**方案模板使用说明**

1. **此方案模版是为研究者撰写方案提供的参考, 使用时请注意涵盖黑体字中的主要方面。**
2. **此方案模板适用于研究者发起的临床预测模型研究方案，包括诊断预测模型和预后预测模型。**
3. **以下凡斜体字括号（）的蓝色部分均为给研究者的提示语，制定完适合本研究的方案后，请勿删去斜体字内容。**
4. **在排版时请注意将方案的封面单独列为一页。**
5. **多中心研究应注意描述多中心情况，我院为参加单位的，应使用组长单位的方案。**

*（以下为方案封面页）*

**研究者发起的临床预测模型研究方案**

**项目名称（中文）：***（请填写项目名称中文）*

**项目名称（英文）：***（请填写项目名称英文）*

研究单位：*（请填写单位名称，如果多中心研究，则列出所有单位并注明组长单位和参加单位；如果仅在我院开展的则此项填写为“复旦大学附属中山医院“）*

研究负责人（签名）：*（本院单中心研究，请填写本院负责人名字，在方案定稿后，研究负责人需在其打印名字后签署姓名和日期）*

申办者：复旦大学附属中山医院

版本号：*（V1.0，统计审查退回修改后更新为V1.1，V1.2；函审/会审后修改更新为V2.0,V3.0*）。

版本日期：*（方案修改后注意更新版本日期*）。

*（以下为方案正文）*

**版本历史/修正历史**

|  |  |  |  |
| --- | --- | --- | --- |
| **文件** | **版本号** | **版本日期** | **修正原因说明和改动部分总结** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**研究者声明**

我将根据中国GCP规定认真履行研究者职责，亲自参加或直接指导本临床研究。我方已阅读并确认此方案，同意此方案的科学性和伦理性。我方将按照中国法律法规、赫尔辛基宣言、中国GCP以及此研究方案规定履行相关的职责，并在经学术委员会和伦理委员会同意后才可实施，除非为保护受试者的安全、权利和利益而必须采取措施。我方将对此研究方案保密。

**研究单位**

主要研究者：

主要研究者（签字）：

签字日期：

1. **方案摘要**

*（以表格形式摘录研究方案的主要内容，突出重点，力求简洁。主要内容包括：标题、研究背景、研究目的、整体设计、研究人群（含入选标准、排除标等）、研究评价指标、预期研究年限、数据来源、数据治理或数据管理、样本量及其确定依据）*

|  |  |
| --- | --- |
| 标题 |  |
| 方案编号（如适用） |  |
| 研究背景 |  |
| 研究目的 |  |
| 研究设计 |  |
| 研究人群 |  |
| 预期研究年限 |  |
| 数据治理或数据管理 |  |
| 样本量估计（详述模型构建和模型外部测试集的样本量情况） |  |
| 模型构建 |  |
| 模型测试 |  |

1. **研究背景**

*（简要介绍研究背景，包括国内外研究的现状和意义，本研究的前期基础等。）*

1. **研究目的**

*（临床预测模型一般可分为****诊断模型****和****预后模型****，简要阐述本研究计划构建的预测模型类型和目标。如果研究仅为既往模型验证，请明确阐述。*

***诊断模型****关注的是基于研究对象的临床症状和特征，诊断当前患有某种疾病的概率；*

***预后模型****关注的是在当下的疾病状态下，未来某段时间内疾病复发、死亡，伤残以及出现并发症等结局的概率。）*

1. **研究人群**
2. 数据来源

*（分别描述训练、验证和测试模型的数据来源（如随机试验、队列研究或注册研究数据等），特别外部测试集的数据来源需描述清楚，若数据来自公开数据库，提供公开数据库的基本信息和使用情况。还应详细描述关键研究日期，包括数据收集的开始时间、结束时间，如果适用还应有随访结束时间。）*

1. 入选标准

*(用清单的方式列出拟参加入选合格受试者的标准，包括疾病的诊断标准(尽可能说明定量检验指标的上、下限)；入选前受试者的相关病史、病程和治疗情况要求；其他相关标准，如年龄、性别等；如果不同人群有区别，分开撰写)*

