临床性能评估：用于放射学图像和放射学器械数据的计算机辅助检测设备的注意事项 - 上市前通知510（k）申请

行业和美国食品药品监督管理局
工作人员指南

**文件发布日期：2020年1月22日**

**文件最初发布日期：2012年7月3日**

如对本指南文件有任何疑问，请致电301-796-2563或发送电子邮件至[Nicholas.Petrick@fda.hhs.gov](../../../%E4%BC%81%E4%B8%9A%E5%BE%AE%E4%BF%A1/WXWork/Administrator/Desktop/%E5%99%A8%E5%AE%A1%E4%B8%AD%E5%BF%83/%E5%8C%97%E4%BA%AC%E5%85%88%E9%94%8B%E5%8D%9A%E6%96%87%E7%BF%BB%E8%AF%91-2021.10.26/%E8%B0%AD-11.11-30%E4%B8%AA-CN/%E9%82%AE%E7%AE%B1%EF%BC%9ANicholas.Petrick%40fda.hhs.gov)联系Nicholas Petrick；或者致电301-796-6661或发送电子邮件至[Robert.Ochs@fda.hhs.gov](../../../%E4%BC%81%E4%B8%9A%E5%BE%AE%E4%BF%A1/WXWork/Administrator/Desktop/%E5%99%A8%E5%AE%A1%E4%B8%AD%E5%BF%83/%E5%8C%97%E4%BA%AC%E5%85%88%E9%94%8B%E5%8D%9A%E6%96%87%E7%BF%BB%E8%AF%91-2021.10.26/%E8%B0%AD-11.11-30%E4%B8%AA-CN/%E9%82%AE%E7%AE%B1%EF%BC%9ARobert.Ochs%40fda.hhs.gov)联系Robert Ochs。

|  |  |
| --- | --- |
|  | **美国卫生与公众服务部****美国食品药品监督管理局****医疗器械和放射健康中心** |
|

**前言**

**公众意见**

电子版意见和建议可随时提交至[https://www.regulations.gov](https://www.regulations.gov/)，供FDA审议。可将书面意见提交至：美国食品药品监督管理局备案文件管理部（5630 Fishers Lane, Room 1061, (HFA-305), Rockville, MD 20852）。

所有意见均应注明备案文件编号FDA-2009-D-0503。在下次修订或更新本文件之前，FDA可能不会对意见采取措施。

**更多副本**

更多副本可通过互联网获取。您也可以通过电子邮件发送请求至CDRH-[Guidance@fda.hhs.gov](../../../%E4%BC%81%E4%B8%9A%E5%BE%AE%E4%BF%A1/WXWork/Administrator/Desktop/%E5%99%A8%E5%AE%A1%E4%B8%AD%E5%BF%83/%E5%8C%97%E4%BA%AC%E5%85%88%E9%94%8B%E5%8D%9A%E6%96%87%E7%BF%BB%E8%AF%91-2021.10.26/%E8%B0%AD-11.11-30%E4%B8%AA-CN/%E9%82%AE%E7%AE%B1%EF%BC%9AGuidance%40fda.hhs.gov)获取本指南的副本。请使用文件编号（1698）注明您所要求获取的指南。

**目录**

1. 引言 4

2. 范围 5

3. 依据 6

4. 临床研究设计 7

4.1 评价范式和研究终点 10

4.2 对照组 12

4.3 阅片情况和随机分组 12

4.4 评定量表 13

4.5 评分 13

4.6 临床阅片者的培训 14

5. 研究人群 15

5.1 数据可合并性 16

5.2 试验数据的重复使用 16

6. 参考标准 18

7. 报告 19

8. 附录 19

8.1 回顾性阅片者研究的潜在偏倚来源 19

参考文献 21

临床性能评估：用于放射学图像和放射学器械数据的计算机辅助检测设备的注意事项 - 上市前通知510（k）申请

行业和美国食品药品监督管理局
工作人员指南

|  |
| --- |
| ***本指南代表美国食品药品监督管理局（FDA）对该主题的当前看法。本文件不赋予任何人任何权利，对FDA或公众不具有约束力。如果替代方法满足适用的情形和法规的要求，则贵司可使用替代方法。如需讨论替代方法，请联系标题页所列负责本指南的FDA工作人员或办公室。*** |

1. 引言

本指南文件向行业、系统和服务提供者、顾问、FDA工作人员和其他人员提供了有关用于放射学图像和放射学器械数据（在本文件中通常称为“放射学数据”）的计算机辅助检测（CADe[[1]](#footnote-2)）器械的临床性能评估的建议。CADe器械是结合了模式识别与数据分析功能（即结合提取自患者放射学数据的数值、测量结果或特性）以用于识别、标记、突出显示或任何其他方式密切关注放射学器械数据的图像部分或方面从而在预期用户（即医生或其他专业医护人员，在本文件中称为“临床医师”）解释患者放射学图像或患者放射学器械数据期间显示异常的计算机化系统。我们考虑了在2008年3月4-5日[[2]](#footnote-3)和2009年11月17-18日[[3]](#footnote-4)召开的放射学器械专家组公共会议期间就有关CADe器械的文件记录和性能试验提出的建议。我们也考虑了2009年10月21日在联邦公报上公布的指南草案收到的公众意见（74 FR 54053）。

此外，FDA发布了最终指令，将用于乳腺X线摄影乳腺癌、超声乳腺病灶、射线照片肺部结节和射线照片龋齿的医学图像分析仪重新分类为II类器械（特殊控制），并遵守上市前通知（510（k））要求（参见医学图像分析仪的重新分类联邦公报最终指令85 FR 3545（2020年1月22日））。本指南提出了可用于遵守21 CFR 892.2070(b)(1)中特殊控制的建议，为清楚起见，在指南中以斜体表示。

