

联合用药数据资源与智能分析研究构想

我的研究定位与合作目标

邓老师您好。我目前负责医院临床药学工作，本科为药学专业，硕士为中药学专业，同时具有抗感染临床药师和信息药师的工作背景。我已有基于 FAERS 开展药物安全研究和发表论文的经历，也有三项人工智能与药物安全相关项目经验。在进一步攻读博士学位时，我希望把临床药学问题、数据库建设、数据提取、机器学习和大数据分析结合起来，形成能够由我持续牵头、也能与团队既有研究相互促进的研究方向。

阅读您及合作研究群的论文和网站后，我的初步判断是，最适合我的位置不局限于提供医院病例或为已有模型做临床验证。我希望从研究对象、数据资源和算法任务的设计阶段就参与并承担主要工作，将医院里的用药问题转化为可计算、可复现、可检验的问题，再以慢病多药联用和抗感染治疗作为研究载体。我的基础实验经验相对有限，因此不准备把必须独立完成细胞、动物和药代实验作为整条研究路线的前提；确需实验时，依托有条件的合作平台，自己承担问题提出、数据与算法研究、结果解释以及临床证据衔接。

我建议以一个相互连接的研究体系展示这种合作可能性：建设条件化联合用药证据资源和纵向临床药学院队，通过药品制剂、用药条件、时间和证据版本将两者连接；围绕这个体系，研究证据抽取、条件化预测、复杂联用表示与动态获益风险分析。我希望将数据资源的科学问题、技术实现和算法研究贯通起来，亲自参与从任务定义到建模分析的完整过程。

以下是供进一步沟通的完整研究布局，目的是说明我能带来什么、哪些工作可以先启动、哪些环节适合与团队共同开展。各模块可以共用数据与方法，但不意味着同时确定多个博士课题。正式研究主题、疾病范围和优先任务，应在了解团队当前安排后共同收敛。

对团队既有工作的理解与拟增加的研究内容

我理解，您及合作的 LinDing 研究群已经具有计算方法、数据库、临床与体检数据、药物组合和组学研究基础。DeepIDC、MDDC 和 ACRM 尤其构成了与我背景直接相关的连续研究基础。我希望承接这些工作，在已有基础上增加条件、真实暴露和患者病程这些研究维度，进一步明确数据与方法可以产生怎样的增量。

已有基础	我的理解	拟由我牵头增加的内容
DeepIDC 注射剂组合预测	建立了基于注射剂配伍参考表的 IDC 标签任务，原文讨论了浓度、溶剂等条件	保留原标签谱系，重申证据域，建立同药对跨条件、跨制剂的预测和冲突解释任务
MDDC 双任务研究	已联合学习 DCI 与 DDI；其中 DDI 为药代变化与不良作用类别，不是患者有无伤害	保留不同证据域的边界，研究标签粒度、来源依赖和跨任务信息迁移何时有效
ACRM 抗菌药组合研究	已综合临床处方与有效性信息，并开展独立回顾队列评价	将组合级信息连接到感染治疗阶段、实际给药、剂量和患者状态，构建新的病程级任务
P450Rdb 及中药资源建设	已有机理知识与文献资源的组织经验	提取人体相关证据、适用条件与证据层级，研究机制线索和临床效应的连接与断点
慢病临床研究与中药研究框架	已有疾病分类、代谢组和单细胞等工作，也已提出暴露、疗效分层与临床回证思路	把实际治疗策略、连续暴露、停换原因及共同干预落实为可复核的数据和分析对象

DeepIDC 原文中部分条件标签被合并为较粗分类，其参考表标签和举例还涉及不同层面的联合用药问题。我拟保留原始 IDC 标签用于基线复现，再逐条核对来源，将配伍条件声明、体内作用警示等重新定位到适当证据域，形成独立的审校标签。MDDC 也保留原标签定义，与新研究的患者事件标签分别管

理。[1-2] ACRM 则为进一步研究病程级方案提供直接基础；其有效性定义和完整分析方法，待结合全文进一步理解后再具体对接。[3]

团队的优势为我提供了方法与机制上的合作基础。我的增量应体现为能够独立提出药学数据科学问题、构建高质量研究资源、实现抽取与建模流程，并把模型的输出解释到具体用药条件和临床任务，而非重复团队已经完成的工作。

