊，同时仍确保所有根据单独术前信息也会被转诊的女性的转诊（参考3）。

未标明评分测试提供了超出单独术前评估的额外信息，贵公司应比较单独术前评估的性能和使用“OR”规则结合的术前评估和评分测试性能。阳性预测值（PPV）和阴性预测值（NPV）（或相等的，炎性和阴性似然比）为比较的基础。其他信息请参见Biggerstaff.3

贵公司应证明贵公司的测试为生物学相关性亚组人群（如，绝经前、绝经后）提供了额外信息或提供了为何不需要此类测定的可接受理由。

3 Biggerstaff，B.J. 比较诊断测试：使用似然比的简单图形。医学统计2000，19：649-663。

考虑测试T1和测试T1及T2组合OR的一般方案。“OR”组合的灵敏度至少与单独T1的灵敏度同样大。“OR”组合的特异性与单独T1的特异性相同或较差。因此，组合“OR”在灵敏度和特异性之间存在固有的权衡。如果结合的特异性显示明显降低，结合单独灵敏度的升高不能证明T1和T2组合（OR）有效。

临床研究的数据应表明

• 组合OR（术前评估和评分测试）的NPV与单独术前评估的NPV相比，有统计学和临床显著的改善；且

• 如果术前评估和评分试验组合OR的PPV与单独术前评估相比有损失，则PPV中的该损失应在临床上可接受。

这些成功标准的逻辑开发如下。考虑在灵敏度和1-特异性绘图上连接点（0，0）和测试T1相应点的直线（参见下图1）。本线条标识与测试T1有相同PPV的测试的性能特征。连接点（1，1）和测试T1相应点的直线表示与测试T1有相同4 NPV的测试性能特征。比较测试T1的性能和测试T1及T2OR组合的性能，有3种可能情景：

情景A：“OR”组合（T1或T2）的2种预测值（PPV和NPV）均大于单独T1的预测值（参见绿色区域）。因此，可以轻易测出结论，组合OR与单独T1相比较好。

4 PPV取决于阳性似然比（PLR）和恶性肿瘤患病率，NPV取决于阴性似然比（NLR）和患病率。对于相同人群中的2种测试比较，PPV和NPV的比较等效于PLR和NLR的比较。



1-特异性

图1

情景B：组合OR的PPV与T1的PPV相比较差，但组合OR的NPV与T1的NPV相比较好（参见蓝色区域）。本区域中，NPV增加量和PPV减少量之间有权衡。成功可总结为，下降后的PPV仍与测试的安全和有效使用一致。

情景C：组合OR的PPV和NPV与测试1的PPV和NPV相比均较差（参见红色区域）。因此，可以轻易测出结论，组合OR与单独T1相比较差。

我们建议贵公司根据病理学结果在与下表类似的表格中汇总临床研究结果，并根据下表所示的不同结果，提供恶性肿瘤可能性的评估（以及95% CI）：

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **非GO术前****评估** |  |
|  |  | **阳性** | **阴性** |  |
|  |  | A | B |  |
| 评分 |  | C | D |  |
| 测试 |  |  |  | N |

对于通过病理学确定恶性肿瘤的受试者以及通过病理学确定无恶性肿瘤的受试者，分别展示评分测试性能和术前评估。



病理学恶性肿瘤

非GO术前评估

阳性

A1

C1

评分

测试

阴性

B1

D1

阳性

阴性

N1

病理学恶性肿瘤

非GO术前评估



阳性

A0

C0

阴性

B0

D0

阳性

阴性

评分

测试

N0

下表显示了应用于所有经非GO医师评估受试者的测试的性能特征。对于单独评估，仅使用术前评估，不参考评分测试结果。对于双重评估（即，“OR”组合），如果术前临床评估、评分测试为阳性或二者均为阳性，则称附件肿瘤为潜在恶性肿瘤。

性能

单次评估

（术前评估）

双重评估

（术前评估“OR”（或）评分测试）



灵敏度

特异性

PPV

NPV

患病率

* 提供单独术前评估的灵敏度、特异性、阳性预测值（PPV）、阴性预测值（NPV）以及95%置信区间；
* 使用决策规则“OR”（或）提供评分测试性能结合术前评估的灵敏度、特异性、阳性预测值（PPV）、阴性预测值（NPV）以及95%置信区间。
* 计算NPV差异和PPV差异以及双侧95%置信区间（可使用引导技术进行置信区间计算）。NPV的改善因为统计学和临床显著，且，如果观察到PPV损失，贵公司应证明本损失的临床可接受性。

此外，贵公司应在下表中展示从非GO评估患者中观察到的不同术前评估和评分测试结果中恶性肿瘤的频率，以及95%置信区间：

恶性肿瘤的频率

95% CI

经非GO医师评估的有附件肿块患者中恶性肿瘤的患病率：

术前单独评估“阳性”