FDA指南文件，包括本指南，并未规定具有法律强制力的责任。相反，指南描述了FDA对该主题的当前看法，除非引用了具体监管或法定要求，否则应仅视为建议。FDA指南中使用的“应该（should）”一词指建议或推荐进行某一事项，并非强制要求。

2. 范围

本文件提供了用于放射学图像和放射学器械数据的CADe器械和用于乳腺X线摄影乳腺癌、超声乳腺病灶、射线照片肺部结节和射线照片龋齿的CADe器械的临床性能评估研究的相关指导。例如，放射学数据包括在患者检查期间通过超声、射线照相、磁共振成像（MRI）、计算机断层扫描（CT）、正电子发射断层扫描（PET）和数字化电影图像产生的图像。如上所述，CADe器械是预期用于识别、标记、突出显示或任何其他方式密切关注放射学器械数据的图像部分或方面从而在临床医师解释患者放射学图像或患者放射学器械数据期间显示异常的计算机化系统。本指南适用于CADe器械，包括CADe器械是组合系统的组成部分的情况，例如组合计算机辅助检测和诊断器械的检测部分。

本指南文件适用于按照21 CFR 892.2050“图像存档和通信系统”进行分类的CADe器械以及按照21 CFR 892.2070“医学图像分析仪”进行分类的器械，包括以下产品代码：

* NWE（结肠计算机断层扫描系统、计算机辅助检测），
* OEB（肺部计算机断层扫描系统、计算机辅助检测），
* MYN（医学图像分析仪）。

FDA希望创建新的产品代码，以在必要时识别和跟踪新的CADe产品类型。[[4]](#footnote-5)

CADe器械可以设计为针对仅一种潜在异常或异常组合或大量多个平行检测方案的独特检测方案，其中每个方案均专门设计为检测患者放射学数据显示的一种潜在异常。例如，本指南范围内的CADe器械包括：

* 设计为确定和提示数字乳房X光造影上的微钙化簇和肿块的CADe算法，
* 设计为确定和提示CT结肠成像研究中的结肠息肉的CADe器械，
* 设计为确定和提示胸腔CT检查上的充盈缺损的CADe，和
* 设计为确定和提示头部MRI研究中的脑部病变的CADe。

为达到本指南的目的，例如，如果图像处理组件、功能、分类算法、培训方法、培训数据集或算法参数发生变化，则我们可以考虑修改CADe算法。有关CADe 510（k）申请的信息，包括可能建议用于CADe 510（k）申请的临床性能评估，参见FDA指南“[用于放射学图像和放射学器械数据的计算机辅助检测设备](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-assisted-detection-devices-applied-radiology-images-and-radiology-device-data-premarket-0) - 上市前通知[510（k）]申请”[[5]](#footnote-6)。

本指南不适用于（1）预期在术中使用的CADe器械的临床性能评估研究，（2）用于可疑癌症病变器械（CADx）的放射学计算机辅助诊断（CADx）软件（按照21 CFR 892.2060进行分类）或者（3）放射学计算机辅助分诊和通知软件器械（按照21 CFR 892.2080进行分类），无论是作为独立器械销售还是与CADe器械捆绑销售，其本身可能受本指南的约束。

对于任何这些类型的器械，我们建议您联系FDA以获得有关监管途径、监管要求的问题反馈，以及有关非临床和临床数据的建议。

3. 依据

*要求按照892.2070分类为医学图像分析仪的器械符合21 CFR 892.2070(b)的特殊控制，包括详细描述预定性能试验方法和数据集的有关要求，以便在设计验证和确认中评估器械是否可以改善阅片者表现（21 CFR 892.2070(b)(1)(ii)）。*

*也要求医学图像分析仪包含性能试验的结果，以证明器械在按照使用说明书使用时能够改善阅片者在适用人群中的表现。性能评估必须依据适当的诊断准确性指标。试验数据集必须包含来自重要队列的足够病例数（例如具有临床意义的混杂因素定义的子集、修正因素、并存疾病和图像采集特征定义的子集），以便能够针对适用人群和成像设备，表征这些个体子集的器械性能估计和置信区间（21 CFR 892.2070(b)(1)(iii)）。以下建议可能有助于证明遵守这些特殊控制。*

建议：以下部分提供了评价CADe系统时，有关临床研究设计、方法、研究人群和报告的建议。本指南提供了应该如何设计CADe器械和开展临床性能评估研究（即充分对照的临床研究）的建议。这些研究可以纳入提交给FDA的510（k）申请。本文件中的建议旨在为CADe器械的开发和试验提供指导，而不是规定可能适用于贵方器械的上市前申请的完整内容或类型。[[6]](#footnote-7)如果您需要FDA对可能适用于器械的分类和监管要求提出建议，则可以按照《联邦食品、药品和化妆品法案》（《FD&C法案》）第513（g）节提交申请。[[7]](#footnote-8)

我们建议，对CADe器械开始独立性能评估和临床性能研究以前，申请FDA对方案进行审评。为申请FDA进行方案审评，可以向FDA提交预申请。[[8]](#footnote-9)

4. 临床研究设计

CADe器械的临床性能评估旨在证明，当预期用户按其拟定标签和说明使用时，器械用于预期用途的临床安全性和有效性。

如以上范围部分所述，CADe器械设计为确定临床医师解释患者放射学图像或数据时可能显示异常的数据。CADe输出与临床医师之间存在较为复杂的关系，因此临床性能可能取决于各种因素，在任何研究设计中应考虑这些因素，包括：

* 在解释过程中使用CADe的时间（例如同期或第二次阅片）；
* CADe标记的物理特征（即尺寸和类型）、边界类型（例如实心、虚线、圆、等高线）、CADe标记与异常的接近程度；
* 用户了解CADe可标记的异常类型；和
* CADe标记的数量。

