人工智能医疗器械软件临床试验设计概要

王卉 md@cirs-group.com

2020.06.23 | 杭州&北京

专注于医疗器械技术服务

注册 | 临床 | 体系 | 合规

微信公众号：CIRS-MD

瑞旭集团（CIRS）-北京西尔思科技有限公司

www.cirs-group.com/md

目录CONTENTS

一、人工智能医疗软件临床试验设计参考资料

二、人工智能医疗器械软件应用分类

三、人工智能医疗软件临床试验现状

四、人工智能医疗软件临床试验设计要素及案例分享

一、人工智能医疗软件临床试验设计参考资料

1.1 参考资料

1. 《深度学习辅助决策医疗器械软件审评要点》
2. 《医疗器械软件注册技术审查指导原则》
3. 《医疗器械网络安全注册技术审查指导原则》
4. 《移动医疗器械注册技术审查指导原则》
5. 《医疗器械临床试验方案设计指导原则》

1.2《医疗器械软件注册技术审查指导原则》

软件安全性级别（YY/T 0664《医疗器械软件 软件生存周期过程》）进行分级

软件安全性级别基于软件损害严重度分为：

A级：不可能对健康有伤害和损坏；

B级：可能有不严重的伤害；

C级：可能死亡或严重伤害。

1.3《深度学习辅助决策医疗器械软件审评要点》

- ✓ AI软件类临床试验设计：
  - 软件的预期用途：辅助决策、辅助筛查、识别、诊断、治疗等-非辅助决策
  - 使用场景和核心功能：前处理、流程优化、常规后处理
  - 软件类型：软件组件、单独软件
- ✓ 确定观察指标、样本量估计、入排标准、随访以及实施机构等要求，来验证软件的安全性和有效性。

二、人工智能医疗器械软件应用分类

2.人工智能软件应用分类

疾病诊断：模拟医生诊断

健康管理：对患者的日常用药行为进行监测、管理

图像分析

手术引导

诊疗规划：狂犬病注射疫苗管理计划软件

2.1 图像分析类产品

对图像进行分析不给出明确诊断意见
（阳性肺结节、病理切片）

II类

对图像进行分析从而给出诊断意见（分析眼底病变）

III类

2.2 手术引导

- ✓ **各类导航系统：**如牙科种植体导航系统、骨科导航系统、达芬奇手术机器人、放疗导航系统。

- ✓ **特点：**软硬件配合使用，两者缺一不可。通过软件对手术进行规划，然后诊疗计划交由硬件进行执行，最后软件再对硬件执行结果进行确认。

三、人工智能医疗软件临床试验现状

3.1 AI医疗软件临床试验现状

- ✓ 目前人工智能类医疗软件临床试验在中国许多地方开展，适应症广泛

已通过临床试验获批的III类有：

- ✓ 冠脉血流储备分数计算软件（科亚医疗CT-FFR ）
- ✓ 心电分析软件（乐普医疗AI-ECG Platform）
- ✓ 颅内MRI、CT软件（安德医疗Bio-Mind）
- ✓ **糖尿病视网膜病变人工智能自动筛查**（北京上工医信）

- ✓ 目前预计**115个**左右AI类医疗软件临床试验试验正在开展。
- ✓ 中国临床试验注册中心备案情况查询：

http://www.chictr.org.cn/index.aspx

3.2 AI医疗软件目前主要开展的临床试验

适应症	病例数	数据库类型	试验设计	试验分组	主要评价指标
肺结节	总例数：900例	CT图像	回顾性诊断试验	金标准组：外部专家 试验组：AI	结节数目一致率
糖尿病视网膜病变	总例数：1000例	眼底图像	诊断试验	金标准组：中心阅片 试验组：AI	灵敏度 特异度 阳性符合率
糖尿病视网膜病变	总例数：1400例	眼底图像	诊断试验	金标准组：3位15-20年眼底医师无时间限制，一致性评分≥2 试验组：初级医师+AI	SEN、FPE、ACC、AUC for ROC
直肠淋巴结	疾病人群：1000例 干扰样本：500例	MRI	诊断试验	对照组：金标准医师判断 试验组：AI	淋巴结个数的一致率