1. 排除标准

*(用清单的方式列出影响研究结果或安全性评估的情况，如可能影响研究的其他治疗、合并疾病和妊娠、认知障碍等。)*

1. 中途退出标准（如适用）

*(用清单的方式列出退出标准。如受试者要求撤回知情同意；研究者从医学角度考虑受试者有必要中止研究等。)*

1. 样本收集和样本要求（如适用）

*（包括样本类型以及样本采集、保存和运输方法，样本采集的时间范围等。某些受试者样本可以来自既往的、其他研究的样本集或无特定用途的样本集。）*

1. **研究设计**
2. 研究设计类型

*（简述研究的整体设计。*

* ***诊断模型****一般采用横断面研究数据，有时候需要依赖随访获取结局（确诊）信息的诊断预测模型也可以采用纵向研究；*
* ***预后模型****可利用队列研究、随机试验、巢式病例对照等研究数据，实际研究中如果使用回顾性队列数据，一定要注意数据存在的潜在偏倚的影响）*

1. 金标准/预后结局指标

* *（****诊断模型****研究中，需要设置“金标准”， 详细描述金标准以及选择金标准的理由，如没有金标准，需说明理由。“金标准”的诊断应该在“盲法”状态下进行，即“金标准”的诊断不能借助预测模型中的预测因子信息，以避免诊断评估偏倚。*
* ***预后模型****研究中，需明确设置预后结局指标，详细描述该结局选择的理由、定义和评估方法，如何以及何时进行评估，如果实施了盲法则需说明具体的实施策略。）*

1. 样本量估计

*（临床预测模型研究也需对样本量估计进行说明，是基于统计方法，还是基于现实考虑（比如使用现有数据）。如基于统计方法计算，建议研究者根据BMJ文献（文献链接：*[Calculating the sample size required for developing a clinical prediction model | The BMJ](https://www.bmj.com/content/368/bmj.m441.long)*)中提出的四步法对研究进行建模部分样本量的合理计算。既往常用的“10 EPV”（10 events per variable）经验法则存在诸多问题，需要结合具体研究情境具体判断，在此不做推荐。）*

1. 数据集定义

*（预测模型的研究一般建议构建三个数据集：训练集（用于模型训练）、验证集（若有，用于模型调优，或称作调优集）、测试集（若有，用于模型性能评估），明确训练集、验证集、测试集的划分方法、划分依据、数据分配比例。训练集原则上需保证样本分布具有均衡性，测试集、验证集原则上需保证样本分布符合真实情况。）*

1. 预测因素

*（清晰定义建立或验证多因素预测模型所使用的所有预测因素，包括如何以及何时测量，如使用盲法进行测量，应描述盲法过程。）*

1. **数据治理/数据管理计划**

*（数据采集需考虑采集设备、采集过程、数据脱敏等质控要求，并建立数据采集操作规范。前瞻性研究要求使用EDC进行数据采集。回顾性研究需详细说明数据采集的时间和人群范围、关键字段。*

*回顾性研究需详细说明数据采集的时间和人群范围、关键字段以及数据治理计划；前瞻性研究要求使用EDC进行数据管理，需明确质量核查和管理计划。）*

1. **缺失数据的处理**

*（在研究者尽可能完善数据质量的前提下，临床研究仍易出现变量信息缺失，缺失数据对于研究结果的影响不容忽略。数据治理及数据管理过程中，应尽可能追踪捕捉遗漏的记录，研究者需明确阐述缺失数据的处理方法，如仅分析完整数据、单一插补和多重插补等，并说明其理由。）*

1. **偏倚的考虑**

*（在预测模型的研究中需考虑偏倚的控制，在方案中应充分考虑各种潜在偏倚及其影响，并制定控制偏倚的有效措施。在预测模型的研究中常见的偏倚包括：因选择性地入选和/或排除数据或失访、退出、剔除、记录缺失等导致的选择偏倚；因测量、数据收集或评价方法的不准确或不一致导致的信息偏倚。建模方法选择不恰当导致的偏倚等等。）*

1. **模型建立与验证**

1. 预测因子筛选与建模方法

*（变量筛选的目的是选择重要的协变量，基于这些协变量可以创建一个最准确地预测感兴趣结局的模型。可以结合临床知识和既往研究，基于传统的逐步回归、信息准则、正则化方法、机器学习等方法进行变量的筛选。模型选用前需考虑是否满足模型或方法的假设。*