临床性能评估应该使用充分对照的研究设计，以排除或限制可能影响器械安全性或有效性结论的不同偏倚（参见**第**[**8**](#bookmark56)**节附录）。**当在实验室内（即在临床场地之外）进行临床性能评估时，因为很难在实验室中复制临床环境，我们尤其建议这样做。可用于评估CADe器械的一些研究设计包括：

* 在实际临床条件下评价器械的现场试验或前瞻性阅片者研究（例如随机对照试验）。现场试验在某些情况下可能不太实际，例如，疾病的患病率非常低，但可能需要入组大量患者的情况。
* 可以使用包含回顾性病例收集并补充病变/异常病例的回顾性阅片者研究替代现场试验。
* 对于一些CADe器械，影响因素试验是进行临床性能评估的另一种替代方案。影响因素试验是一种补充了患者病例的回顾性研究，包含难度比常规临床实践更高但是仍然属于器械适用人群的影像学结果（或者其他图像数据）（参见**第**[**5**](#bookmark45)**节研究人群）。**应注意，由于试验与临床实践之间的疾病患病率不同（和难以进行影响因素试验的病例），与临床实践相比，使用样本补充很可能改变阅片者在试验中的表现。

CADe器械的临床性能评估通常使用多阅片者多病例（MRMC）研究设计进行，其中由一组临床阅片者（即在MRMC研究中评价放射学图像或数据的临床医师）在多个阅片条件或方式下（例如未揭盲的阅片者 vs. 借助于CADe的阅片者）评价图像数据。MRMC设计可以采用“完全交叉”，以便所有阅片者对所有病例进行独立阅片。该设计对特定数量的病例具有最高的统计把握度。然而，非完全交叉研究设计可以接受，例如，在前瞻性研究中，多位临床医师无法解释同一患者的数据。

无论您是否决定使用完全交叉研究设计，我们都建议使用MRMC评价范式，以使用上述其中一种研究设计评估CADe器械的临床性能。申请中应包含完整的临床研究设计方案。预先规定统计分析是获得一致和令人信服的科学证据的关键因素。我们建议提供：

* 研究设计描述；
* 描述如何收集成像数据（例如成像器械的品牌和型号及成像方案）；
* 方案副本，包括以下内容：
* 要进行试验的假设和研究终点，
* 检查确认试验所需的任何假设计划，
* 如果不符合要求的假设，要使用的替换程序/检验，
* 表明认定临床研究成功应该符合哪些假设的研究成功标准，
* 特定病例样本量的统计和临床证明，
* 特定阅片者人数的统计和临床证明，
* 图像解释方法及其与临床实践的关系，
* 随机分组方法，和
* 阅片者任务，包括使用的评定量表（参见**第**[**4.4**](#bookmark37)**节评定量表**）；
* 阅片者资质和经验；
* 阅片者培训描述；
* 统计分析计划（即终点、统计方法），其中描述：
* 定义事实的过程（参见**第**[**6**](#bookmark52)**节参考标准**），
* 使用的评分技术的详细信息（参见**第**[**4.5**](#bookmark39)**节评分**）；和
* 支持拟定设计的初探性研究的任何结果。

有效估计CADe器械的临床性能取决于完善的研究设计。完善的临床研究设计的各个方面应包含：

* 研究人群（病变和正常病例）适当代表适用人群或从适用人群中取样；
* 研究设计避免CADe效应发生混淆，例如阅片过程效应；
* 样本量足以证明性能声明；
* 事实定义适用于性能评估，在适当情况下，在研究分析中正确考虑参考标准中的不确定性；
* 在数据集中表示适当的数据队列；
* 阅片者的选择应确保他们代表临床用户中的适用人群；和
* 成像硬件的选择应确保它们与当前的临床实践一致。

4.1 评价范式和研究终点

应该选择可证明CADe器械的有效性的研究终点（即，在目标人群的重要组成部分中，在提供充分的使用说明及危险使用的警告时，将器械用于预期用途和使用条件可以达到具有临床意义的结果）。[[9]](#footnote-10)将根据器械的预期用途选择主要和次要终点，并在开始评价以前固定下来。除临床作用点的敏感性（Se）和特异性（Sp）以外，基于受试者工作特征（ROC）曲线或ROC变体（例如自由响应受试者工作特征（FROC）曲线或地点特定受试者工作特征（LROC）曲线）的性能指标很可能作为候选终点。基于ROC的终点能够评价各种操作点范围内的器械。ROC曲线是Se vs. 1-Sp的图片，总结了器械或临床医师的诊断性能。FROC曲线是Se vs. 每个图像集合中的假阳性标记数量的图片。FROC指标总结了在分析中考虑每例患者的疾病位置和多个疾病部位时的诊断性能。有关这些评估范式的其他详细信息，参见Wagner等人[[10]](#footnote-11)和IRCU报告79[[11]](#footnote-12)。通过报告Se和Sp对，可以在临床阈值或阅片者将采取行动的临界点对器械进行评价。Se定义为患有疾病/状况/异常的患者人群的试验显示阳性的概率，Sp定义为正常患者人群的试验显示阴性的概率（即无疾病/状况/异常的患者）。[[12]](#footnote-13)

应根据临床阅片者的明确临床测定（例如召回或无召回）估计Se和Sp，而不是间接源自用于生成ROC曲线的评定数据。可以在一项阅片者研究中同时收集有关ROC、Se和Sp的数据。例如，在阅片期间，可以首先让阅片者给出在临床上合理的二元响应，以用于Se和Sp评价，然后立即遵守与用于ROC评价的二元响应一致的评定回答。

您可以使用不同的总结性能指标评估阅片者使用CADe器械的有效性（根据特定的器械和临床适应证，这些指标可能不同）。例如，这些指标包括：

* ROC曲线的面积、部分面积或其他指标，
* FROC曲线的面积、部分面积或其他指标，
* LROC曲线的面积、部分面积或其他指标，
* 阅片者Se和Sp（或Se和召回率[[13]](#footnote-14)），和
* 阅片者定位准确性。

