

基于疾病登记的真实世界数据应用 指导原则（征求意见稿）

2023 年 11 月

目录

一、引言	1
二、疾病登记的定义	1
三、疾病登记数据的应用场景	3
(一) 产生补充或关键临床证据	4
(二) 为临床试验设计提供依据	4
(三) 作为单臂试验的外部对照产生临床证据	4
(四) 用于上市后安全性监测和/或有效性评价	5
(五) 丰富临床试验的实施方式	5
四、数据库的建立	6
(一) 疾病登记的建立过程	6
(二) 疾病登记计划与设计	7
(三) 数据库构建	13
(四) 疾病登记数据的获取、录入与质量控制	15
(五) 数据库成熟	18
(六) 数据库融合与扩展	18
(七) 数据安全与伦理	20
(八) 早期融入 PFDD 理念	21
五、从临床研究角度对疾病登记数据的评估	22
(一) 研究目的的实现	22
(二) 疾病登记建设的规范性	22
(三) 研究人群代表性	23
(四) 随访期、随访问隔和时间窗	23
(五) 数据质量及其控制	24
(六) 数据使用的合规性	24

六、疾病登记数据的递交及与监管机构的沟通交流	25
(一) 疾病登记数据的递交	25
(二) 与审评机构的沟通交流	25
参考文献	26
附录 1: 词汇表	30
附录 2: 中英文词汇对照	34

基于疾病登记的真实世界数据应用指导原则 (征求意见稿)

一、引言

使用真实世界证据 (real-world evidence, RWE) 支持药物监管决策日益受到重视。目前, 开展真实世界研究 (real-world study, RWS) 面临的首要挑战是真实世界数据在质、量和可及性方面难以有效支持监管决策所需要的临床证据。疾病登记 (disease registry) 通常根据登记目的进行前瞻性设计, 针对性地、系统地、适时地、较长期地收集特定人群的疾病及其临床特征数据, 实施中注重质量控制, 其数据质量相对较好, 是真实世界数据 (real-world data, RWD) 的重要来源之一。如何充分利用现有的质量较好的数据资源, 以及如何战略性布局建立质量优良的疾病登记, 对于产生充分的真实世界证据并用于支持监管决策至关重要。

本指导原则将阐述如何建立疾病登记, 以及如何评估和使用已有的疾病登记数据, 旨在为基于疾病登记的真实世界数据应用提供指导。

二、疾病登记的定义

“登记”一词所对应的英文“registry”, 同时有“注册”、“注册登记”等不同翻译。为了与临床研究中的药物注册 (drug registration) 相区分, 本指导原则采用“登记”这一词汇。

23 登记是指根据事先确定的登记目的，有组织的、系统的、
24 根据所定义的随访/观察时限和时点中心化地收集与人口学、
25 疾病、暴露、诊疗、结局等相关的患者个体层面的纵向数据
26 所形成的数据库，也称为登记系统，登记数据库是登记形成
27 的最终成果。登记通常是前瞻性地收集数据的，也可以按照
28 前瞻性设计方案融合既往产生的数据。基于登记所产生的数
29 据开展的研究，称为登记研究（registry-based study）。

30 登记又称为病例登记（patient registries），针对患者的数
31 据进行收集，包括患者的疾病状况，接受的治疗或所处的暴
32 露，以及使用的药品。登记研究根据研究的侧重点不同可分
33 为疾病登记、医疗产品登记（product registries）和健康服务
34 登记（health service registries）。其中疾病登记可以是单一疾
35 病登记（如乳腺癌登记、高血压登记、戈谢病登记），也可以
36 是某一类疾病登记（如癌症登记，心血管病登记），还可以是
37 国家或地区规模的一类或多类疾病登记（如中国国家罕见病
38 注册系统）。由于登记有多种含义，考虑到药物临床研发与评
39 价的实际，如无特别说明，本指导原则仅针对单一疾病登记
40 进行阐述。

41 登记数据的最大优势是根据特定研究目的前瞻性地建
42 立特定疾病人群队列，针对性地确定所要收集的数据以及收
43 集数据的方法，能够形成较完整的纵向数据，包括产品临床