3.3 AI医疗软件目前主要开展的临床试验

适应症	病例数	数据库类型	试验设计	试验分组	主要评价指标
髋关节疾病	疾病人群：600例 干扰人群：600例	X-ray片	回顾性诊断试验	金标准组：3位15-20年高级别放射科医师无时间限制，一致性评分≥2 试验组：初级放射科医师+AI分析	SEN、FPE、ACC、AUC for ROC
肠息肉结节	总例数：2400例	肠镜片	观察研究	对照：常规结肠镜检查 对照组：常规结肠镜检测+AI	结肠息肉检出率（PPP）
肺结节	疾病人群：1000例 干扰样本：50例	CT片	诊断试验	对照组：金标准病理切片 试验组：AI	SEN、FPE、ACC、AUC for ROC

3.4 AI医疗软件目前主要开展的临床试验

适应症	病例数	数据库类型	试验设计	试验分组	主要评价指标
胰腺	总例数：504例	超声图像	回顾性诊断试验	金标准组：病理检查 试验组：超声内镜AI辅助诊断	检查的盲点率
糖尿病视网膜病变	总例数：10000例	眼底图像	诊断试验	金标准组：3位具有10年以上诊断经验的眼底病医师根据标准7视野眼底照相结果得出的一致性诊断结论，一致性≧2. 试验组：AI	敏感性、特异性、ROC曲线下的面积
胰腺	总例数：100例	细胞病理	诊断试验/预试验	金标准组：病理医师对EUS-FNA标本的检查 试验组：EUS-FNA现场快速细胞病理人工智能辅助评估系统	SEN, SPE, ACC, AUC of ROC

3.5AI医疗软件目前主要开展的临床试验

适应症	病例数	数据库类型	试验设计	试验分组	主要评价指标
髋关节疾病	疾病人群：600例 干扰人群：600例	X-ray片	回顾性诊断试验	金标准组：3位15-20年高级别放射科医师无时间限制，一致性评分≥2 试验组：初级放射科医师+AI分析	SEN、FPE、ACC、AUC for ROC
肺癌	目标人群：8000例 干扰人群：1000例	CT片	诊断试验	对照：术后的石蜡病理诊断结果 对照组：医学影像人工智能技术	基于CT图像的影像组学特征/学习特征
肺癌	目标人群：100例	4DCT片	观察研究	对照组：人工勾画 试验组：智能勾画	ACC、真阳性、放射性肺炎发生率