首先区分四种不同的联合用药问题

联合用药包含多个层面的研究问题。同一组合在不同层面可以得到不同结论，研究中需要保留这种差异。

证据域	真正回答的问题	关键条件与终点	需要保持的证据边界
体外配伍与稳定性 DCI	特定制剂在指定接触条件下是否出现物理或化学变化	制剂、浓度、溶媒、温度、接触方式和时长；外观、颗粒、含量与降解	结论限于所测条件和终点，患者伤害另需证据
体内 PK 或 PD 相互作用	一种药如何改变另一种药的暴露或药效	方向、剂量、顺序、人群、酶或转运体；AUC、Cmax 及药效指标	药代或药效变化与临床伤害分别判定
患者安全事件与 ADE 证据	暴露后发生了什么安全事件，其关联证据多强	实际暴露、事件定义、时间窗、对照、分母和归因	观察到事件与认定药物归因分层保存
临床疗效	在特定适应证和治疗阶段，某策略相对对照产生何种结果	治疗起点、共同治疗、病情、随访与临床终点	临床终点与处方频次、体外协同、药师判断分别报告

例如，两药不适合同管接触，并不意味着不能经符合要求的不同给药安排用于同一患者；未见外观改变也不等于已经证实化学稳定。我的数据和算法将分别输出各域结论及适用条件，允许“有证据支持”“有证据反对”和“信息不足”并存。疗效与安全先分别研究，在具有明确权重和验证依据时再讨论综合权衡。

患者安全层先记录暴露后观察到的事件，再另存药物相关性裁定；未完成归因的事件不直接标为确定 ADE。

这一划分同时约束数据库和算法：训练时显式保存未知状态，区分说明书声明与实验测量，分别管理潜在风险警报与实际事件。清楚的任务语义为多任务学习、知识融合和临床迁移提供可解释基础。

两类相连资源与一条证据连接链

条件化联合用药证据资源

我拟首先建设一个主题明确、逐条可溯源的证据数据集。首期围绕抗感染药及慢病常见合用药，不追求一次覆盖全部药物。其基本单位是一条有来源、有条件、有终点的证据记录；同一药对可以对应多个制剂、实验条件、研究人群和不同域的结论。

核心结构包括成分、产品、来源、研究、证据记录、暴露方案、条件、终点和证据关系。成分表保留标准名、别名和外部映射；产品表保留剂型、规格、厂家、组成及可获得的辅料信息；来源表记录 DOI 或说明书标识、版本、许可和原文定位；研究表识别同一实验或队列的多次报告。暴露方案允许三种以上药物，保留组合整体及其成员关系，药对作为派生视图。复方总体结论始终与原方案关联。

每条记录保留原值与标准值、原始结论与结构化结论、抽取和审校状态。证据之间可以是支持、条件不同、真正矛盾、共享研究或后续更正。由此可以研究“数据库冲突”有多少由浓度、制剂或检测方法不同造成，有多少是相同条件下尚不能解释的矛盾，并量化来源依赖对结论的影响。

纵向临床药学院列与共同数据模型

第二类资源围绕真实治疗过程，优先考虑慢病多药患者及其住院抗感染阶段。我拟复用 OMOP 等已有模型的基础结构，增加本项目需要的药学扩展，而不另造一个包揽所有医院信息的标准。[9] 研究对象包括医嘱、发药、实际给药、检验、微生物、临床状态、治疗阶段、停换药、药师干预和结局。

其中最重要的区分是同期开立、实际共同给药、同管接触和药理作用窗口。同日存在两条医嘱只说明共处方；其中一种可能被取消或从未执行。两种输注时间重叠，如果没有管腔与接触记录，也不能判为实际混合。门诊发药可以支持推定暴露，却不能证明每日服用。这些层级分别保存，并注明已观察、推断或未知。

我还会保留发生时间和信息可得时间。例如检验在上午采样、下午回报，上午的决策模型不能提前使用该结果；停换药后的原因记录不能倒灌为先前预测特征。每次治疗变化形成可回放的方案阶段，按具体研究定义时间零点、当时可见信息、候选策略和结局窗口。每个队列因此都有明确规则和可评价的测量性能，能够被复用和独立检验。