44 研发全生命周期数据的捕获，以及体现以患者为中心的药物
45 研发（patient-focused drug development, PFDD）理念的患者
46 体验数据（patient experience data, PED）的收集，如患者报
47 告结局（patient-reported outcome, PRO）数据的收集等。在
48 采集数据方面可以根据前瞻性设计建立结构化和标准化的
49 数据库，并利用性能不断提升的数字化工具高效获取数据。

50 疾病登记在建立过程中如果设计不严谨或执行质量不
51 佳，均可能导致登记数据应用的局限，特别是不同偏倚的产
52 生。例如，设计时重要变量的遗漏或缺乏严格的规定（如测
53 量方法、随访周期和时间点）会导致重要信息的缺失或不可
54 用，引起多种偏倚或影响对某些偏倚的校正；入选/排除标准
55 不当会影响目标人群的代表性或产生选择偏倚；病情较重者
56 比病情较轻者更倾向于入组而导致人群代表性受限；失访率
57 偏高可能导致的选择偏倚；持久的质量控制得不到保障导致
58 数据质量问题，所引发的多种问题和偏倚风险；多种原因导
59 致的不同研究中心间的数据异质性问题（如医疗习惯与治疗
60 水平、社会经济背景、自我疾病管理能力）可能引起多种偏
61 倚，并带来研究设计和结果解释上的挑战，等等。

62 三、疾病登记数据的应用场景

63 登记研究具有较广泛的应用场景，包括发现疾病的过程，
64 了解治疗和结局的变化，探查预后的影响因素，刻画诊疗习

65 惯，监测安全性风险，评估治疗效果，等等。在药物研发领
66 域，基于真实世界诊疗环境的疾病登记数据可用于上市前和
67 上市后评价，通常在儿科、罕见疾病、肿瘤等疾病中应用较
68 多，常见应用场景包括但不限于以下情形：

69 **（一）产生补充或关键临床证据**

70 基于疾病登记数据产生的真实世界证据可作为补充或
71 关键临床证据支持药物监管决策。例如，在儿科、罕见疾病、
72 肿瘤等疾病登记中收集超说明书使用的数据，可能用于支持
73 新增适应症的监管决策。

74 **（二）为临床试验设计提供依据**

75 利用登记数据可以为临床试验设计提供一定依据，例如，
76 样本量估计时参数的确定依据，病例选择的入选/排除标准的
77 制定，体现临床价值且能够灵敏反映临床疗效的研究终点的
78 确定，关键的观测时间点、时间窗和间隔的选择，最小临床
79 意义差别的探索，患者就医行为和治疗偏好的刻画等。此外，
80 如果疾病登记有相关基因或生物标记物的完整记录，将有助
81 于精准定位试验的目标人群。

82 **（三）作为单臂试验的外部对照产生临床证据**

83 单臂研究设计通常应设置外部对照，外部对照采用的形
84 式有历史对照、平行对照等，其来源和选择至关重要。外部
85 对照对真实世界数据的质量和样本量均有较高要求，外部对

86 照的基线和影响结局的关键变量（研究相关的治疗除外）应
87 与试验组相近且数据缺失应在可接受的范围内，如果有设计
88 和执行良好的高质量疾病登记数据，可以被优先考虑作为外
89 部对照，因为它相较电子病历等其它数据源数据更容易满足
90 外部对照的要求且在数据的获取和治理方面更具有可行性。

91 **（四）用于上市后安全性监测和/或有效性评价**

92 疾病登记具有长期性、及时性和纵向性特点，通过对药
93 品上市后的长期监测和数据不断积累更能体现其优势。疾病
94 登记随着病例数量的增加和观察时间的延长，更有利于全面
95 监测药品的安全性事件和发现罕见不良事件，以及评价在随
96 机对照临床试验中难以获取的较长周期或罕见的临床终点。
97 基于疾病登记，还可以设计不同的上市后产品的产品登记。
98 通过对药品上市后登记数据的分析，不仅可以洞察真实诊疗
99 环境下的药品疗效，还能支持不同的治疗策略（如剂量、疗
100 程、用药频率等）下的疗效评估，为指导临床实践提供依据，
101 还可能为说明书修订提供依据。