尽管我们认为替代性能指标也可能是合适的，但是我们建议在主要分析中使用ROC总结性能指标。如上所述，当使用ROC总结性能指标时，我们建议在分析中包含Se和Sp作为次要终点。报告Se和Sp（或Se和召回率）可能提供附加信息，以了解器械对临床实践的预期影响。当考虑替代性能指标时，我们也建议您联系FDA。

对于基于ROC/FROC/LROC曲线下面积或ROC/FROC/LROC曲线下部分面积的研究终点，我们建议提供实际曲线的图片，并在可能时提供参数和非参数分析方法的总结性能信息。对于仅仅依靠一种ROC分析的潜在局限，请参见Gur等人[[14]](#footnote-15)。最后，您应检查分析中使用的所有方法，以适当确定拟合数据的充分性。

将根据预期用途和器械对临床实践的预期影响，选择基于病变、患者或其他单位指标的性能指标（作为主要或次要终点）。除非您想提出具体的性能声明，否则无需使其他单位分析达到统计学意义。

我们建议描述统计评价方法，并提供以下结果：

* 总体阅片者表现；
* 相关混杂因素或修正因素（例如病变类型、病变大小、病变位置、扫描方案、成像硬件、并存疾病）的分层性能（参见**第**[**5**](#bookmark45)**节研究人群**）（除非提出具体的子集性能声明，否则每个队列的把握度无需达到统计学意义）；和
* 解释阅片者变异性、病例变异性和事实变异性或其他变异性来源（适当情况下）的置信区间（CI）。

我们建议对分析软件进行识别和确认。[[15]](#footnote-16)您应该提供所用分析方法的参考，明确软件实现，并在适当情况下指明版本号。根据预期用途和使用条件，某些经确认的MRMC分析方法（在文献中找到的或者在网上获得的示例）可能不适用于您的器械评价。[[16]](#footnote-17),[[17]](#footnote-18)

真阳性、真阴性、假阳性和假阴性CADe标记的定义应与器械的预期用途和参考标准的表征一致（参见**第**[**6**](#bookmark52)**节参考标准**）。

4.2 对照组

我们建议根据对照方式评估CADe器械的临床性能。在CADe性能评估中，使用传统临床解释（即不借助于CADe器械进行解释）的研究对照组通常是最适当的相关对照。对于作为第二位阅片者的CADe器械，另一种可能的控制方法是由两名临床医师重复进行阅片。这些控制方法或者与实质等同CADe器械的直接比较通常适用于确定实质等同性。其他对照组可能是有效的。我们建议在开始临床研究以前联系FDA以讨论选择的对照组。

研究对照组应使用相同的阅片方法作为器械组，并与临床实践一致。如果本身不是相同的病例，一些病例群体应全部为研究组，以尽量减少潜在偏倚。对于包含每个研究组中的不同病例的设计，我们建议提供描述和流程图，以显示如何将患者和阅片者随机分入不同的小组。

4.3 阅片情况和随机分组

临床评价中的阅片情况应与器械的预期用途一致。以下举例说明了可以纳入CADe临床评价的阅片情况[[18]](#footnote-19)。

* 仅供第二位阅片者使用的器械（序贯设计）：
* 不使用CADe器械的传统阅片（即单纯阅片者）；
* 在进行传统解释后立即显示CADe输出的第二次阅片（该阅片可以在传统阅片的同一次阅片中进行，称为“连续”阅片方案）。
* 用于同时阅片的器械（交叉设计）：
* 不使用CADe器械的传统阅片（即单纯阅片者）；
* 在解释过程中随时都有CADe输出可以使用的同期或同时阅片（在传统阅片的另一次阅片中进行该阅片）。

应该对阅片者、病例和阅片场景进行随机分组，以尽量减少性能指标的偏倚。我们建议描述随机分组方法，并提供相关流程图。其中一种随机方法是在阅片者研究的区组设计中使用拉丁方阵原理。[[19]](#footnote-20)

在多次阅片的情况下对相同病例进行多次阅片，我们建议每次阅片至少间隔4周，以免发生记忆偏差，但是间隔更长时间也是可取的。如果要缩短或延长阅片的间隔时间，我们建议提供支持拟定时间间隔的数据。

4.4 评定量表

应该使用传统医学解释，并报告病变位置、范围和患者管理。当阅片者评定患者的病变和/或疾病状态时，基于ROC的终点（参见**第**[**4.1**](#bookmark21)**节评价范式和研究终点**）可能支持使用更精细的评定量表（例如7分或100分量表）收集数据。我们建议对阅片者进行使用评定量表的培训（参见**第**[**4.6**](#bookmark42)**节研究受试者的培训**）。

4.5 评分

我们将确定阅片者解释与真实状况（例如疾病状态）之间的一致性的程序称为评分过程。评分过程和评分定义是对CADe器械进行临床评估的重要组成部分，应该对其进行描述。我们建议描述确定阅片者解释与实况调查过程中确定的真实状况相对应的过程（即依据、定义和标准）（有关实况调查过程的信息参见**第**[**6**](#bookmark52)**节参考标准**）。

在本文件中，我们描述了临床性能评估的评分情况。使用不同类型的评分方式评价器械的独立性能，见FDA指南“用于放射学图像和放射学器械数据的计算机辅助检测设备 - 上市前通知510（k）申请。”[[20]](#footnote-21)

临床研究的评分过程应与CADe标记的异常和器械的预期用途一致。在开始评价以前，应该描述并确定评分过程。在评分过程的描述中，我们建议表明评分基础：

* 电子或非电子方式；
* 阅片者标记的边界、面积或体积与参考标准的边界、面积或体积的物理重叠；
* 阅片者标记的面心与参考标准的边界或空间位置的相关性；
* 参考标准的面心与阅片者标记的边界或空间位置的相关性；
* 审查阅片者的解释；或
* 其他方法。