连接两类资源的证据桥

连接链包括药品制剂词典、条件匹配、时间对齐和版本追踪。证据资源告诉我们“某些条件下已知什么”，患者队列告诉我们“这个人当时发生了怎样的暴露、哪些信息已可见”。连接结果应返回适用、不适用或缺少关键条件，并给出原因。

我希望比较三种连接方式，即仅按药名、按药对标签、按完整条件与时间连接，评价事实匹配正确率、可判断覆盖率及复核工作量。这既能独立形成证据语义与数据连接研究，也为后续算法提供明确输入。桥接成功首先说明证据语义和患者条件能够可靠对应，预测输出的临床价值再由相应研究评价。

在原型阶段，我可以将以下查询实现为可重复运行的分析，直接展示数据库如何支持科学问题。

研究查询	需要回答的内容
同药对跨条件比较	哪些结论差异由制剂、浓度、溶媒和检测方法解释，哪些仍属未知
共处方与真实暴露对照	有多少医嘱组合真正执行，时间重叠与可确认接触的分母有何不同
慢病合并感染的证据适用变化	新增抗感染治疗及器官功能变化后，哪些既有证据适用，哪些缺少关键条件
机制到患者证据的链路	哪些 P450 关系已有相关人体 PK 与安全证据，哪些仍停留在体外或其他物种

我能够主导的数据提取与工程研究

将证据抽取做成可评价的方法任务

我拟建立从检索、来源登记、解析、实体链接、条件抽取、质量校验到版本发布的完整流程。先形成检索边界和标注手册，使用小范围主题证据试标；随后冻结独立测试集，避免边开发边改变评价标准。所有字段应能回到原文页、表、脚注或段落，不允许模型补造未报告的剂量、浓度和人群。

抽取过程以规则和词典为可解释基线，进一步比较文本模型或经授权的结构化模型辅助抽取。研究重点是联合识别药品、制剂、数值、单位、否定和条件关系。例如“某浓度下未见沉淀”必须保留浓度和检测范围，不能变成普遍相容结论；“不推荐同时使用”也不能遗漏否定或同时这一条件。

实体规范化中，我将重点处理中文通用名、商品名、盐型、复方与具体制剂的映射。精确规则先行，模糊匹配产生候选，无法确定的记录进入人工队列，保留一对多关系。单位换算同时存原值、标准值和公式，不可换算则明确缺失。数据去重不仅比较 DOI，还识别同一实验转载、共享队列和数据库二次转述，避免把相同证据多次计算。

我拟承担数据字典、抽取脚本、链接规则、错误分析和基线实现，组织双人独立标注及争议裁决。评价包括字段级准确性、数值单位完全匹配、条件和否定遗漏、制剂误合并、来源定位正确率及每条有效记录的人时。高风险错误单独统计，与总体 F1 同时报告。除困难样例外，保留概率抽样估计总体错误，避免只评价主动学习挑出的少量特殊记录。

把医院数据整理转成暴露测量研究

我拟设计可复现的院内 ETL，将原始系统记录、标准化事件、质量隔离、药学扩展模型和冻结分析队列分层。每个主键和转换规则都可追溯，增量更新能够处理迟到记录、医嘱修订、取消和退药，不用覆盖原始事实的方式“清洗”数据。大数据处理优先落实到联结正确性、窗口计算、增量更新和版本管理，再依据实际规模选择计算架构。

相应科学问题是，粗略共处方、真实给药和药理风险窗口产生多大暴露误分类，这种误差如何改变患者纳入、事件分母和模型表现。可以在获批样本中盲法核对病历及执行记录，评价暴露识别的敏感度、阳性预测值、时间误差和治疗阶段边界；再比较不同暴露定义下的研究结论。精细化未必在所有任务中都有足够收益，新增字段的采集成本和缺失也需要共同报告。

公开原始文献、说明书和合法可用的第三方资源可以支持首期证据任务。医院数据则在批准后留在受控环境内处理，身份关联和项目标识生成由医院可信服务负责。可公开的成果以字典、代码、标注规范、合成示例和经审核汇总为主。FAERS 作为独立证据来源，用于候选风险与证据覆盖，保留自发报告的偏倚和不确定性，不据报告数量估计发生率。[10]