102 **（五）丰富临床试验的实施方式**

103 利用疾病登记还可以配合干预性临床研究的实施，显著
104 提高临床研究的效率。例如对于病例数足够大的疾病登记系
105 统，通过对病例的临床特征分析，根据研究方案中的入选/排
106 除标准快速筛选病人；在病人招募的基础上，还可实施随机

107 化，例如心脏病登记系统 **SWEDEHEART** 就被用于药品和器
108 械研发的临床试验中对招募病人进行随机化分配，被称为基
109 于登记的随机对照临床试验（**registry-based RCT/ register-**
110 **based, R-RCT**）。

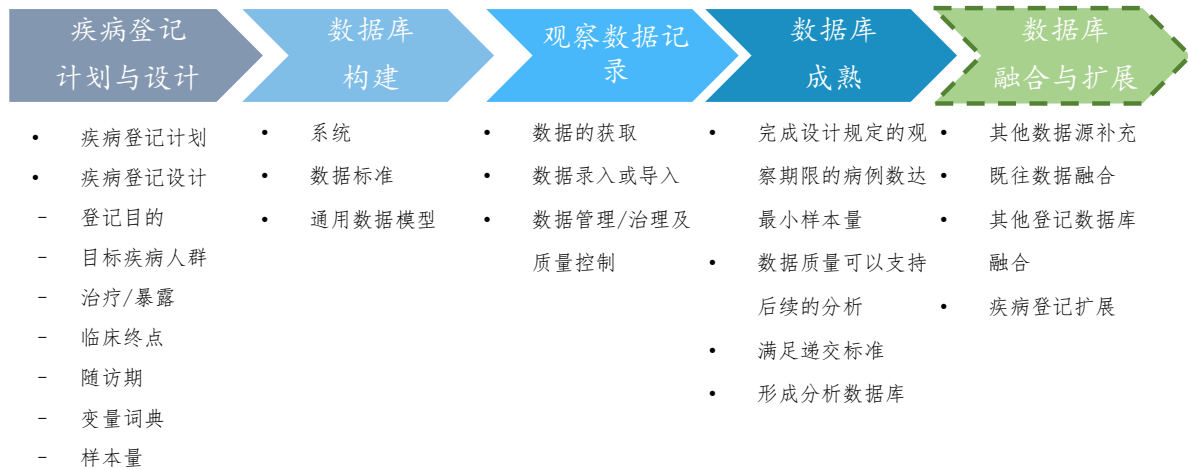
111 登记数据有时会涉及到以上多个应用场景，例如基于疾
112 病自然史登记系统不仅可以为研究设计提供依据，如入选/排
113 除标准、干预阶段、研究终点、随访时限及时间点等的合理
114 选择，同时可以在上市后研究中帮助精准定位特定的获益人
115 群，识别和开发生物标志物，还可以作为单臂试验的外部对
116 照等。罕见疾病自然史登记系统如果设计良好，可以为罕见
117 疾病研究提供有力支持，包括获得人口学、疾病特征、疾病
118 进展等基本信息。

119 **四、数据库的建立**

120 **（一）疾病登记的建立过程**

121 疾病登记的建立过程如图 1 所示。首先制定疾病登记计
122 划，并根据整体计划进行具体设计。接着根据设计的要求构
123 建数据库。建库完成后开始收集和录入数据，并在这一过程
124 中进行数据的质量控制，当数据库成熟后可用于分析目的。
125 疾病登记的数据库还建议具备数据的融合与扩展功能，一方
126 面可以融合既往数据和其它登记数据；另一方面也可以根据

127 研究需要和数据获取的可能，在疾病种类、观察期限、变量
128 等方面进行扩展。



129 **图 1. 疾病登记的建立过程**

130 **(二) 疾病登记计划与设计**

131 **1. 疾病登记计划**

132 制定疾病登记计划需要依次考虑若干关键问题，包括但
133 不限于：

134 • **明确登记目的：**应明确开展疾病登记研究的整体目的
135 是什么，例如哪个疾病领域、何种规模、基于此疾病登记可
136 以开展哪些研究以及这些研究可能回答的临床问题。

