

# 有效丰富真实世界证据的策略

## 富集研究概述

随着医药行业愈来愈关注使用真实世界数据（RWD）形成真实世界证据，在对需要收集一手数据的新研究进行规划时，人们常常会提出这样一个问题：能否从既有RWD中获取一些研究所需数据以实现研究目的？

目前，越来越多研究项目在数据收集时，将为该研究专门收集的一手数据与既有的患者层面的真实世界数据相结合，这样的研究被称为“富集研究”或“混合研究”<sup>[1-4]</sup>。富集研究整合了二手数据（如电子病例记录 EMR 或医保支付数据）和直接从研究中心、患者或医疗机构收集的一手数据，不仅增加可供研究的数据量，还能在单个研究中解决更多的科学问题<sup>[5]</sup>。

包括美国食品和药物监督管理局（FDA）和欧洲药品管理局（EMA）在内的监管机构，也愈来愈多地将药械获批前后的研究中产生的 RWD 作为支持监管决策的考虑依据。近期，FDA 发布了利用电子病历记录和医保支付数据支持药品和生物制品监管决策的指南草案。该草案对于如何在整个数据生命周期中联接 RWD 并确保数据质量提出了具体建议<sup>[6]</sup>。

### 富集研究方法可应用于多种研究设计，并有助于提高效率

富集数据收集方法可应用于一系列研究设计和研究目的，包括虚拟临床试验、疾病和产品注册登记研究、扩展研究、

上市后安全性研究（PASS）、上市后有效性研究（PAES）、实效性随机对照研究和药物使用研究（DUS）等（图 1）。

富集研究还可以提高研究数据收集的效率。与传统的、仅仅基于研究中心的观察性研究相比，从既有数据来源获取一些研究数据元素可能会减少从研究中心和/或患者收集一手数据的工作量。减少一手数据的录入也会影响研究的执行；在某些情况下，通过减少需要溯源核查的数据元素数量，可以减少数据监察时间，或者借由通过二手数据库的更新，获得一些所需的纵向数据进行长期的随访，减少联系患者次数。

在 IQVIA 的一项初级市场调研中，研究人员向来自美国、德国、英国、法国的 210 名医生描绘了两种观察性研究情景：富集和非富集。非富集情景假设所有研究变量通过病历回顾手动收集。富集情景假设利用现有的病历设计出较短的病例报告表（CRF），使医生的手动数据录入时间减少 50%。有 66% 的受访医生认为，通过整合二手数据，缩短观察性研究数据收集所需的时间，更能提升他们参与研究的意愿<sup>[7]</sup>。

在某些情况下，二手数据（如 EMR 数据或登记数据）可用于预填写电子病例报告表（eCRF）。使用这种方法，不仅减少了人工转录，提高数据质量，还可以简化获取研究数据的过程。

图1：可使用富集数据收集方法的研究设计



这些研究设计通常的证据需求按以下颜色表示：医疗事务（青色）、监管（桔色）、HEOR/HTA（深蓝色）

## 有效丰富真实世界证据的策略

过往已有许多证据显示整合电子健康记录（EHR）和研究系统进行数据采集是可实现的<sup>[8]</sup>。近来监管单位对数据库间信息互通的授权，更提升了将 EMR 与研究系统进行整合作为解决方案的可能性。

### 一手数据和二手数据来源的类型

富集方法中，首要步骤是确定所需数据的来源。数据源的最佳选择取决于具体的研究问题、关注的区域及医疗现状。这些因素决定了整合一手数据和二手数据在患者层面连接的策略。

一手数据可分为两大类：患者报告数据和医生或研究中心报告数据。这两类一手数据可以使用多种临床结局评价（COA）工具收集。患者报告数据通常可用患者报告结局（PRO）表收集，也可以由监护人等非临床观察人员或基于临床表现评价工具来收集<sup>[9,10]</sup>。近年来，随着数字技术的发展，越来越多的联网设备和应用程序都可以用来收集患者报告数据。研究中心报告数据通常采用 eCRF 或研究指定的其他的数据录入方式，由医务人员或其他研究中心工作人员采集和记录。