围绕数据资源形成可由我承担的算法研究

条件与来源约束的证据抽取和冲突识别

第一项算法任务可直接从公开资料启动。输入为正文、表格、脚注和来源信息，输出带原文定位的结构化证据及缺失状态。我要检验，加入领域词典、数值单位校验、条件关系和来源依赖后，能否比规则或无约束抽取减少关键条件遗漏和错误归并。

方法上可采用分层抽取，再以约束校验器检查实体、单位、关系和域边界；主动学习优先选择高不确定或高风险记录送审。对冲突则先区分条件不同与同条件矛盾，再判断是否只是共享原始研究。基线包括规则、通用抽取、无来源去重和无条件聚合；测试按研究家族、来源和时间分隔。词典、说明书和数据库声明可进一步作为弱监督标注函数，逐项记录来源、投票或弃权，比较多数投票与可处理来源相关性的弱标签模型，并始终用独立人工金标准评价。研究成果既包括主题资源，也包括抽取、链接和冲突分析方法，并同时量化质量、成本及可迁移性。

从静态药对预测到条件化兼容性与适用域学习

第二项任务承接 DeepIDC 和 MDCC。输入由药物表示扩展为成分与产品、浓度、溶媒、接触方式、时间和检测终点；输出为相应条件下的预测、适用域及不确定性。这里要检验的是，同药对的条件表示是否真正改善跨条件和跨制剂泛化，而不只是增加参数后提高随机留出分数。

我拟先建立指纹加条件的树模型、药对模型加规则匹配等强基线，再与团队合作比较条件编码、交互表示和多任务结构。不同证据域设置独立输出，必要时只共享部分表示，检验共享机制信息的收益和负迁移。对于未覆盖的新条件，模型可以拒判；评价应同时报告覆盖率和风险，尤其关注离线错误可相容判断，以及在给定风险水平下仍能可靠回答的比例。

数据划分将分别回答已知药物新药对、单药未见或双药均未见的冷启动、新制剂、新条件和新来源等不同问题。同一原始实验的重复记录不能分散到训练与测试。缺条件的记录可用于弱监督或辅助任务，但不得混入条件真值测试集。我愿承担标签体系、数据划分、简单与机器学习基线、消融实验和错误分析，结合团队在分子表示与多任务方法上的积累共同改进模型。

从药对叠加到复杂方案与患者时序建模

第三项任务面向三种以上药物和动态患者状态。输入为实际方案、剂量途径、持续时间、既往暴露及当时可见的肝肾功能 and 病情，输出为预设时间窗内特定结局的风险或需复核程度。科学问题是，完整方案和时序信息能否提供超出“两两 DDI 数量加患者指标”的增量，以及这种增量来自哪类数据和患者。

我拟比较规则风险计数、时间依赖回归、梯度提升与方案集合或超图表示；只有在数据支持时再引入时序网络。模型应区分长期背景治疗、感染阶段新增药与停换后残留作用，保留缺失及观察覆盖信息。药对证据可以作为先验或特征，不能将知识库警告直接当患者结局标签。

评价按患者和治疗阶段分组，进一步做时间留出，条件允许再跨院检验。插补、标准化、特征选择、采样和调参均在开发折内完成，阈值与概率校准使用独立开发校准集，最终测试集保持锁定。除区分度外，报告校准、固定复核工作量下检出率、严重漏报、不同药物负担和肾功能亚组表现，并与不使用复杂暴露信息的模型比较。若复杂模型收益有限，也可通过预设分析识别暴露信息和模型复杂度的适用边界。

复杂联用模式挖掘与风险证据排序

我还希望发挥数据挖掘与 FAERS 研究经验，从三药以上组合及治疗顺序中发现值得验证的关系。输入是保留时间与患者状态的方案阶段，先用频繁项集、闭合模式和序列模式形成候选，再检验其是否提供超出药物数量与两两关系的额外信息。发现阶段与确认阶段分开，预先限制候选规模和多重比较，报告有效暴露机会，而不把常见组合天然排在最前面。