137 • **选择疾病领域：**可以是单一疾病、同一类疾病或多类疾
138 病，下文主要围绕单一疾病展开。

139 • **定义目标疾病人群：**根据所选的疾病领域定义目标疾
140 病人群，并考虑本疾病登记的代表性。

141 • 定性长期或短期研究：综合考虑科学性、实施的现实性
142 和经费支持程度，拟定观察期的时限，定性长期或短期研究，
143 尤其鼓励长期研究。

144 • 数据源：充分考虑数据来源的充足性、适用性和可及性。

145 • 数据安全与伦理：需考虑数据系统的安全性及数据使
146 用的合规性和所涉及的伦理问题。

147 • 各相关方合作方的确认：疾病登记需要多学科的协作，
148 需确认各参与方角色和职责，组建项目团队，以保障项目的
149 顺利实施。鼓励患者的积极参与以及项目团队同患者组织的
150 合作。

151 • 可行性论证：综合各种因素进行可行性论证，包括项目
152 的科学意义和临床价值、数据的获取和质量保证、经费是否
153 充足、患者的依从性、数据使用的安全性与合规性、伦理风
154 险等，以判断项目开展的可行性。

155 • 项目执行计划：即项目管理与实施计划，包括组织实施、
156 经费预算与管理、进度管理、人员管理、质量控制、沟通机
157 制、风险管理等，还应包括数据使用和共享机制的考虑。

158 2. 疾病登记设计

159 在《药物真实世界研究设计与方案框架指导原则（试行）》
160 中有关于真实世界研究设计的详细论述，本指导原则仅根据
161 疾病登记特点重点阐述几个方面。

162 (1) 登记目的

163 根据具体疾病及其目标人群、治疗、结局和主要数据来
164 源，简要阐明所要建立的疾病登记，包括区域（如果有）、针
165 对的疾病、研究期限、病例规模等，以及基于疾病登记所要
166 回答的临床问题，例如，罕见疾病的自然史、临床效果评价、
167 预后预测、安全性监测、药物经济学评价等。

168 (2) 目标疾病人群

169 首先应定义目标疾病人群，明确所依据的诊断标准以及
170 对应的疾病编码（如 ICD 编码）。然后根据目标疾病制定病
171 例的入选/排除标准，进而选择登记系统的入组人群。入组人
172 群是目标疾病人群的子集，严格意义上讲，应该从目标疾病
173 人群中通过随机抽样产生，但现实情况是随机抽样往往难以
174 实施，从而会影响其代表性，因此，需要提供患者招募及研
175 究中心、基线特征等信息，便于后续对研究人群代表性的评
176 估。

177 (3) 治疗/暴露

178 在疾病登记中，治疗策略和治疗模式由临床实践所决定，
179 对于治疗或暴露变量应尽可能详细记录，如药物治疗的剂量、
180 频次、给药途径、疗程、药物组合方案（如果有）等，以及
181 药物的商品名和生产厂家；如果是物理治疗（如放疗或激光
182 治疗），应给出具体的治疗参数和设备及其型号。由于疾病登

183 记是针对疾病所建立的登记，除非有特殊考虑，通常在建设
184 过程中并不特别限定治疗药物或治疗手段，而针对某种产品
185 /治疗策略的临床研究则是通过对疾病登记数据库的可行性
186 评估后（如该产品/治疗策略的样本量是否足够）开展的，之
187 后才有研究组和对照组的划分，同时伴随治疗也将被定义。

188 (4) 临床终点

189 针对某种疾病的临床研究，不同研究回答的临床问题不
190 尽相同，相应地在确定主要和次要临床终点时也不尽相同。
191 因此，在疾病登记的设计中难以确定临床研究主次要终点，
192 但需要较长远地考虑该疾病登记将要支持开展哪些临床研
193 究，这些研究会涉及到哪些重要的临床终点，包括有效性、
194 安全性和药物经济（如需要）的主要和次要终点。

195 (5) 随访期

196 如果是短期研究，则随访期的确定首先要依据疾病特点
197 和重要临床终点，同时还要兼顾经费的承受能力。如果是长
198 期研究，需要定义随访期的下限，而对随访期的上限则没有
199 特别要求。