富集方法可使用多种来源的二手数据，最常见的来源包括：电子病历记录（EMR）、医保支付数据和患者注册登记数据。这些数据源的优势体现在各个方面，包括可用的变量、医疗服务及其覆盖人群、临床数据的细节程度以及医疗系统或支付方的考虑因素。这些方面的权重因研究而异。在选择合适的二手数据源过程中，还应考虑：

- 当地隐私保护法规
- 二手数据源的管理和使用规定中，关于连接一手和二手患者数据方法的规范
- 连接不同数据源时如何保护患者隐私
- 使用和连接二手数据的患者知情同意
- 伦理审查委员会（IRB）伦理审查和相关指南

### 可联接的数据结构和格式

标准化的数据格式、编码和结构有利于多个来自不同医疗系统的二手数据源互相整合，并将其应用于研究中。

近期美国的政策改变（如美国国家卫生信息技术协调办公室ONC发布的“治愈法案最终规则”<sup>[11]</sup>）正在加速EHR数据的可及性和信息互操作性，预期其他国家也将积极跟进。该法规于2020年3月发布，详细规定了基于HL7快速医疗互操作性资源（FHIR）的医疗信息互通技术标准，同时还建立了美国互操作核心数据集（USCDI）；标准化的健康数据分类和数据形式可进一步实现全国范围内电子健康数据的互操作化<sup>[12]</sup>。