*如使用人工智能算法，需提供所用算法的名称、类型（如有监督学习、无监督学习，基于模型、基于数据，白盒、黑盒）、结构（如层数、参数规模）、输入输出数据类型、流程图、算法编程框架、运行环境等基本信息，并明确算法选用依据，包括选用的理由和基本原则。若组合使用集成学习、迁移学习、强化学习等，亦需提供算法基本信息以及算法选用依据。）*

2. 模型评价

*（模型在构建好以后，需要评估预测模型的性能，查看预测感兴趣结局的效果如何。模型的评估指标建议根据研究者需求进行选择，常用的评价指标包括区分度（discrimination）和校准度（calibration）指标，也可增加灵敏度、特异度、阳性预测值、阴性预测值、混淆矩阵及其衍生参数等多维度反映模型性能。也可考虑对模型的临床适用性进行评价，增加决策曲线、综合判别指数（IDI）、净重分类指数（NRI）等指标。模型的训练方式包括但不限于留出法和交叉验证法，需说明模型选用依据和基本信息。若有模型的调优，需明确优化策略和实现方法。若涉及多个预测模型的比较，需采用适宜统计方法进行比较。）*

3. 模型验证

*(在构建好预测模型后，需要对其进行验证，以确保模型的稳定性以及对提供数据来源的人群能进行高质量的预测，并且可以适用于不同的人群。需基于测试集对构建的预测模型进行评估，综合考虑假阴性与假阳性、重复性与再现性、稳健性、实时性等适用评估要求，以证实模型性能满足研究设计目标。)*

4. 模型更新（如适用）

*（如果有，描述验证模型后进行的任何模型的更新（如再校准等）方法过程。）*

5.模型可解释性

*（所建模型应当具备可解释性，如基于机器学习建立的模型，还应提供模型可解释性分析的方法。）*

6.模型展示方法（如适用）

*（如果有，说明如何展示预测模型，包括预测模型的公式、列线图、网页计算器、建议评分表等方法。）*

1. **质量控制**

*（一般而言，临床预测模型的研究中需要特别关注数据采集和治理过程的质量控制。数据采集过程，需设置数据采集操作规范文档，根据数据采集方式明确采集设备、采集过程、数据脱敏等明确质控要求。数据整理/治理过程，明确数据清洗、数据预处理的质控要求。对于一些需要进行数据标注的研究，需提供数据标注操作规范文档，明确标注资源管理、标注过程质控、标注质量评估、数据安全保证等要求。若数据来自公开数据库，提供公开数据库的基本信息和使用情况。在模型构建和验证角度，需规范整个数据分析过程。）*

1. **研究相关伦理学**

1. 伦理委员会审核

*（本方案和书面知情同意书及与受试者直接相关的资料必须提交伦理委员会，获得伦理委员会书面批准后方可正式开展研究。研究者必须至少每年（如果适用）向伦理委员会提交研究年度报告。在研究中止和/或完成时，研究者必须书面通知伦理委员会；研究者必须及时向伦理委员会报告所有研究工作中发生的变化（如方案和/或知情同意书的修订），并且在未获得伦理委员会批准之前不得执行这些变动，除非是为了消除对受试者明显且直接的风险而做出的变更。在发生这类情况时，将通知伦理委员会。）*

2. 知情同意

3. 获得知情同意的程序

*（研究者必须向受试者或其法定代理人提供易于理解的并且经伦理委员会批准的知情同意书，并给与受试者或其法定代理人充分的时间考虑本项研究，在从受试者获得签署的书面知情同意书之前，受试者不得入组。 在受试者参与期间，将向受试者提供所有更新版本的知情同意书以及书面信息。知情同意书应作为临床试验的重要文档保留备查。）*

*（此部分请按具体情况自行修改，根据实际情况进行书写，如果已经申请免除知情同意书请说明申请免除的理由，如果有部分免除知情同意书，或者口头知情等情况必须写清楚）*

1. **保密措施**

*（通过本项目研究的结果可能会在医学杂志上发表，但是我们会按照法律的要求为患者的信息保密，除非应相关法律要求，患者的个人信息不会被泄露。必要时，政府管理部门和医院伦理委员会及其有关人员可以按规定查阅患者的资料。）*

1. **研究项目的预期进度和完成日期**
2. **参考文献**