200 (6) 变量词典

201 规范的疾病登记应有专门的变量词典。变量词典围绕数
202 据元素进行创建，其主要内容包括变量定义、变量维度、变
203 量名与标签、变量类型、变量单位、测量、关联关系、合法

204 值（如参考范围）、分类与转化、随访的时间点等。

205 • 定义/标准术语: 每一个变量应该有相应的标准的定义，
206 如果引用的是其它的标准术语库，应注明出处。

207 • 变量维度: 在变量词典中，通常会将变量划分归属于若
208 干维度，如人口学、病史、诊断、治疗或暴露、疗效结局（含
209 主要终点和次要终点）、实验室检查、安全性事件、生活方
210 式与饮食、社会与环境因素等。应合理划分变量的维度。

211 • 变量的名称、标签与类型: 每个变量的名称、标签和类
212 型需要按照统一标准设定，设定时应注意变量名称的最大长
213 度限定。

214 • 变量类型: 建议在变量词典中明确定义变量的采集形
215 式，在设计变量类型时尽量使用标准化的形式，如数值型、
216 选项型（分类型）等，不鼓励使用开放式文本。对于选项型
217 的变量，还应对选项标签进行规范化编码。在同一疾病登记
218 中，尽量使用统一的编码原则及顺序。

219 • 关联关系: 定义变量之间的关联关系，如疾病诊断与检
220 查结果、治疗的关联。

221 • 合法值: 定义变量采集的合法值范围，也称为有效范围，
222 注意变量的合法值范围指的是变量值对于疾病登记是否合
223 理、有效，需区别于临床检查值的正常值/参考值范围。实验
224 室检查结果的参考值范围在研究过程中可能会定期更新，可

225 单独管理，避免出现变量字典的频繁修改。

226 • 衍生数据: 定义衍生数据的产生方法，如根据年龄产生
227 的年龄分组变量，根据身高和体重产生的体重指数（BMI）
228 等。

229 • 随访的时间点: 每个变量都需要确定其基线和随访性
230 质。对于随访采集的变量，不同变量的随访周期和时间点不
231 完全相同，需要明确每一个变量的随访周期和时间点。

232 (7) 样本量

233 一般而言，在疾病登记设计阶段，需要根据最迫切需要
234 回答的一个或多个具体的研究问题估计样本量，并确保数据
235 库最终的样本量不低于设计确定的样本量。在实际操作中，
236 疾病登记在保证最低样本量的前提下，通常不设上限，而且
237 随着登记数据的积累，样本量会不断增大。需要注意，最初
238 的登记设计的样本量未必能够满足所有在疾病登记建设过
239 程中提出的一些新的临床研究，研发人员应评估疾病登记数
240 据库当时或扩大规模后能否满足新提出的临床研究对样本
241 量的要求。

242 (8) 数据来源

243 疾病登记的数据可以包括自身采集的数据和与外部融
244 合的数据，其来源又可分为医疗环境和非医疗环境采集的数
245 据。在医疗环境下，患者住院或根据随访安排来医院复诊，

246 可产生日常的和随访的诊疗和各种检查数据。在非医疗环境
247 下，数据主要来源于院外随访、PRO/ePRO 和监测设备（如
248 可穿戴设备）。

249 (三) 数据库构建

250 疾病登记数据库的构建需要充分考虑互操作性的要求，
251 从系统、数据标准、通用数据模型(common data model, CDM)
252 等维度实现不同数据库、系统之间的通讯、数据交换及信息
253 使用。

254 1. 系统

255 为保证数据质量，建议采用经验证的电子数据采集
256 (electronic data capture, EDC) 系统进行疾病登记的数据采
257 集及数据管理/治理。疾病登记所使用的数据库系统应满足
258 EDC 系统的一般要求：①安全的物理和网络环境；②系统运
259 行的稳定性及数据安全性；③用户角色管理及权限控制；④
260 稽查轨迹，即数据在系统中录入并保存后，所有的稽查轨迹
261 均应被记录且不允许被删除或修改；⑤标准化操作程序。具
262 体要求可参考《临床试验的电子数据采集技术指导原则》。此
263 外，还应满足对互操作性能的要求，即可以通过定义数据交
264 换的结构和格式，实现系统间的数据传输及业务信息交换。