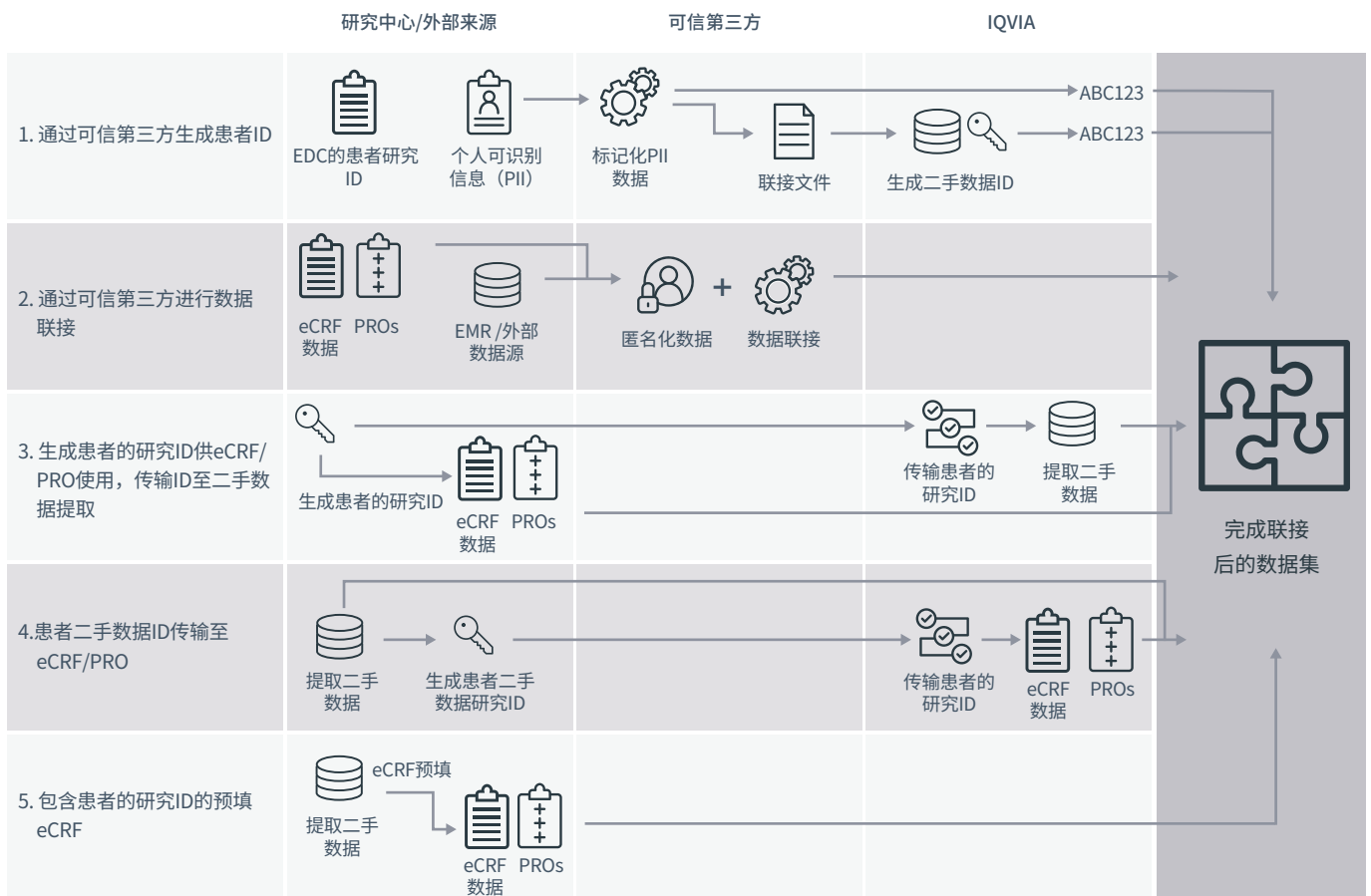
### 联接研究数据源的策略

在患者层面连接一手和二手数据需要考虑技术可操作性、患者隐私保护和数据整合方法（即整合不同来源的数据并提供用户一个整合的数据集）。图2展示了几种数据联接策略，这些策略可以简化下游数据管理步骤，以整合收集的一手数据和二手数据。考虑到技术系统的差异和二手数据源的不同要求，这些数据联接策略允许相关各方统一操作流程，同时保护患者隐私。在发布研究结果时，还需严谨报告数据联接的方法、对联接结果的确认和验证，以及联接的准确性<sup>[13]</sup>。



图2. 数据联接策略

- 考虑到当地数据收集技术的可行性以及数据隐私保护法，富集研究设计可采用不同的数据联接方法
- 所有数据联接策略均应保证患者隐私安全
- 受信任的第三方可按照数据所有者、地方政府和伦理委员会的要求，执行/协调数据联接过程，并克服某些外部数据提供商的合同限制
- 将二手数据中患者的研究ID集成到eCRF/PRO，有助于已收集数据的直接确定性联接