对于依赖于审查阅片者的解释的评分，我们建议提供参与阅片者的人数、资质、经验和专业知识水平、在参与评分过程前向其传达的具体说明以及在评分过程中使用的任何具体标准。当多名阅片者参与评分时，应该描述结合各个解释以确定总体评分的过程，或者如何将他们的解释结合到性能评价的过程，包括如何解决不一致的结果。

4.6 临床阅片者的培训

我们建议提供详细说明并对参与研究的临床阅片者如何使用CADe器械进行培训，并提供如何参与临床研究的详细信息。培训内容应包括器械描述和器械使用说明。对于专业化的阅片说明或规则（例如在审评CADe标记时，更改初始无CADe解释的规则），我们建议您根据阅片任务、临床工作流和医学实践论证临床相关性。

我们也建议对阅片者进行使用评定量表的培训（参见**第**[**4.4**](#bookmark37)**节评定量表**），尤其是这类评定量表通常不用于临床实践的情况。这种培训有助于避免获得错误或无法解释的结果。我们建议阅片者培训包括根据研究设计方法对一组有代表性的正常和异常病例进行评定，并利用不属于试验数据库的病例。

5. 研究人群

可以根据定义明确的入选和排除标准前瞻性或回顾性收集患者数据（即病例）。建议提供收集病例的方案。应注意，在临床试验过程中收集的病例应与器械开发期间使用的病例无关。可接受的数据收集方法是根据每家参与采集机构的入选和排除标准收集连续病例。

对于高效且不繁重的典型病例数据集，允许补充病变/异常病例。补充病例可能影响阅片者表现，因此应对补充范围与研究设计的偏倚进行权衡。也可以在研究人群中补充包含某些成像结果（或其他图像数据）的患者病例，这些成像结果对于临床医师较为困难，但是仍然在器械适用人群的范围内。补充通常称为影响因素试验，会导致研究设计发生偏倚。因此，并非所有的影响因素人群开发方法都适用于您的临床评估。例如，当评估设计为帮助检测各种息肉大小的通用CADe器械时，在研究人群中补充含有小结肠息肉的病例可能是适当的。然而，我们不建议根据CADe器械的性能补充病例。

应该从器械的适用人群对研究人群进行采样，以便病例集包含适当范围的病变/异常和正常病例。在使用影响因素试验时，从适用人群中的队列进行选择抽样是适当的。研究的样本量应该足够大，以使研究具有充分的把握度检测拟定性能声明的统计学意义。在适当情况下，研究应包含来自重要队列的足够病例数（例如具有临床意义的混杂因素定义的子集、修正因素和并存疾病），以便获得这些个体子集的临床性能估计和置信区间。除非提出具体的子集性能声明，否则每个队列的把握度不必达到统计学意义。在提出多个性能声明时，需要对多个子集的试验进行预定的统计调整。

在描述研究人群时，我们建议提供具体的信息，在适当情况下包括：

* 患者人口统计学数据（例如年龄、种族、人种）；
* 相对于使用CADe的患者病史；
* 患者疾病状态和放射学检查的适应证；
* 放射学检查条件（例如技术）（包括是否在使用造影剂的情况下进行检查、每例患者的造影剂类型和剂量、患者体质指数、辐射暴露、MRI图像的T加权）和采集的视图；
* 如何收集成像数据的描述（例如成像器械的品牌和型号和成像方案）；
* 采集机构；
* 如适用，处理机构（例如患者数据数字化）；
* 病例数量：
* 疾病病例的数量，
* 正常病例的数量，
* 用于确定疾病状态、位置和范围的方法（参见**第**[**6**](#bookmark52)**节参考标准**）；
* 按照相关混杂因素或修正因素分层的病例分布，例如病变类型（例如增生性 vs. 腺瘤状结肠息肉）、病变尺寸、病变位置、疾病分期、器官特征（例如乳房组成）、并存疾病、成像硬件（例如品牌和型号）、成像或扫描方案、采集部位和处理部位（如适用）；和
* 患者数据的临床、成像和病理学特征与目标人群的比较。

5.1 数据可合并性

如果在其他要求中，海外数据适用于美国人群和美国医学实践，并且由具有公认能力的临床研究者进行研究（21 CFR 814.15），则可以只根据海外临床数据和符合批准标准来批准上市前批准申请。应该判断为什么非美国数据反映的是美国人群在疾病发生、特征、医学实践和临床医师能力方面的预期结果。根据标准临床研究设计，应该从统计学和临床的角度判断来自多个临床试验机构的数据的可合并性。我们建议上市前通知申请遵循有关海外数据和数据可合并性的类似质量数据规范。如果想要将海外临床数据作为上市前申请的基础，鼓励联系FDA。

5.2 试验数据的重复使用

FDA承认采集用于CADe评估的数据有难度。FDA还认识到，如果多次使用同一个试验集，阅片者和CADe算法可能隐匿或明显地调整试验数据。如果在两次试验之间未进行算法培训（例如，如果申办者正在研究新的提示类型的影响），则不了解既往研究结果的一组新阅片者可能解决使用相同试验数据集的问题。如果在两次试验之间进行了CADe算法培训，或者如果使用新的CADe算法，则FDA鼓励使用新采集的独立试验病例评价CADe系统（即，FDA不鼓励重复使用试验数据评价器械性能）。如果考虑在评价CADe器械时重复使用数据，应该证明重复使用任何部分的试验数据不会导致CADe性能的估值发生不当偏倚，并保持试验数据的完整性。