265 在疾病登记数据库的构建及配置时，应当充分考虑在多
266 个数据库/数据系统间进行数据交换的需求，根据不同数据库

267 /数据系统的属性及数据结构，提前定义数据交换标准，并进
268 行传输测试，以确保信息的准确传输。

269 2. 数据标准

270 在疾病登记的建设过程中，可能需要融合多种不同来源
271 的临床数据。标准化的数据结构与格式是数据交换及共享的
272 基础，为了实现数据库的融合与扩展，提高数据的可交换性
273 及数据质量，同时为提高注册监管审评效率，在构建疾病登
274 记数据库时建议优先考虑行业通用的数据标准，例如
275 **CDASH (Clinical Data Acquisition Standards Harmonization)**。

276 从更广义的角度，数据标准涵盖疾病登记建设的各个环节，
277 如方案设计、数据采集、分析、交换、递交、报告撰写等，
278 在构建疾病登记数据库时应当充分考虑相关标准适用性及
279 数据结构的兼容性。用于上市申请时，最终递交的数据应符合
280 《药物临床试验数据递交指导原则（试行）》要求。

281 对于通用数据标准未覆盖/不适用的数据模块，在疾病登
282 记的系统层面也应制定统一的标准，并在相同或不同的疾病
283 队列中应用。

284 3. 通用数据模型

285 在疾病登记的数据库及数据采集方案的设计阶段，应清
286 晰地定义该疾病队列适用的通用数据模型，明确数据源与通
287 用数据模型的映射关系，包括变量的定义、源数据的提取/转

288 换/加载规则、多个数据库的数据转换标准等。有关数据通用
289 模型的详细介绍，可参见《用于产生真实世界证据的真实世
290 界数据指导原则（试行）》。

291 **（四）疾病登记数据的获取、录入与质量控制**

292 登记数据主要由登记系统内部以前瞻性收集的方式生成，
293 也可能融合外部数据，如既往数据或其它系统的数据。本节
294 主要介绍登记数据如何由系统内部生成。

295 **4. 数据的获取**

296 疾病登记数据主要来源于医院内系统数据、随访数据、
297 PRO 数据和日常的个体监测数据等。

298 （1）院内系统数据：医院内部系统的数据，特别是电
299 子病历数据，是疾病登记数据最主要的来源。通常，无论患
300 者是否住院，其基线数据、诊疗数据、各种检查数据等，均
301 产生于院内系统。

302 （2）随访数据：通过随访患者获取数据是保障数据纵
303 向性的重要手段。根据研究计划，通过定期或不定期随访、
304 门诊或通讯随访等方式，获取患者的疾病状态、临床终点、
305 治疗等数据。随访数据通常以纸质或者电子的病例报告表
306 （case report form, CRF）作为主要采集工具，也可以考虑采
307 用经验证的技术手段直接导入。

308 (3) **PRO 数据**：PRO 数据有纸质和电子记录两种方
309 式，后者称为电子化患者报告结局（**electronic patient-reported**
310 **outcome, ePRO**），目前较为常用。ePRO 系统可与电子病历
311 系统或 EDC 系统对接，形成个体水平的完整数据流。有关
312 PRO 数据的详细介绍，可参见《患者报告结局在药物临床研
313 发中应用的指导原则（试行）》。

314 (4) **个体监测数据**：借助移动设备（如智能手机、可
315 穿戴设备、动态心电监测仪等）实时采集个体监测数据在数
316 据收集的便利性和即时性等方面优势明显，不仅可以丰富数
317 据的获取手段，同时也丰富了疾病登记数据库。鼓励数据库
318 建设中采用先进的且可靠的数据采集技术。

319 5. 数据录入或导入

320 数据的采集包括人工录入和自动化导入两种方式。人工
321 录入的数据通常包括纸质或电子的随访数据、纸质 PRO 数
322 据等，由具有资质的、经授权的人员将采集的数据录入至
323 EDC 系统。录入工作正式开展前，需要根据疾病登记计划与
324 设计的要求，制定明确的填写指南和具体的填写说明，并对
325 相关工作人员进行数据填写及录入的培训。