# 有效丰富真实世界证据的策略

## 富集研究案例分享

根据数据来源的可用性和研究需要，富集研究会使用不同类型的一手数据和二手数据。表1展示了针对不同研究问题的四类富集研究示例。

表 1：使用富集方法进行患者层面数据联接的研究示例

	DISCOVER 研究	“WE’ VE GOT YOUR BACK” 研究	CARE 研究	SNDS研究 <sup>1</sup>
<b>证据需求</b>	了解一线治疗失败的2型糖尿病（T2D）患者的特征、治疗方式，并评价这些方案对改善血糖控制和长期生活质量（QoL）的影响	评估腰背痛患者在随访期内的生活质量（QoL）和疼痛变化情况	描述医院和急性护理环境之外的COVID-19状况。确定高风险亚组、严重程度分布、治疗模式和疗效，以及与轻症或快速康复相关的因素	研究法国T2D患者中慢性肾病（CKD）的流行病学情况，并对患者的人口学及临床特征、患者结局（临床未满足需求）、治疗模式、资源使用和成本等特点进行分析和描述
<b>一手数据</b>	ePRO和eCRF	ePRO	ePRO	不适用，本研究未使用一手数据。 注：IQVIA可将ePRO和eCRF与SNDS进行联接
<b>二手数据</b>	<ul style="list-style-type: none"> <li>• 北欧（丹麦、挪威、瑞典）：IQVIA 电子病历提取工具和国家登记信息</li> <li>• 法国：Disease Analyzer (IQVIA EMR Panel)</li> <li>• 加拿大：IQVIA Electronic Medical Record Extraction</li> </ul>	Egton 医疗信息系统 (EMIS Health) EMR	LRx/Dx (IQVIA 医保支付数据)	<ul style="list-style-type: none"> <li>• SNDS (法国健康保险数据库)</li> <li>• France Disease Analyzer (IQVIA EMR) 数据库</li> </ul>
<b>附加价值</b>	<ul style="list-style-type: none"> <li>• 通过EMR和登记信息的数据关联，填补疾病知识不足</li> <li>• 数据驱动的研究中心和患者识别 (EMR界面)</li> <li>• 大样本纵向队列，使用富集方法纳入超过1000名患者</li> </ul>	<ul style="list-style-type: none"> <li>• 便捷的移动应用程序实时收集患者症状数据</li> <li>• 在较少依赖医师的前提下，有效识别和招募患者</li> <li>• 减少一手数据收集的负担</li> </ul>	<ul style="list-style-type: none"> <li>• 与药房支付数据的联接可以提供配药的宝贵细节，且无需耗费患者更多时间</li> </ul>	<ul style="list-style-type: none"> <li>• 使用EMR 数据了解T2D 患者的肾功能损害特点</li> <li>• 基于EMR-SNDS样本开发CKD分期的预测模型，并将其应用到SNDS的所有数据</li> </ul>
<b>规划的考虑因素</b>	由于二手数据源的异质性和限制性(例如北欧登记信息数据的滞后)，会导致复杂的数据流和数据整合	在全科医生的日常工作中，整合不同技术 / 软件，有效、及时地招募患者。	二手数据因国家 / 地区而异（如美国的 LRx/Dx 支付数据）	SNDS数据的授权大约需要4到6个月的时间，而实现数据访问则需要额外的4到6个月

<sup>1</sup> 尽管IQVIA在法国具备实施富集方法的能力，但SNDS研究案例并没有采用富集研究的方法（一手数据和二手数据的联接）。此处为说明性示例。

## DISCOVER 研究

### 证据需求

许多地区都缺乏关于 2 型糖尿病 (T2D) 治疗和控制的 RWD。现有的二手数据作为单一来源无法包括患者完整病程的详细数据和其他关键指标,如糖化血红蛋白、血压和体重可能不完整或难以获取。

### 解决方案

DISCOVER 研究是一项为期 3 年的全球性、前瞻性、观察性研究,旨在补充 T2D 管理方面知识的不足,尤其是对于一线治疗失败的患者<sup>[4]</sup>。

该研究的主要研究目的是描述使用二线降糖治疗的 T2D 患者特征、疾病管理的变化和临床结局。研究在五个国家(加拿大、法国、挪威、瑞典、丹麦)采用了富集方法,将 EMR, 注册登记数据库(如可用)与简化的 CRF 和患者报告数据(QoL、治疗满意度,经验证的抑郁量表)相关联<sup>[14]</sup>。通过富集方法, IQVIA 在这些国家共招募到 1005 名患者。

数据来源和联接策略在不同区域存在差异:

- 在法国,二手数据从 IQVIA 专有的 EMR (Disease Analyzer France) 中提取,一手数据在研究随访期间按规定的时间间隔收集。医生完成填写定制的 eCRF 后,会为每个研究患者生成一个匿名的研究编号,医生在研究访视期间还会将纸质 PRO 发放给患者(根据患者的研究编号相应地对 PRO 预先编号)。Disease Analyzer 的 EMR 数据和 eCRF 数据进行确定性关联。纸质 PRO 经扫描电子化