如果希望FDA考虑重复使用任何试验数据，应该控制工作人员访问试验子人群和个体病例的性能结果的途径。因此，可能需要设置“防火墙”，以确保监管评估团队之外的人员（例如，算法开发人员）完全不了解放射学图像和放射学数据，也不了解任何相关的总结性能结果。应该在产品的整个生命周期内保持试验数据的完整性。这类似于涉及临床试验数据监测委员会的数据完整性，以及使用防火墙将负责提出临时方案变更的人员与了解中期比较结果的人员隔离开来。有关临床试验数据监查委员会的数据完整性的更多信息，请参阅FDA指南“[临床试验数据监查委员会的设立和运作-临床试验申办者](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/establishment-and-operation-clinical-trial-data-monitoring-committees)”的第4.4.1.4节[[21]](#footnote-22)

为了尽量降低调整试验数据和保持数据完整性的风险，当打算重复使用任何试验数据时，我们建议开发审核跟踪并采取以下控制措施：

* 从随时间增长的大型数据库随机选择数据；
* 使用过时的图像采集方案或设备获取的数据不再代表当前的实践；
* 对病例可用于评估的次数设置了较小的固定限制；
* 保持“防火墙”以对数据访问进行严格控制，从而确保监管评估团队以外的人员，尤其是参与算法开发的任何人，不会访问数据（即只向评估团队以外报告总结性能结果）；
* 保存数据访问日志，以便在访问数据时进行跟踪，包括哪些人访问了数据、试验条件和性能结果总结的记录；和
* 在每项新的临床阅片者研究中使用新的阅片者。

审核跟踪的目的包括：（1）确定适当规定了培训和试验集的病例，以确保在培训与试验集之间不会发生数据泄密；（2）确保提前确定新的CADe算法（即使用试验集以前）；和（3）提供有关范围的信息，以用于相同的试验集或子集，从而测试其他CADe算法或设计，包括上报FDA的结果以及未报告的结果。我们推荐的管理措施旨在大大降低在后续研究中使用与之前的CADe算法相同的试验数据集评价新的CADe算法的可能性。

如果重复使用试验数据，除总体性能以外，还应该报告相关子集的试验性能。这些子集包括：（1）与之前使用的试验集重叠的新CADe算法的试验集部分；和（2）之前从未使用过的试验集部分。由于这些子集比新的CADe算法的整个试验集更小，这些子集的置信区间会更宽。然而，平均性能的趋势分析有助于表明是否根据之前使用的试验集部分调整CADe系统。

6. 参考标准

对于本文件的目的，患者数据的参考标准（在成像领域通常也称为“金标准”或“真实状况”）表明是否存在疾病/状况/异常，是否包含这些属性，作为疾病/状况/异常的范围或位置。我们将患者参考标准（例如疾病状态）的表征称为实况调查过程。

我们建议提供实况调查过程的依据，并表明是否基于：

* 其他器械的输出；
* 确定的临床测定（例如活检、特定实验室检查）；
* 随访临床成像检查；
* 除成像以外的随访医学检查；或
* 审查临床医师（即临床实况调查者）的解释。

我们也建议描述用于确定该参考标准的方法（例如根据病理学或根据标准治疗确定）。对于依赖于审查临床医师解释的实况调查，我们建议提供：

* 参与临床实况调查者的人数；
* 其资格鉴定；
* 经验和专业知识水平；
* 在参与实况调查过程以前，向其传达的说明；
* 在确定疾病/状况/异常和标记疾病/状况/异常的位置和范围时，临床实况调查者使用的所有可用的患者临床信息；和
* 在实况调查过程中使用的任何特定标准。

当涉及多位临床实况调查者时，应该描述结合他们的解释确定总体参考标准的过程以及您的过程如何解释参与实况调查过程的临床医师的分歧的过程（真实变异性）（参见**第**[**4.1**](#bookmark21)**节评价范式和研究终点**）。应注意，参与实况调查过程的临床医师不应该参与CADe器械的中心临床性能评估。

7. 报告

可以在FDA指南[“报告评价诊断试验研究结果的统计指南-行业和FDA审评者指南](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-results-studies-evaluating-diagnostic-tests-guidance-industry-and-fda)”[[22]](#footnote-23)的指导下报告性能结果。我们建议以电子版的形式提交在任何研究统计分析中使用的数据，包括：

* 患者信息，
* 疾病或正常状态，
* 并存疾病，
* 病变尺寸，
* 病变类型，
* 病变位置，
* 病变分期，
* 器官特征，
* 成像硬件，
* 成像或扫描方案，
* 成像和数据特征（例如使用扫描胶片的CADe的数字化架构差异相关的特征），
* 和统计分析。

有关以电子形式提交数据的更多信息，请查看[FDA的网页](https://www.fda.gov/medical-devices/premarket-submissions/clinical-data-premarket-submissions)。[[23]](#footnote-24)

8. 附录

8.1 回顾性阅片者研究的潜在偏倚来源

尽管具有实际价值，但是评价CADe器械的回顾性阅片者研究可能生成CADe性能的估计值，潜在统计偏倚可能存在一种或更多来源。统计偏倚表示研究的性能估计与适用人群的真实性能不一致的趋势。通常通过标准的研究设计尽量减少或至少减轻许多统计偏倚来源。阅片者研究的一些潜在偏倚来源包括：

* 选择偏倚：受试者（或阅片者）样本不代表目标人群。
* 疾病谱偏倚：研究的受试者（或阅片者）样本不包括目标人群的完整疾病谱。
* 存在缺陷的参考（金）标准偏倚：根据是否存在相关疾病（例如乳腺癌、结肠息肉）对受试者进行分类的情况下，参考程序不是100%准确的。
* 验证偏倚：只根据受试者是否符合参考标准所验证的相关情况，对诊断性能进行统计分析。
* 阅片顺序偏倚：在比较两次或更多试验时，阅片者的解释受到其完成试验时对结果的记忆的影响。
* 上下文偏倚：在异常情况的样本患病率与目标人群的患病率相差较大时，阅片者的解释可能受到影响，导致诊断准确性估计发生偏倚。