326 ePRO 数据、医院随访系统数据、来自移动设备的个体
327 监测数据有时可以直接导入至 EDC 系统。数据导入前需明
328 确源数据与 EDC 数据库的映射关系及数据导入机制，并通

329 过导入测试。所有的规则及操作过程均应保留详细的文件记
330 录，确保数据处理流程的透明性及数据的可溯源性。

331 6. 数据管理/治理及质量控制

332 回顾性收集的数据通常需要进行数据治理，而前瞻性收
333 集的数据则需进行数据管理。数据管理/治理流程的关键环节
334 包括但不限于：数据管理/治理计划的撰写、CRF/数据采集表
335 的设计、数据库的设计及构建、数据采集及录入、数据核查
336 及核查管理、医学编码、数据审核、数据库锁定、数据储存
337 与传输、质量控制等。此外，数据治理的主要流程还包括：
338 个人信息保护及数据安全性的处理，通用数据模型的建立，
339 数据提取、转化等。具体的数据管理/治理流程，可参见《药
340 物临床试验数据管理工作技术指南》和《用于产生真实世界
341 证据的真实世界数据指导原则（试行）》。

342 数据质量控制是确保研究数据完整性、准确性和透明性
343 的关键。针对疾病登记数据的特点，建议在数据质量控制时
344 考虑以下几点原则：①建立质量控制标准操作规程（**standard**
345 **operation procedure, SOP**）；②制定完善的质量控制计划，明
346 确质量控制工作的范围、频率、流程及质量标准；③在疾病
347 登记计划中明确数据采集的时间点和实施顺序；④针对相关
348 人员进行数据采集的培训和指导，包括数据填写及录入规则、
349 相关数据标准及通用数据模型，数据填写、存储、传输的方

350 式和要求等；⑤保证数据处理过程记录的完整性；⑥针对可
351 能发生的脱落，积极采取措施控制，尽量将脱落率控制在可
352 接受范围以内。

353 (五) 数据库成熟

354 疾病登记在数据库建成后，随着数据的不断收集和沉淀，
355 数据库逐渐成熟。数据库成熟是指数据库中已经生成的数据
356 可以支持供监管决策的产生真实世界证据的统计分析，具体
357 而言应至少具备以下四点：①完成设计规定的观察期限的病
358 例数达到了最小样本量的要求；②数据质量可以支持后续的分
359 析，即满足 ALCOA+CCEA 原则（ALCOA：可归因性
360 (Attributable)，易读性(Legible)，同时性(Contemporaneous)，
361 原始性(Original)，准确性(Accurate)；CCEA：完整性
362 (Complete)，一致性(Consistent)，持久性(Enduring)，可获得
363 性(Available When Needed)；③数据应可以衍生为符合监
364 管递交标准的数据格式，可参见《药物临床试验数据递交指
365 导原则（试行）》；④最终可用于衍生为满足不同分析需要的
366 分析数据库。

367 (六) 数据库融合与扩展

368 在疾病登记的建设过程中，除了自身数据库的建设外，
369 还可以融合其它数据，包括既往已有的数据和其它登记数据

370 库。此外，还可以根据新的研究需求或新条件的产生扩展疾
371 病登记。

372 1. 数据库融合

373 如果登记系统不能从内部所掌握的数据源获取完整的
374 登记研究所需的信息，就需要从其它数据源补充，例如，在
375 以死亡为主要终点的肿瘤研究中，以医院管理信息系统为主
376 要数据源的疾病登记往往难以包含完整的死亡事件，需要从
377 其它数据源（如国家 CDC 的死亡登记系统）融合这一信息。
378 在与其它数据源进行融合时，应确保链接数据系统的互操作
379 性和准确、一致、完整的数据传输。

380 当与其它登记数据库的融合时，由于数据库结构和数据
381 标准不完全相同，需要考虑使用或研发通用数据模型并就融
382 合的可行性进行评估，包括关键变量的匹配程度、数据的适
383 用性和可追溯性等。