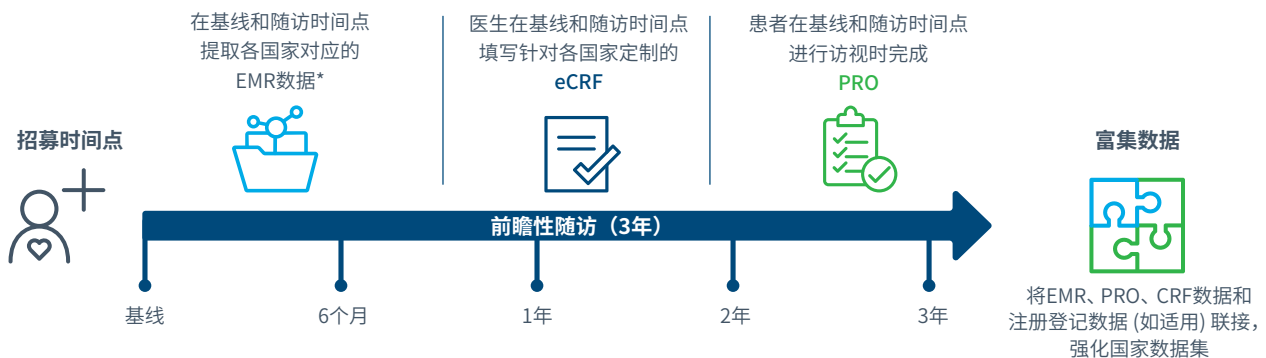
后被传输给外部合作方,以录入研究数据库,通过研究 ID 即可联接其他数据源。

- 在加拿大,该研究采用 IQVIA 的 EMR 提取技术工具获取参与研究中心的数据。所有临床信息均使用 EMR 采集,不会收集额外的医生报告数据。通过电子化流程收集 PRO,患者会收到一封包含研究问卷网址联接的电子邮件,可在研究方案规定的时间点完成问卷调查。
- 在北欧多国,该研究采用 IQVIA 的 EMR 提取技术工具获取参与研究中心的数据,使用为各国定制的 eCRF 按照预定的时间间隔获取数据,同时还收集了纸质版的 PRO。通过研究者另外收集的患者的个人识别信息实现数据联接。挪威和瑞典将 EMR、PRO 和 eCRF 数据发送给各国卫生管理部门。卫生管理部门将研究数据与国家登记信息数据联接,并在最终数据集中删除特定的个人识别信息,创建新的患者 ID 进行匿名化处理。在丹麦,由于在研究过程中法律的变更,研究数据和 EMR 数据与国家登记数据联接的步骤未能完成。

### 研究结果

通过将患者层面的一手数据和二手数据联结,富集方法能够获取申办者感兴趣的所有研究终点。在这项多国研究中,通过对每种研究方法收集上来的全部数据变量进行比较,我们发现,富集方法相较于非富集方法可以减少高达 27% 的 eCRF 数据录入量,缩短了参与研究的医生和研究中心其他人员收集数据的时间<sup>[15]</sup>。

图 3: DISCOVER 研究示意图



注\* 在前瞻性随访期间,为了进行中期分析而进行了数据联接。由于一些国家 EMR 数据存在滞后,数据提取时间点存在差异。



## 有效丰富真实世界证据的策略

### “We’ ve Got Your Back” 研究

#### 证据需求

这项极具创新价值的研究是一项观察性、前瞻性队列研究，研究对象是在英国初级卫生健康场所接受治疗的急性或慢性腰背痛患者。这项研究的目的是在英国试点富集数据收集模式，同时评估患者的生活质量(QoL)、腰背痛症状的变化，并量化其对生产力和经济负担的影响。