当为研究选择的受试者（病例）不代表适用人群时，会引起*选择偏倚*。回顾性选择具有可用图像的受试者会引起选择偏倚。对受试者进行随机或连续采样可能消除或缓解选择偏倚。选择参与研究的阅片者也可能发生选择偏倚。阅片者应反映使用器械的阅片者比例。经过类似培训或具备类似临床经验的少数阅片者（例如来自相同诊所的阅片者）可能无法推广到预计使用CADe器械的整个阅片者人群。

*疾病谱偏倚*是选择偏倚的特殊形式，其中研究受试者代表适用人群中的受试者的部分特征（即研究缺失重要亚组）。

在CADe研究中补充难以诊断的受试者（即影响因素试验）可能导致CADe效果发生偏倚。例如，如果CADe对这些困难病例最有用，则可以使用相对于适用人群的影响因素试验补充CADe效果。然而，由于对在重要亚组中评价CADe具有重要意义，并且在适当使用时可以减少研究设计的样本量，所以鼓励进行影响因素试验。

在CADe研究中补充具有异常状况的受试者可能导致CADe阳性和阴性预测值的估计值存在偏倚，因为这些取决于异常状况的患病率。预测值的“正确”估计值取决于研究之外的患病率估计值。如果阅片者“泄密”研究的增强患病率，并更改其阅片行为，这会导致上下文偏倚，增强患病率也可能间接导致Se和Sp的ROC曲线下面积（AUC）的估计值发生偏倚。

对图像集进行回顾性阅片也可能导致CADe性能发生偏倚。偏倚问题是阅片行为可能相对于实际临床实践发生变化，因为阅片者知道他们参与患者管理不受诊断阅片影响的研究。

确定是否存在相关状况的真实状况可能存在几个偏倚来源。根据参考标准确定真实状况。*存在缺陷的参考标准偏倚*表示将一些受试者错误分类为有或没有某种状况的参考标准。例如，如果在初步评价期间诊断为异常状况阴性的受试者未进行随访以确认他们确实无状况，则真实状况的确定不完全。

*验证偏倚*表示缺失某些受试者的真实状况。如果在统计分析中仅包含已确定真实状况的受试者，则CADe性能的估计值可能发生偏倚。

回顾性阅片者研究的研究设计可能产生统计偏倚的潜在来源。例如，在连续阅片设计中，要求阅片者最初在不使用CADe的情况下依次阅片，然后作为第二名阅片者，在CADe的帮助下进行阅片。对无辅助阅片与无辅助阅片结合CAD辅助阅片进行比较。由于两种形式的读数相近可以最大限度地减少阅片者间变异性，连续阅片设计值得考虑。然而，如果器械使用说明书（IFU）中的阅片条件与阅片者研究不同，可能担心CADe对阅片者诊断性能的影响可能与给阅片者更多时间对每个病例进行阅片所产生的影响相混淆。

另一个问题是，与CADe辅助阅片相比，阅片者可能未充分回想起无辅助阅片，导致CADe的效果增强。如果这一问题是真的，缓解方法是对少部分病例进行随机分组，只在最初的无辅助阅片模式下进行阅片。只在进行初步无辅助阅片以后揭示随机分组。应该为每位阅片者分别进行病例随机分组，以确保一些阅片者向所有病例进行CADe辅助阅片。

连续阅片设计的替代方案是*交叉设计*。在无辅助和CADe输出辅助下对病例进行阅片，在两个独立的阅片过程中间隔洗脱期，以消除阅片者对图像的记忆。一半病例随机分入A组，另一半随机分入B组。在第1次阅片过程中，A组病例无辅助，B组病例在CADe的辅助下进行阅片。在第2次阅片过程中，病例“交叉”为另一种方式（A组辅助阅片，B组无辅助阅片）。在两种模式的比较中，消除了特定阅片过程对阅片的任何影响。然而，与连续阅读设计相比，两次阅片均导致CADe效果的估计值发生额外的变异性。交叉设计可能尤其适用于同期阅片。也可推广到用于评价超过两种模式（例如，如果除无辅助阅片以外，CADe本身也有两种或更多模式）。