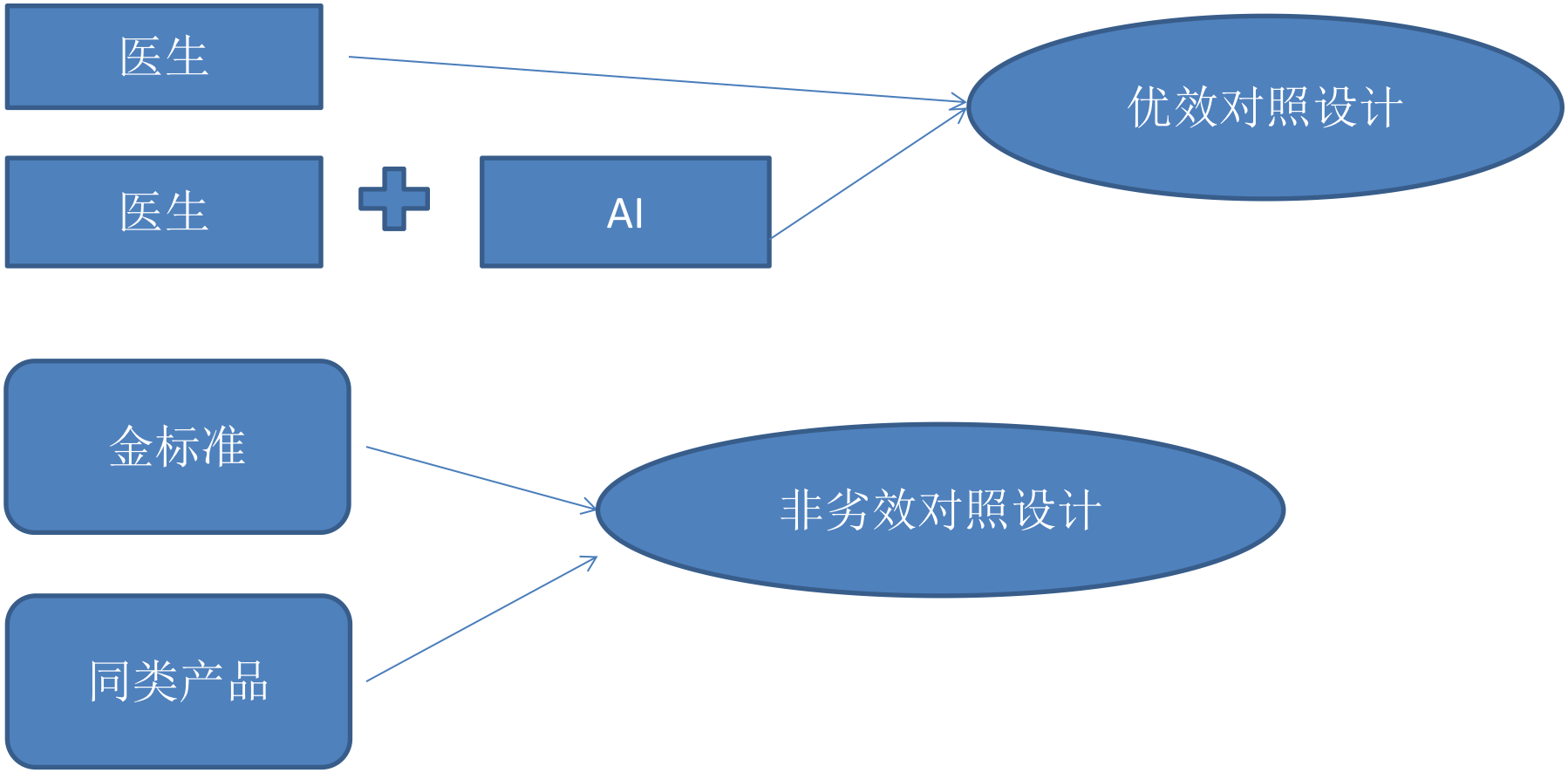
3.5AI产品目前主要开展的临床试验

适应症	病例数	数据库类型	试验设计	试验分组	主要评价指标
骨质疏松	激素治疗组：1500 对照组：1500	血清I型原胶原N-端前肽	干预试验（管理系统）	试验组：激素治疗 对照组：维生素D+钙	血清I型原胶原N-端前肽
不孕症	试验组：137例 对照组：137；例	患病人群	平行随机对照（随访系统）	试验组：人工智能系统语音随访 对照组：常规电话岁阿芳	随访花费时间
髋关节置换术	试验组：40人 对照组：40人	患病人群	随机平行对照（手术辅助决策）	试验组：AI HIP术前规划结合3D打印导板辅助完成手术 对照组：术前二维模板测量结合传统手术方法完成手术	FJS-12评分

四、AI医疗软件临床试验设计要素及案例分享

4.1.1 AI医疗器械临床试验设计类型

考虑到用户的差异性，可选择多阅片者多病例（MRMC）试验设计



4.1.2 AI医疗软件临床试验设计类型

- 诊断试验
- 前瞻性队列研究（观察）
- 回顾性研究（病例对照）
- 双向研究（前瞻性、回顾性）
- 干预研究
- 糖网筛查系统
- 狂犬病疫苗接种随访系统
- 影像诊断
- 影像辅助诊断
- 手术导航系统、疾病健康管理系统

4.2.1 对照的选择

- ✓ **金标准：**指在现有条件下，公认的、可靠的、权威的诊断方法。（组织病理学检查、影像学检查、病原体分离培养鉴定）--------采用**第三方阅片方式**

- ✓ **同类产品名：**与待评价方法具有可比性的已上市产品------所谓可比性指的是适应证或适用范围、基础算法、使用条件基本相似的同类产品。

- ✓ 由于目前图像分析类人工智能产品的算法各不相同，且已上市的同类产品很少，所以目前图像分析类人工智能产品还是选择**“金标准”**作对照。

4.2.2 定量还是定性样本

定量试验：产品主要是用于测量图像中靶病灶的相关参数，-----**数量、长度、体积等。**

定性试验：产品主要是根据图像分析结果对疾病、诊断、预后进行分类，如是否有患病、是否需要转诊、是否需要干预等。**其特点是：指标都是分类变量、基本上没有单位、常常计算其比例。**