#### 解决方案

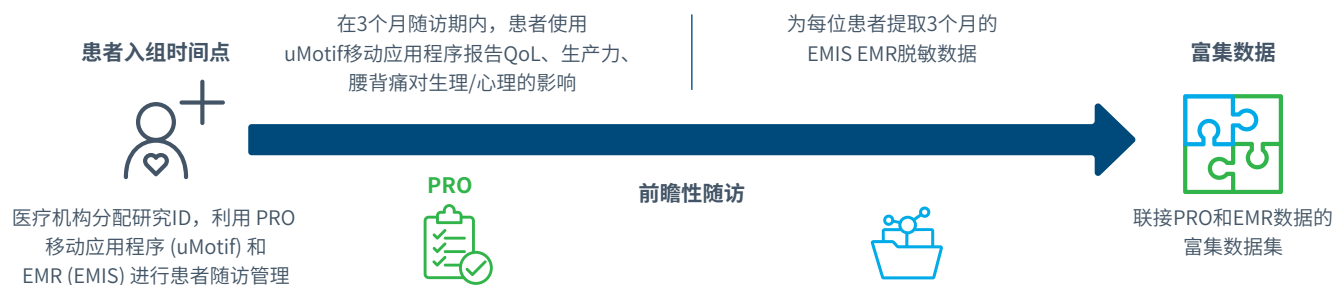
该研究由英国国家卫生研究院 (NIHR) 临床研究网络、英国利兹大学、uMotif 和 Egton 医疗信息系统 (EMIS Health) EMR 共同合作完成。在该研究中，EMR 会根据纳入标准自动识别符合条件的患者，协助医生进行招募。该研究共招募了 135 名患者。患者被招募后，EMR 中会生成并记录一个唯一的患者识别 ID，患者也会收到一个可用于下载 uMotif 移动应用程序的连接。

下载 uMotif 应用程序后，患者输入自己的 ID，就可以提交参与研究的电子知情同意书。通过 uMotif 应用程序，患者在 3 个月随访期内报告 QoL、生产力和腰背痛对生理 / 心理的影响数据。在研究结束时，通过患者 ID 联接患者 EMR 数据和 PRO 脱敏数据。

#### 结果

WGYB 研究整合了由患者直接提供的数据和他们的 EMR 数据。从 EMR 数据中收集到的数据超过了研究所需全部数据的三分之一，减轻了医生的负担。研究完成时，有 42% 的参与医生报告称，他们在研究中的投入低于预期，有 75% 的医生表示，会推荐使用应用程序收集研究知情同意书<sup>[16]</sup>。对于患者而言，uMotif 应用程序的优点是能够通过记录症状和追踪疾病趋势或恶化情况，有效地协助医患管理病情。与传统的日志报告方法相比，直接收集实时产生的患者数据也有助于得到更准确的症状信息<sup>[17]</sup>。

图4. We’ ve Got Your Back 研究



### CARE 研究

#### 证据需求

COVID Active Research Experience (CARE) 研究旨在描述在医院和急诊护理环境之外的 COVID-19 状况。这项直接面向患者的纵向、观察性富集研究，收集了与健康状况相关的信息，包括症状进展和严重程度、疫苗、非处方药和处方药的使用以及膳食补充剂等。该研究目的在于研究疫苗的安全性和有效性，描述 COVID-19 的长期特征，确定哪些因素可降低 COVID-19 的暴露风险或促进更快康复。