参考文献

1. Pepe, M.S., Statistical evaluation of medical tests for classification and prediction. Oxford Press, 2003.

2. Zhou, X., Obuchowski, N.A., McClish, D.K., Statistical Methods in Diagnostic Medicine. Wiley, 2002.

1. 计算机辅助检测的缩写词CADe在整个领域中可能未获得普遍认可。在此处表示本文件中讨论的特定器械类型。 [↑](#footnote-ref-2)
2. [https://wayback.archive-it.org/7993/20170405191951/https:/www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevic](https://wayback.archive-it.org/7993/20170405191951/https%3A/www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevic)[es/MedicalDevicesAdvisoryCommittee/RadiologicalDevicesPanel/ucm124890.htm](https://wayback.archive-it.org/7993/20170405191951/https%3A/www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/RadiologicalDevicesPanel/ucm124890.htm) [↑](#footnote-ref-3)
3. <https://wayback.archive-> [it.org/7993/20170405191949/https:/www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/RadiologicalDevicesPanel/ucm147063.htm](https://wayback.archive-it.org/7993/20170405191949/https%3A/www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/RadiologicalDevicesPanel/ucm147063.htm) [↑](#footnote-ref-4)
4. 有关产品代码的更多信息，请参阅FDA指南“[医疗器械分类产品代码](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/medical-device-classification-product-codes-guidance-industry-and-food-and-drug-administration-staff)”（<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/medical-device-classification-product-codes-guidance-industry-and-food-and-drug-administration-staff>）。 [↑](#footnote-ref-5)
5. [https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-assisted-detection- devices-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-assisted-detection-%20devices-) [applied-radiology-images-and-radiology-device-data-premarket](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-assisted-detection-devices-applied-radiology-images-and-radiology-device-data-premarket) [↑](#footnote-ref-6)
6. 对于本指南中的CADe器械，510（k）申请是最常见的申请类型。指南 [用于放射学图像和放射学器械数据的计算机辅助检测设备 -](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-assisted-detection-devices-applied-radiology-images-and-radiology-device-data-premarket-0) [上市前通知510（k）申请](../../../%E4%BC%81%E4%B8%9A%E5%BE%AE%E4%BF%A1/WXWork/1688850039802833/Cache/File/2022-03/%E4%B8%8A%E5%B8%82%E5%89%8D%E9%80%9A%E7%9F%A5510%EF%BC%88k%EF%BC%89%E7%94%B3%E8%AF%B7)（[https://www.fda.gov/regulatory-information/search-fda-guidance-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-assisted-detection-devices-applied-radiology-images-and-radiology-device-data-premarket) [documents/computer-assisted-detection-devices-applied-radiology-images-and-radiology-device-data-premarket](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-assisted-detection-devices-applied-radiology-images-and-radiology-device-data-premarket)）提供了有关510（k）申请的器械描述和独立性能评估的其他信息。 [↑](#footnote-ref-7)
7. 《FD&C法案》的第513（g）节（21 U.S.C. 360c（g））提供了获取FDA对可能适用于器械分类和监管要求的意见的方法。 [↑](#footnote-ref-8)
8. 有关预申请的更多信息，请参阅FDA指南“[医疗器械申请反馈和会议申请：Q申请计划](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/requests-feedback-and-meetings-medical-device-submissions-q-submission-program)”（<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/requests-feedback-and-meetings-medical-device-submissions-q-submission-program>）。 [↑](#footnote-ref-9)
9. 参见21 CFR 860.7（e）。 [↑](#footnote-ref-10)
10. Wagner, R. F., Metz, C. E., and Campbell, G., Assessment of medical imaging systems and computer aids: A tutorial review. Acad. Radiol., 2007. 14(6):723-48. [↑](#footnote-ref-11)
11. ICRU Report 79, Receiver Operating Characteristic Analysis in Medical Imaging. Oxford University Press, 2008; 8(1).(ISSN 1473-6691). [↑](#footnote-ref-12)
12. Altman, D.G., Practical Statistics for Medical Research.Chapman & Hall, London, 1991. [↑](#footnote-ref-13)
13. 召回率指返回或召回以进行额外医疗评估的患者（包括患病和非患病患者）百分比。 [↑](#footnote-ref-14)
14. Gur, D., Bandos, A.I., and Rockette, H.E., Comparing Areas under Receiver Operating Characteristic Curves:Potential Impact of the Last Experimentally Measured Operating Point.Radiology, 2008; 247(1):12–15. [↑](#footnote-ref-15)
15. 有关MRMC分析软件的更多信息，例如，请参见Obuchowski, N. A., Beiden, S. V., Berbaum, K. S., Hillis, S. L., Ishwaran, H., Song, H. H., and Wagner, R. F., Multi-reader, multi-case ROC analysis:An empirical comparison of five methods. Acad. Radiol., 2004; 11(9); 980-995. [↑](#footnote-ref-16)
16. 有关MRMC文献参考，例如，请参见：Metz, C. E., Fundamental ROC analysis. Handbook of Medical Imaging, Vol. 1. Physics and Psychophysics. SPIE Press, 2000; Chapter 15, 751-769; Wagner, R. F., Metz, C. E., and Campbell, G., Assessment of medical imaging systems and computer aids: A tutorial review. Acad. Radiol., 2007; 14(6);723-48; Obuchowski, N. A., Beiden, S. V., Berbaum, K. S., Hillis, S. L., Ishwaran, H., Song, H. H., and Wagner, R. F., Multi-reader, multi-case ROC analysis: An empirical comparison of five methods. Acad. Radiol., 2004; 11(9); 980-995. [↑](#footnote-ref-17)
17. 对于根据经验证技术分析MRMC数据的软件的在线访问，例如，请参见：LABMRMC软件和通用ROC软件，芝加哥大学：<http://metz-roc.uchicago.edu/> （对于准连续或分类数据）；爱荷华大学MRMC软件：<https://perception.lab.uiowa.edu/OR-DBM-MRMC-program-manual>（对于分类数据）；OBUMRM软件：<https://www.lerner.ccf.org/qhs/software/> [↑](#footnote-ref-18)
18. Obuchowski NA, Meziane M, Dachman AH, Lieber ML, Mazzone PJ, What's the control in studies measuring the effect of computer-aided detection (CAD) on observer performance? Acad Radiol., 2010; 17(6):761-7. [↑](#footnote-ref-19)
19. Bailey, R.A., Design of Comparative Experiments. Cambridge University Press, 2008. [↑](#footnote-ref-20)
20. [https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-assisted- detection-devices-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-assisted-%20detection-devices-)[applied-radiology-images-and-radiology-device-data-premarket](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-assisted-detection-devices-applied-radiology-images-and-radiology-device-data-premarket) [↑](#footnote-ref-21)
21. [https://www.fda.gov/regulatory-information/search-fda-guidance-documents/establishment-and-operation- clinical-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/establishment-and-operation-%20clinical-)[trial-data-monitoring-committees](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/establishment-and-operation-clinical-trial-data-monitoring-committees) [↑](#footnote-ref-22)
22. [https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting- results-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-%20results-)[studies-evaluating-diagnostic-tests-guidance-industry-and-fda](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-results-studies-evaluating-diagnostic-tests-guidance-industry-and-fda) [↑](#footnote-ref-23)
23. <https://www.fda.gov/medical-devices/premarket-submissions/clinical-data-premarket-submissions> [↑](#footnote-ref-24)