4.3 试验目的

- 确认软件在一定适用范围内的安全性和有效性
- 适用范围不要图多，应选择最准确描述------

“辅助诊断”的描述

“诊断”的描述

“分析”的描述

“识别”的描述

非诊断类：导航、随访、健康管理

4.4.1试验对象

- ✓ **目标人群：诊断类：**
- ✓ 包括“阳性”患者和“阴性”患者
  - 试验组：全部为“阳性”患者或有相关体征的患者这些患者疾病应覆盖疾病各阶段，轻、中、重病人都应该有。
  - 对照组：全部为“阴性”患者或没有相关体征的患者，阴性病例还应当选择易于与阳性病例混淆的疾病。

4.4.2 试验对象

- ✓ **目标人群：非诊断类：**
  - 手术导航：导航系统手术人群和传统手术人群
  - 健康管理：接受AI系统慢病管理的人群和普通慢病管理人群
  - 诊疗规划系统：接受AI系统诊疗的人群和普通电话管理人群

4.4.3试验对象

- 为鼓励创新并降低临床试验成本，临床试验可使用**回顾性数据**，但应在设计时考虑并严格控制偏倚问题，原则上应当包含多个不同地域的临床机构(非训练数据主要来源机构)的同期数据。
- **高风险软件：**适用范围变更应当开展临床试验,其他情况原则上可使用回顾性研究
- **中低风险软件：**可使用回顾性研究

4.5.1 评价指标

- **诊断指标：**
- 定性指标：灵敏度、特异度、符合率和Kappa值。
- 定量指标：离群点检查、回归拟合方程、Pearson相关系数、Bland-Altman图。
- 次要指标：敏感性/特异性衍生指标、ROC/AUC衍生指标、组内相关系数、Kappa系数、时间效率、数据有效使用率等指标作为观察指标

由于诊断试验的特殊性，上述评价指标对产品的评价都是很重要，而且各自的侧重点常常不一样，所以有时图像分析类人工智能产品很难确定主要评价指标。

4.5.2 评价指标

- 非诊断类指标：有效率、时间花费、具体数值（血糖控制水平、髋关节评分）

4.6.1 样本量计算

- 定性的样本量估算需要分阳性受试者和阴性受试者两部分，分别按照以下公式计算：

$$n = \frac{\left[Z_{1-\alpha/2}\right]^2 P(1-P)}{\Delta^2}$$

- 由于实际目标人群中的阴阳性比例与理论值的比例不一样，这就导致实际参与筛选的患者要多于理论值。

4.6.2 样本量计算

- 假设P rev=50%，n=189，Z1-β取正态分布曲线下95%对应的界值1.645，计算得至少纳入412例病人。因此，我们可以有95%的把握认为412例病人中至少有189例受试者患**眼底病变**。

$$Z_{1-\beta} = \frac{(N_{total} \times P_{rev}) - n}{\sqrt{N_{total} \times P_{rev} \times (1 - P_{rev})}}$$

4.6.3 样本量计算

注意

文献中的灵敏度和特异度可能并不适用于人工智能研发公司的软件。

测试数据集用于测试软件的诊断性能，类似于小样本试验。灵敏度和特异度，可以供参考给统计师进行样本量估算。

4.7 注意事项

- 重复使用受试者数据
- 受试者知情同意签署
- 高质量图像获取率
- 图像采集设备的适配性
- 精密分析
- 金标准评价者培训
- 临床试验的风险

4.7.1 注意事项

1. **重复使用受试者数据：**存在重复使用同一名患者不同时期的图像数据：QA应避免出现
2. **受试者知情同意签署：**前瞻性样本需获得知情，具体情况需要与机构进行讨论执行
3. **高质量图像获取率：**方案严格入排和图像标准，并对图像拍摄者进行培训

4.7.2 注意事项

4. **图像采集设备的适配性：**明确图像格式的要求

5. **精密分析：**同一个人同一个时期内的多幅照片分析一致率

6. **金标准评价者培训：**必要时可以选择一些高年资的医生参与“金标准”判断

4.7.3 注意事项

7.临床使用风险

- 假阳性：误诊，过度医疗风险
- 假阴性：漏诊，快速进展疾病风险
- 人因与可用性：用户操作错误
- 进口软件：人种、流行病学、临床诊疗准则等差异

Thank You!

本次会议资料将通过以下微信平台发布

CIRS-MD

CIRS

总部地址：杭州市滨江区秋溢路288号东冠高新科技园1号楼11层 310052
北京地址：北京市西城区宣武门西大街28号大成广场7门1109-1111室
网站：www.cirs-group.com (中文) www.cirs-md.com (EN)
电话：0571-8720 6527
邮箱：md@cirs-group.com