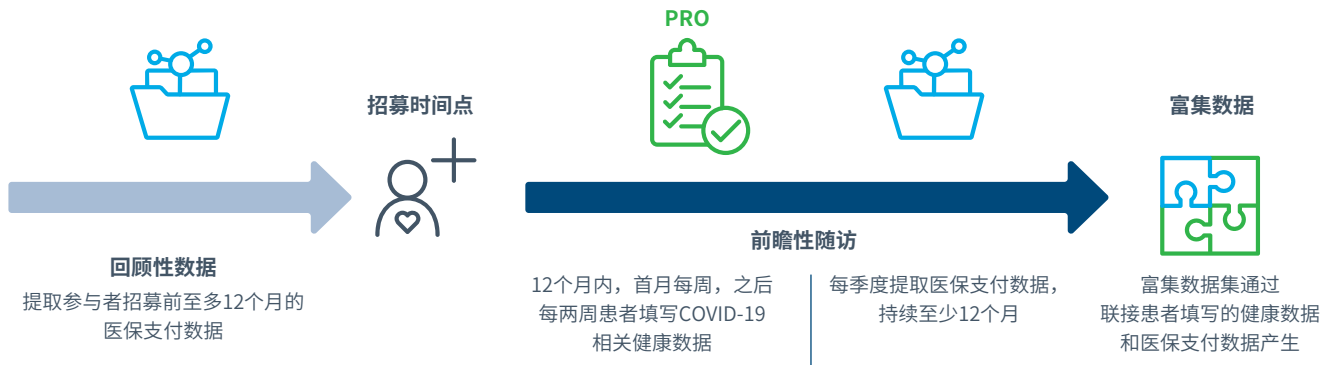
#### 解决方案

CARE 项目在美国正在进行中。成年志愿者被邀请在 12

个月内每两周提供简要信息。目前，招募仍在继续 ([www.helpstopCOVID19.com](http://www.helpstopCOVID19.com))。

2020 年 4 月至 2021 年 6 月期间，有 3 万余人被招募。这些来自社区的志愿者签署了知情同意书，登记参与本研究。他们通过填写推送到智能手机、平板电脑或电脑的加密电子问卷，上报与 COVID-19 相关的经历。知情同意书还允许将经去识别化处理的、个人生成的健康数据与医疗和药房保险支付数据联接起来，以整合更多有关医疗服务利用、处方药品使用 (包括 COVID-19 的治疗) 和健康结局的信息。患者填写的健康数据与医保支付数据进行了确定性联接，并使用了专有算法和可信的第三方，以保护参与者身份并遵循患者数据隐私规定。全球 COVID-19 疫情的流行病学和患者的真实世界情况还在不断变化，富集方法使得 CARE 研究能够以灵活的方式产生相关证据。

图5. CARE 研究



## SNDS 研究

### 证据需求

法国健康保险数据库（Système National DesDonnées-de Santé, SNDS）全面登记了法国近6600万居民的所有医保支付信息。其主要关注领域包括健康监测（包括疾病流行病学）、分析（辅助）医疗实践、政策实施监测、卫生经济学和药物流行病学<sup>[18]</sup>。然而，SNDS 缺乏详细的临床数据以及临床辅助检查的结果。因此可以通过与全科或专科医生在常规诊疗中填写的 IQVIA EMR 数据进行概率性联接，以获得详细的临床数据，从而弥补 SNDS 的这一局限性。

SNDS 研究是一项观察性、回顾性研究，旨在评估 2014 年至 2018 年期间法国 T2D 患者中 CKD 的年患病率和发病率。这项研究案例联接了两个二手数据源：SNDS 数据和 IQVIA EMR，因此作为示例被纳入到本文中（并不是富集方法）。但 SNDS 具备与一手数据源进行联接的强大功能，这可以通过下文描述的解决方案来实现。

### 解决方案

为了实现联接，IQVIA 向 CNAM（负责 SNDS 数据处理的第三方）提供需要与 SNDS 数据进行联接的 EMR 数据集，以及一个专用联接数据集，其中包含用于患者匹配的个人识别信息（PII）。

此处使用的联接方式是概率联接，针对每个患者 ID，使用患者的特征变量（性别、年龄、居住地区）以及看全科医生的就诊日期进行联接：需要至少 3 次不同的就诊日期才能够对患者进行匹配，就诊日期越多，联接也就越准确。

在该研究中，EMR 可提供包含了实验室检查结果（如血清肌酐、估算的肾小球滤过率）的丰富临床数据，用以评估肾功能、确定 CKD 分期，并估计 CKD 在 T2D 患者中的流行病学情况。次要研究目的是描述患者的人口学特征、病史、医疗资源使用和相关费用，以及不同 CKD 分期患者的用药管理与心血管事件发生之间的关系。探索性研究目的是使用 EMR-SNDS 联接数据集开发预测模型，研发一种算法，用以识别法国 SNDS 数据库里全体 T2D 患者中的 CKD 患者，和 / 或对 CKD 进行分期。