同一模式可能反映病情更重、某种适应证或医院处方习惯。我拟在明确疾病和治疗阶段内比较，控制可测混杂，分析基础药物频率及重叠时间，再在时间留出或独立队列验证。FAERS 可提供另一路候选信号，但须处理报告版本、重复、药物角色和适应证差异；医院事件分析与自发报告信号分别估计，最后在证据层汇合。若三药或顺序模式不能超越两两基线，也能界定复杂挖掘在哪些范围缺乏增量。这项工作由我牵头完成候选生成、数据处理、传统挖掘基线及独立验证设计。

从风险预测走向治疗策略的相对获益研究

风险高不等于改用另一方案就会获益。需要回答治疗选择时，我拟另行明确比较策略、时间零点、纳排标准、随访和估计目标，先判断两种策略是否具有临床可比性和数据重叠，再讨论主动对照队列、目标试验模拟及异质性分析。预测任务与因果问题分别设计，不能将 SHAP 解释或倾向评分调整直接作为因果证明。

方法可从预设分层和可解释模型开始，条件满足时比较双重稳健估计或因果机器学习，并检验暴露误差、停换药、失访和未测混杂的影响。我的重点是将真实用药策略与测量过程落实到分析中，并亲自参与建模和敏感性分析；我与统计学和算法同事共同加强识别假设、推断与方法设计，在数据、方法和临床解释三个环节持续承担研究工作。

慢病多药联用作为持续研究场景

我认为慢病方向适合建立可复用的纵向研究基础。团队已有糖尿病联合治疗综述和疾病风险研究，但疾病识别与治疗方案相对效果不是同一任务。[5-6] 我希望优先围绕一个明确的加药、换药或联合治疗决策，研究在复杂共病和多药背景下，哪些信息影响疗效、安全和治疗持续性。

以降糖治疗为例，首先确定相同治疗阶段的两种临床可接受策略，记录既有治疗、剂量、肾功能、伴随药物、依从性证据和停换原因，再分别分析预设时间窗的血糖改善、特定安全事件和持续用药。具体药物和终点由实际可观察性及研究价值决定，不提前承诺覆盖所有降糖药或并发症。

为使讨论有具体起点，我准备了一项默认配置：成人 2 型糖尿病患者二甲双胍背景下新增 DPP4 抑制剂或 SGLT2 抑制剂，以首次加药处方为起点，研究约 90 日 HbA1c 及记录中的治疗持续性。先评价暴

露和联用信息的预测增量；确认两种策略临床均可接受、识别假设与共同支持后，再另立相对效应分析。附件“示范卡一”列明180日回溯、60至120日随访窗口、特征可见性、失访处理、比较步骤与交付物。这是可调整的研究配置，不是对所有患者的选药建议。

这里最值得研究的是，看似“无应答”是否来自药理反应不足、实际暴露不足、不耐受退出、共同治疗变化或随访选择。它既需要药师定义，也可以形成测量误差和治疗异质性的方法问题。模型比较应分层增加信息，即常规临床指标、实际药物暴露、条件化联用证据，最后才是可选组学。评价重点放在相对获益识别、校准和决策信息的改善，同时分析增加变量和数据成本是否值得。

已有TriMaster及治疗选择模型，为个体化降糖治疗和收益风险整合提供了重要比较基础。[11-12]我希望检验的增量是复杂联用与暴露质量在所选人群和情境中是否具有额外价值，以及模型在哪些患者中失效。若缺少足够随访，可先完成暴露与治疗持续性研究；若候选策略缺乏可比性，则不强行做疗效因果推断。

慢病队列还可以与抗感染阶段自然相连：长期治疗患者入院后增加抗菌药，病情和肾功能随时间变化，既有相互作用证据的适用性也会改变。同一数据基座因此能服务不同任务，但每项研究仍单独定义人群、起点和终点，避免把全部住院经历混成一个大样本。

抗感染治疗作为最贴近我经验的应用入口

抗感染方向可以优先选择一个感染场景和明确复评时点，例如经验治疗转向病原学指导治疗的阶段。研究关注的不仅是“两个抗菌药是否常被联用”，而是患者当时是否存在联合治疗目的、剂量与暴露是否合适、药敏何时可见、哪些风险需要复核，以及随后发生了怎样的安全和疗效结果。

默认示范场景为成人非ICU肺炎患者首次静脉抗菌给药后48小时复评：识别当时需复核的剂量、重复覆盖或监测问题，并单独预测之后72小时的肌酐变化。后者只是连续实验室表型，不当作AKI诊断或DDI因果结局。附件“示范卡二”给出风险集、实际给药起点、双药师标签、可见信息、缺失与出院截断、阶梯模型和可交付研究包，使方案可以直接讨论与试跑。

