

北京大学临床数据科学家临床数据分析研修班

招生简章

项目编号: 2024-12(校)

【项目内容】

健康医疗大数据的发展日新月异，体现在数据资源和种类的不断丰富，以及分析策略和方法的不断进步。当下，研究者不仅可利用自行建立的临床随访队列研究开展病因学和疾病预后探索，还可方便地获取公开数据资源，回答感兴趣的研究问题。不同于传统以 CRF 表获取的规则的观察性研究数据，既有电子健康档案数据在形式上呈现纵向且不规则的特征，如何有效利用该类型数据开展研究较大的挑战。此外，近年来兴起的机器学习算法，以及多种环境因素与疾病表型的关联研究，均为提升临床研究的深度和广度提供了机遇。

本次工作坊课程聚焦上述热点问题，旨在帮助研究者：

1. 利用既有电子健康档案数据，设计流行病学观察性研究，开展规范的数据清理与整合，熟悉常用的疾病预后研究、疾病风险预测模型建立和验证的范式；
2. 关注电子健康档案数据不规则重复测量资料的特征，学习利用纵向数据分析方法开展研究（线性混合效应模型和轨迹分析）；
3. 了解如何利用常用的机器学习算法开展临床研究（增强回归树，随机生存森林等）；
4. 学习基于两种类型的观察性研究数据（时间序列和横断面调查），解析环境因素（空气污染、气候变化）对人群健康影响的分析思路、研究设计和研究方法。
5. 通过标准数据元，学习中医药领域临床研究数据治理思路，结合大语言模型提升数据治理效率。

【办学单位】

北京大学医学继续教育学院

【学术支持单位】

北京大学健康医疗大数据国家研究院、北京大学第一医院

【招生对象】

本次工作坊为希望独立开展临床研究的临床工作者、医疗健康其他领域工作者量身打造，也适用于健康数据科学相关从业者。包括但不限于：

1. 临床医生、医学相关专业研究生和科研人员：为研究课题增加新内容，运用多种类型的统计/机器学习方法丰富研究内容，提升课题/论文质量；
2. 科研管理人员：深入理解基于常规观察性数据开展临床研究的思路 and 策略，为做好科研课题的规划和管理提供助力；
3. 数据科学家：了解临床数据管理和临床研究设计常规，理解前沿数据分析工具与技术与健康医疗领域的应用空间与潜力，促进跨学科合作。

【培训地点】

腾讯会议线上授课+北大医学部线下授课。

【培训时间】

2024 年 6 月 17、19、21、25、27 日晚 18:30-21:00 线上授课，仅 6 月 29 日周六全天（8:30-17:30）线下课程，共计 20 学时。

【课程安排】

1. 第一次课（线上，18:30-21:00，主讲人：王晋伟）：基于既有电子健康档案数据建立临床随访队列研究；基于流行病学观察性研究设计，开展探索疾病发生和发展危险因素，及建立疾病预后风险预测模型等问题的研究；
2. 第二次课（线上，18:30-21:00，主讲人：李承龙）：重复测量数据的清理与整合方式；统计学分析的招式解析——从方差分析到混合效应模型；从纵向数据中挖掘指标潜在的发展模式——动态轨迹分析模型；
3. 第三次课（线上，18:30-21:00，主讲人：张斐斐）：以科研任务为导向的临床数据资料分析实战：介绍既有电子健康档案数据资料清理、分析要点；常用机器学习算法在临床研究中应用；介绍增强回归树，随机生存森林等机器学习算法的原理以及在医学研究中的应用；
4. 第四次课（线上，18:30-21:00，主讲人：杨超）：以实例带入，讲解环境因素（空气污染、气候变化）对人群健康影响的分析思路、研究设计、研究方法，使学员了解环境因素与人群健康相关研究的基本概念、分析策略和几种常见的统计分析模型，能够基于手中已有数据进一步拓展研究维度，挖掘新的研究视角；
5. 第五次课（线上，18:30-21:00，主讲人：彭苏元）：中医药数据治理与标准化

策略：介绍标准数据元，讲解如何通过 GB/T 国标进行临床研究数据标准化数据治理（如名医经验、经典名方、中药饮片等）；并介绍大语言模型在中医药领域研究中的应用，讲解如何运用大语言模型提升数据治理效率；

6. 第六次课（线下实操，8:30-12:00,13:30-17:30，主讲人：王晋伟、李承龙、张斐斐、杨超、彭苏元）：显示实例操作，包括：

1.于临床随访队列开展关联分析和风险预测模型构建的 SAS 编程实践 王晋伟
以实例解析 Cox 比例风险回归模型进行关联分析和建立预测模型、区分度和校准度评价、模型校正、打分表制作。

2.纵向随访队列的混合效应模型和群组轨迹模型的 SAS 编程实践 李承龙
以实例解析线性混合效应模型(MM)、广义线性混合效应模型(GLMM)、群组轨迹模型(GBTM)。结合实例讲解 GBTM 相较于其他轨迹模型的优劣和使用注意事项。

3.以科研任务为导向的临床数据资料分析实战 张斐斐
通过示例数据，展示对应于具体科研任务的不同数据类型的整理与分析方法，并展示常用机器学习算法在临床研究中的应用(增强回归树，随机生存森林等)。

4.环境因素对人群健康影响的数据分析实践 杨超
进行已发表环境流行病学领域高水平论文领读、并进行数据清理、统计分析和软件包使用复现(时间序列数据和横断面调查)。

4. 中医药领域数据分析处理实战 彭苏元
以示例讲解舌诊、脉诊、证候数据分析方法，对大模型文章精读解析。

【主讲教师】

王晋伟 北京大学第一医院 副研究员（项目负责人）

李承龙 北京大学健康医疗大数据国家研究院 助理研究员

张斐斐 北京大学健康医疗大数据国家研究院 助理研究员

杨 超 北京大学第一医院 助理研究员

彭苏元 中国中医科学院中医药信息研究所 助理研究员

课程助教：王福琳、王万州等。

【培训证书】

学员完成全部课程学习，考评合格，将颁发北京大学结业证书，加盖北京大学钢印。

【报名缴费】

1. 费用：3600 元/人（含证书工本费，不含线下课程餐费）

注：学员完成全部课程学习，考评合格，将颁发北京大学结业证书，加盖北京大学钢印。

2. 报名缴费流程：

1) 报名：扫描下方二维码填写报名信息

2) 缴费：报名后，将培训费用于截止日期前全额对公汇入以下账户：

账户名称：北京大学医学部

开户银行：工商银行北京东升路支行

账 号：0200 0062 0908 9112 565

用途（汇款时请备注）：健康数据科研+学员姓名

注：将缴费凭证（转账截图/汇款凭证照片）通过微信或邮件发送给报名联系人崔老师（联系方式见下），确认报名成功。

3. 报名缴费截止时间：2024 年 6 月 14 日。

注：小班授课，人数有限，额满即止（以缴费顺序为准）。



（扫描二维码，填写报名信息）

【联系方式】

报名联系人：王老师

电话：010-82806531，15101084860（微信号：15101084860）

邮箱：pku_dshm@bjmu.edu.cn

地址：北京市海淀区学院路 38 号北京大学医学部校内北大医学科技楼