除了二手数据集 SNDS 和 IQVIA EMR 之间的联接外，IQVIA 还可支持 SNDS 数据与患者和 / 或医生报告的数据进行富集，相关工作流程完备，工作人员也训练有素。在富集研究中，获得患者知情同意后，IQVIA 会从电子数据采集（EDC）平台上收集一手数据和 PII 变量。SNDS 收到每个患者 ID 对应的 PII 变量后，会通过概率性联接在国家数据集中识别出研究患者。每个患者 ID 对应相关医保支付记录将被传回至经过认证的 IQVIA 安全云端，再通过患者 ID 与一手数据（PRO 和 / 或 CRF）进行整合。

## 讨论

富集方法的潜在价值和益处，可结合以下因素进行考虑：访问和联接二手数据的可行性（考虑到全球不同的与隐私保护相关的法律和法规）、技术能力，以及该方法是否可以在研究纳入患者数量、研究中心数量和国家数量几方面带来效率提升。在联接二手数据源的启动和准备阶段，如果研究中的患者数量较少或在某国家或某个数据源的患者数量较少，则通常无法保证在技术、数据访问费用、隐私评估等方面的效率。因此，在启动计划的研究之前一般会建议先进行可行

# 有效丰富真实世界证据的策略

图 6. SNDS 研究案例



请注意：上图中关于与一手数据联接的内容描述了IQVIA目前具备的联接功能，而本SNDS研究中并未使用该功能。此处为说明性示例。

性评估，尤其需要关注研究目的和证据需求，以解决富集方法会涉及的任何问题和风险（关键要点如下）。

### 富集研究可行性的关键要点和考虑因素：

- 确定关注地区和患者群体的二手数据来源
- 满足研究目的的二手数据中数据各要素的可及性和完整性，或者相关变量是否可以从医生或患者处获得
- 患者层面的二手数据和一手数据可通过确定性或概率方法联接，且需遵循当地法律，尤其是隐私保护
- 用以支持数据采集和联接技术解决方案的适用性（如：用于标记的患者联接软件、支持 FHIR 进行 EHR 数据提取和整合的平台）
- 可通过使用二手数据提高效率的研究中心和患者的数量及分布（多国研究需按国家划分）

基于当前趋势，富集研究面临的进一步挑战和机遇包括：

**保证 EHR 信息互操作性 / 系统成熟度**——为了未来能够在已整合的传输网络之外更有效地启动富集研究，并保证其成本效益性，推进建立在美国和其他国家的信息互操作性标准至关重要。利用 FHIR 应用程序编程接口 (API) 等先进的技术平台的设计，将有助提升相关功能并大大缩短访问患者电子健康数据时启动和准备所需的时间。

**通过患者访问健康记录**——随着信息互操作性的发展，法规也赋予患者对个人 EHR 的直接所有权和访问权。通过患者可直接从多个医疗机构访问其 EHR，从而部分弥补了复杂医疗卫生系统的碎片化问题。

**提高医疗机构 EMR 系统中研究工具的集成度**——在医疗机构的 EMR 工作流程中嵌入支持识别符合研究条件的患者的工具，可以使得研究者在患者的常规就诊时即可对符合条件的患者进行招募。创建集成的 EMR 工具有助于更有效、更完整地获取 COA，例如风湿病患者的疾病活动度测量，也有助于在富集研究中更多的使用 EMR 数据。

**向患者和照顾者提供数字健康应用程序**——移动设备上有着越来越多方便易用的数字健康应用程序可用于研究数据收集和 / 或疾病管理，这为整合电子健康数据和患者报告的一手数据提供了更多的可能性，以实现富集研究的初衷。

更多信息，敬请联系：

#### 谢洋 博士

IQVIA艾昆纬副总裁、亚太区卫生  
经济与真实世界研究负责人  
yang.xie@iqvia.com

#### 殷铮 博士

IQVIA艾昆纬大中华区真实世界  
研究负责人  
angela.yin@iqvia.com



本文参考文献请扫描二维码  
进行查阅