术前单独评估“阴性”

单独评分测试 “阳性”

单独评分测试 “阴性”

术前评估“阳性”和

评分“阳性”

术前评估“阳性”和

评分“阴性”

术前评估“阴性”和

评分“阳性”

术前评估“阴性”和
评分“阴性”

相同信息应展示为似然比及其95% CI，按照以上阐明的恶性肿瘤频率进行制表。似然比（结果）= Pr（结果|恶性肿瘤）/ Pr（结果|无恶性肿瘤）。似然比与预测值不同，其不依赖于疾病的患病率。

*亚组分析：*贵公司应证明绝经前和绝经后患者分别分析中，类似于总体人群所述的分析，双重评估的NPV与单独评估的NPV相比有统计学和临床线束改善。

*其他信息：*贵公司应提供在按照肿瘤分期或组织病理学所见进行分组的患者中，贵公司评分测试的描述性统计学制表。

***妇科肿瘤学家评估的患者人群中的结果***

卵巢附件肿块评估评分测试系统预期用于将进行手术的有盆腔肿块女性。测试适用于作为确定转诊决策的辅助工具。贵公司的临床研究应避免对可能由非GO研究中心选择性入组的患者进行评估而导致的可能偏倚。（例如，一些非GO



医师可能因各种原因自动将患者转诊至GO，而不考虑术前评估。在此类研究中心贵公司的临床试验入组将导致有潜在偏倚的患者代表。）贵公司可选择提供GO评估患者的额外信息，以证明贵公司的测试性能中不会出现阳性偏倚。本数据经食品药品监督管理局审查，预期在GO评估组中性能不会减弱。

E. 其他良性和恶性疾病下的预期值

目标人群可能有与癌症无关的各种疾病，但此时显示尚未识别到卵巢肿块。这些其他疾病（在一些情况下已说明分析物的真实测定值[如，为测试组件的免疫分析]）可戏剧性地影响评分测试结果并混淆其解释。贵公司应在单独分析物试验显示有疾病状况的患者中，以及可能并发出现良性和恶性疾病的患者中，证明贵公司的测试评分结果。这些疾病的示例有：子宫内膜异位、盆腔炎性疾病、糖尿病、贫血、自身免疫疾病如克罗恩病、SLE和类风湿性关节炎、心脏病、肝炎、肾病和营养不良，以及各种癌症如宫颈癌、肺癌、乳腺癌和结直肠癌。

F. 参考间期

可提供在表明健康女性中的参考值，但此类女性不能是评分测试预期用途人群的一部分。对于每种分析物和评分测试结果，任何分析值均应包括跨越贵公司测试年龄范围的女性，且应评估至少120例绝经前女性和120例绝经后女性，除非贵公司能够证明2个人群之间无任何差异。如可能，除高加索人和非裔美国人外，贵公司还应纳入其他种族（拉丁美洲人和亚洲人）。贵公司应提供每例女性的评分并调查评分和年龄的关系。

G. 评分中所含单独分析物的相关性

贵公司应证明评分测试使用中每种分析物纳入的合理性。一个选择是使用来自临床研究的数据，证明评分计算中纳入的单独分析物可为卵巢恶性肿瘤提供有用信息。对此，应进行ROC分析：对于每种单独分析物，展示单独分析物的ROC曲线，并计算单独分析物ROC曲线下的面积以及置信区间（应正确处理多样性问题）。此外，对于每种单独分析物，展示单独分析物的ROC曲线和相同图表上评分的ROC曲线。如果临床研究的数据未证明相同的单独分析物可为卵巢恶性肿瘤提供有用信息，贵公司应证明为什么将这些分析物纳入评分测试的计算中。



**IX.标签**

上市前通告必须包括足够详细的贴标，以满足21 CFR 807.87（e）的要求。在体外诊断器械引入州际贸易前，体外诊断器械的最终贴标必须也符合21 CFR 809.10的要求。下表所列的内容不包括标签要求的所有元素，但其旨在辅助贵公司制备可满足这些要求的标签。

预期用途

预期用途应指明测试手段是什么，将使用测试的临床适用症是什么，以及测试预期的特定人群是什么（如适用）。预期用途应指明测试是否为定性或定量。

黑盒警告

考虑到卵巢癌测试方面历史和当前未满足的医疗需求，食品药品监督管理局得出结论，本器械的标签外使用有风险。为解决本风险，制造商应在卵巢附件肿块评估评分测试系统的贴标、广告宣传和推广材料中提供关于标签为使用风险的通知。制造商必须强调以下风险：