在资源层，我可以牵头将医嘱、执行、微生物、肾功能、既往抗菌暴露和方案变更重建为治疗阶段。在算法层，先比较常规规则、简单临床模型、药对信息及患者状态加方案表示，检验新增信息能否提高对预设安全事件或可行动药理学问题的识别。ACRM若可合法获得冻结模型和必要说明，可作为合作基线；在缺少原模型接口时，采用可复现公开基线建立新的病程级任务。

标签分成两层。药师独立、对模型输出盲法裁定“是否存在需复核的药理学问题”，用于研究 workflow；客观安全事件和感染疗效按各自定义与时间窗分析，用于患者结果研究。药师判断、建议采纳和患者结果分别记录；器官损伤保留感染进展等其他解释，由预设标准和独立裁定确定事件性质。

我还希望研究模型的拒判与复核排序。面对新病原、少见组合、缺关键检验或明显偏离训练人群的患者，提供可解释的证据不足提示可能比强行推荐更可靠。评价可包括每百治疗阶段的告警量、固定药师工时下高价值问题检出、严重漏报和拒判覆盖。先以回顾性或静默方式确定可靠性，再在必要审批和适当设计下讨论前瞻人机协同；既有研究已开展抗菌处方决策干预，因此简单“模型加药师提醒”不作为创新主张。[13]

条件化输入研究可作为同场景的相邻任务：仅在真实管路、制剂与执行条件可得时研究实际不相容接触；否则先做证据冲突和可回答性基准。新的化学兼容性结论必须依靠合格实验，模型只可帮助选择值得验证的条件组合。

中药学背景可以形成有条件的拓展

中药方向与我的训练相符，但我希望围绕具体产品和真实暴露展开，而不再建一个范围过大的成分靶点图谱。团队已有中药资源和临床到基础再回到临床的研究框架，近期框架也已涉及成分暴露、动态轨迹

和疗效分层。[4,7-8] 我能够进一步承担的是，将这些问题转成产品、处方组成、实际给药、共同治疗、临床表型和安全结局的数据结构与分析。

如果选择一种医院确有稳定使用、团队也认可的制剂或方剂，可在主资源中增加产品与组方版本、厂家剂型、剂量、批次可得性和加减变化。消渴方、消渴丸、玉泉丸及不同来源同名产品分别管理；文献成分、实测产品成分与患者体内暴露也分层保存。

研究可先比较临床指标与实际暴露基线，再检验预先锁定的代谢物或分子 PANEL 是否增加疗效分层信息，并评估跨时间、跨批次的稳定性。共同化药调整、停药不耐受和采样时间均需记录。临床疗效为核心终点，分子变化用于补充解释，并检验两者之间的关联及不一致。

没有标准化产品和可追踪疗程时，不急于采集昂贵组学；需要采样和机制验证时，另行落实伦理、样本质量控制与实验合作。血液代谢组也不能直接替代团队胰岛单细胞数据。这样，中医学背景能够成为产品与暴露研究的专业优势，而不要求我承担当前不具备条件的全链条药物开发。

可以分阶段交出的研究资产

我希望用可检查的研究资产和分析结果推进合作，依据实际证据逐步扩大工作。公开证据资源可以先行，医院队列和应用研究以审批、字段质量及问题可识别性为进入条件。

阶段	我拟主导的工作	可向团队展示的成果	后续判断依据
公开原型	选定主题边界，设计词典、数据模型、证据卡和抽取基线	真实来源小样例、合成患者时间线、可执行查询与许可清单	能否保留关键条件并复核到原文
主题证据集	检索去重、双人审校、抽取与冲突识别、冻结测试	可复用数据集、标注规范、性能和工时报告、条件化基准	证据覆盖和质量是否支持拟研究任务
院内可行性与暴露研究	获批后核查字段、实现 ETL、核对实际给药和治疗阶段	药学扩展模型、数据质量图、暴露误分类分析	是否有足够真实执行、观察覆盖和可靠终点
聚焦算法与疾病研究	选择一个主任务，完成基线、方法改进和独立评价	可复现代码、锁定分析、方法或疾病应用研究	新信息或新方法是否产生可解释的增量
有依据的拓展	跨时间或机构复用，必要时增加组学或前瞻评价	迁移分析、更新资源、获批验证成果	是否值得增加数据、实验与组织成本