* 无附件盆腔肿块的女性（即，癌症“筛选”）不是卵巢附件肿块评估评分测试系统预期用途人群的一部分。卵巢癌筛选测试的假阳性结果相关公共卫生风险在医学文献中有明确描述，并纳入了与非必须测试和手术相关的发病率和死亡率。阴性筛选结果的风险也包括测试和治疗卵巢恶性肿瘤失败导致的发病率和死亡率。
* 如果使用测试结果确定已知有卵巢附件肿块的患者中的手术需要，根据疾病患病率和类型调整后的类似风险升高。
* 如果使用除本特殊控制指南中所述“OR”规则以外的规则，来自卵巢附件肿块评估评分测试系统的结果将造成因未转诊进行肿瘤学评估和治疗导致的发病率和死亡率。

为解决标签外使用的风险，卵巢附件肿块评估评分测试系统的标签、广告宣传和推广材料均应包含使用以下模板的文字或等效文字的注意事项框：

注意事项：[测试名称]不得在无独立临床/放射学评估时使用，且预期不可用作筛选测试或用作确定患者是否应继续进行手术。错误使用[名称]可带来进行非必须测试、手术和/或诊断延迟的风险。

**测试原则**

贵公司应描述本类型器械中使用的测试组件（特异性试验、校准器和仪器）或测试方法学。

**警告和注意事项**

贵公司应纳入所有针对贵公司测试的警告和注意事项，包括影响样本的条件、贵公司测试的组件中任何其他适用制造商说明书中指定的条件、以及潜在实验室危险。

标本和试剂 ─ 稳定性和储存

贵公司应说明贵公司测试所用的样本矩阵、样本处理的说明以及稳定性信息（包括储存和温度）。如果贵公司的测试系统包括单独免疫分析，样本稳定性和储存声明应受限于贵公司测试系统中最不稳定组件（试验）的性能声明，除非贵公司已提供验证数据证明。此外贵公司还应汇总每种单独分析物的储存和稳定性数据。

**测试组件**

贵公司应提供贵公司测试系统中所需要的具体试验列表，包括校准器和控制。对于这些与贵公司测试相关试验的性能，贵公司应为用户提供任何期望的汇总，包括但不限于测量范围、测量单位和质量控制措施。

**程序**

本章节应包括程序从样本处理到结果报告的明确和简明的说明。应提供具体且充分的说明，包括软件安装的任何故障处理建议。用户可参见组件包装说明书；然而，也应纳入每种试验程序的一般总结。

**结果解释**

贵公司应明确定义可能的结果范围、使用的具体临界点和/或模糊区、这些临界点之间的结果意义，并解释用户在有任何模糊结果的情况下应做什么（如，重复）。贵公司应指明结果是否应根据年龄、月经状态和其他因素而进行不同解释。

**局限性**

贵公司应明确描述标签中的任何和所有局限性。本章节应包括操作者或医师需要在使用测试前了解的适当局限性。

除与贵公司测试相关的任何局限性和警告外，卵巢附件肿块评估评分测试系统还应包含声明：在阳性术前评估背景下阴性测试结果不得排除肿瘤学转诊。

**临床性能研究**

贵公司应在说明书中纳入贵公司研究中所有可评估受试者的人口统计学特征和病理学汇总。贵公司应纳入贵公司研究设计和研究中结果的汇总。本章节应纳入术前评估（单独评估）、测试（单独评估）和二者结合（双重评估）的描述性能（灵敏度、特异性、NPV、PPV和95% CI）。如适用，应包括基于月经状态、病理学和分期的结果。贵公司应根据这些研究总结贵公司的结论。

**分析性能结果**

贵公司应提供贵公司的评分测试系统中分析性能结果的汇总。贵公司提供的结果只能为为完整结果，而非单独分析物的结果。本数据应包括（如适用）精密度（可重复性/再现性）、数值测试评分结果范围、干扰、交叉反应性和矩阵比较。

**参考值和预期值**

这些章节应包括无疾病女性和有其他良性和恶性疾病的女性中贵公司测试结果的第5和第95百分位范围。信息还应包括用于测定这些值的所用人群的样本数、年龄、疾病和人口统计学。

**X. 参考文件**

1.ACOG Practice Bulletin No. 83:Management of Adnexal Masses.Obstetrics & Gynecology.110(1):201-214, July 2007.

2.Kondratovich M., Yousef WA.Evaluation of accuracy and optimal cutoff of diagnostic devices in the same study.Joint Statistical Meeting.2005.ASA Section on Statistics in Epidemiology; p.2547-2551.

3.Gostout BS, Brewer MA.Guidelines for Referral of the Patient with an Adnexal Mass. 2006.Clinical Obstetrics and Gynecology 49(3):448-459.