作为原型工作量，可以先用约 20 至 30 个成分和约 100 条证据卡试跑字典、审校和查询流程。这是用于验证流程的设计性启动范围；正式研究规模还需结合数据盘点和精度要求确定。随后依据四域原始证据数量、人工成本与研究目标再定资源边界。若关键条件缺失较多，可将证据缺口及其对研究可回答性的影响作为分析内容。

我预期形成的成果是相互支撑的：数据资源及其规范、抽取与链接方法、暴露测量和条件化算法、聚焦疾病的应用研究。每项都需要独立问题与充分证据才考虑成文，避免将同一工作机械拆分。对医院数据不能公开的情况，仍可发布字典、队列定义、质量规则、合成数据和可复现代码，使研究资产可以被团队后续工作利用。

我希望承担的责任与进一步讨论重点

在上述布局中，我愿意对药学研究对象、四域标签、证据条件、暴露定义、数据质量和主要分析结果承担学术责任；在工程与方法上，主导词典和数据库设计、数据提取与 ETL、队列构建、基线建模、误差分析和版本发布。在已有机器学习与数据分析经验上，我希望与团队共同推进更复杂的条件表示、跨域学习和因果估计，并通过亲自完成可复现实现、分析与论文写作持续深化方法能力。

团队既有的分子表示、多任务建模、资源建设和组学研究积累，可以与这些工作深度结合。临床、护理、静配和信息部门帮助核实暴露与数据来源，统计学合作者共同审查估计目标和识别假设，机制或兼容实验由有条件的平台承担。我已有的医院与 GCP 相关资源支持，可用于研究质量管理、终点裁定、随访组织和必要的临床研究协作；这与我的数据和方法工作相互配合。具体数据权限、伦理、样本及外部合作仍按项目分别落实。

我希望进一步与您讨论，团队当前更需要首先推进哪一个可产生独立贡献的任务：条件化联合用药证据及抽取方法、面向慢病复杂联用的暴露与动态分析，或抗感染治疗阶段的方案表示与风险研究。三者共享资源基础，也有各自可独立启动的部分。我的倾向是先用公开证据和原型实现展示技术工作能力，再围绕团队最有价值且条件成熟的临床问题确定主线。

最终，我希望形成的个人研究特点，是能够把药学知识转成严谨的数据与算法问题，并将方法结果重新放回真实用药情境中解释。这样的研究路径既充分利用我的抗感染、中药学、信息药学与药物安全背景，也能为团队留下可持续更新的数据资源、方法工具和后续合作基础。

主要参考文献与资源

- [1] DeepIDC 注射剂组合预测研究 Clinical Pharmacokinetics 2022 原文 DOI
- [2] MDDC 联合用药双任务研究 IEEE BIBM 2024 原文 DOI
- [3] ACRM 抗菌药组合预测研究 International Journal of Antimicrobial Agents 2024 原文 DOI PubMed 摘要
- [4] P450Rdb 原始论文与 TCM2COVID 原始论文 P450Rdb 全文 TCM2COVID 索引
- [5] 糖尿病及其并发症联合治疗获益风险综述 Frontiers in Endocrinology 2023 原文
- [6] 糖尿病风险体检数据研究 Information Fusion 2021 原文 DOI
- [7] 中药疗效关联质量框架与分子轨迹研究思路 药学学报 2021 及 2024 质量框架 DOI 分子轨迹 DOI
- [8] 中药临床基础研究与分子整体状态框架 药学学报 2026 临床基础框架 DOI 中尺度框架 DOI
- [9] OHDSI OMOP 共同数据模型 v5.4 官方规范
- [10] FDA 公开上市后安全监测数据说明 官方说明
- [11] TriMaster 分层降糖治疗随机交叉研究 Nature Medicine 2023 原文 DOI
- [12] 五类降糖药个体化治疗选择模型 Lancet 2025 原文 DOI
- [13] INSPIRE 抗菌药决策干预随机试验 JAMA 2024 原文 DOI