384 与既往已有的数据融合时还需要考虑，应按照前瞻性设
385 计的随访计划一并跟踪观察融合的既往数据的病例，特别是
386 观察期未达计划要求的病例。如果既往数据的质量不佳，如
387 关键变量缺失率太高、记录不准确等，或目标人群、关键变
388 量的定义和测量与本疾病登记无法取得一致，即使数据可获
389 取，也不宜将其融合。

390 2. 数据库扩展

391 随着科学研究进展和数据获取的可能，疾病登记也可能
392 以不同方式扩展。

393 • 疾病种类增加：例如，出于共病研究的需要，在原有的
394 糖尿病登记队列中加入慢性肾病登记队列；考虑同类病种的
395 便利性，在冠心病登记队列增加心衰登记队列。

396 • 队列观察期限延伸：对于某些观察期较短的疾病登记，
397 由于原研究计划的局限（如因为观察期太短而不足以估计中
398 位生存时间）或研究条件的改善（如获得充足的经费支持），
399 可能会修改原研究计划将观察期显著延长，因为观察期越长，
400 队列研究的价值越高。

401 • 变量增多：随着科学研究的进展，伴随着新的诊断方法
402 和预后因子（如新的生物标记物）的产生，新的重要终点评
403 价方法的出现，新的分类标准的提出等，疾病登记应及时加
404 入相应的新的变量，以紧跟科学前沿的需要。

405 **（七）数据安全与伦理**

406 应高度重视疾病登记的安全问题。用于分析的疾病登记
407 数据应是经匿名化处理的数据，且不包含敏感个人信息，如
408 姓名、证件号码、联系电话、所属单位、家庭住址、办公地
409 址等。对基于 EDC 系统建立的疾病登记，在数据管理的操作
410 端和数据经处理后最终形成的分析数据库，患者的识别码是
411 研究编号，通常在数据产生端（如原始病历或 HIS 系统）将

412 患者个人信息通过编码转化而成。如果已有的疾病登记未经
413 匿名化处理，则需经匿名化处理后才具备数据共享条件，否
414 则会涉及个人信息安全的法律问题。有关数据安全的具体措
415 施，可参见《用于产生真实世界证据的真实世界数据指导原
416 则（试行）》。

417 数据使用的伦理要求可参照国家卫生健康委员会《涉及
418 人的生物医学伦理审查办法》等管理要求执行，需要经过伦
419 理审查。伦理审查除了审核研究方案的科学性和合理性之外，
420 还会关注研究对象的知情同意问题，知情同意应平衡考虑患
421 者权益保护与实施可行性。

422 **（八）早期融入 PFDD 理念**

423 患者参与的疾病登记具有积极意义。例如，在确定研究
424 终点、关键变量及评估标准时，征求患者的意见和需求，可
425 以更好地反映患者关心的问题，如生活质量、满意度和治疗
426 依从性等。在登记的设计中，还可以考虑疾病管理功能，例
427 如，提供线上的用药和监测指导、心理支持、康复训练等，
428 帮助患者更好地管理自己的疾病，使其在药物研发过程中获
429 益。基于 PFDD 的理念，疾病登记系统可以为患者提供关于
430 所患疾病、治疗选择和参与研究的机会的信息，帮助患者提
431 高对自身疾病的认识。

432 将 PFDD 的理念融入疾病登记及其研究，有助于提高患
433 者满意度、改善临床实践及药物研发的质量和效率。

434 五、从临床研究角度对疾病登记数据的评估

435 只有经评估后适用的真实世界数据才有可能产生支持
436 监管决策的真实世界证据，对疾病登记数据的适用性评估可
437 参阅《用于产生真实世界证据的真实世界数据指导原则（试
438 行）》。以下主要针对疾病登记数据阐述评估的要点。

439 （一）研究目的的实现

440 首先需要评估疾病登记数据是否支持拟开展项目的研
441 究目的的实现，即基于这些数据经分析后所产生的 RWE 是
442 否能够回答研究所要回答的临床或科学问题。可以根据对一些
443 关键信息的评估做出初步判断，例如，疾病的诊断方法是
444 否可靠，病例的选择是否与研究目的一致，研究组和对照组的
445 治疗信息是否详尽，重要临床结局的测量是否可信，重要
446 临床结局的随访期是否够长，关键变量的记录是否完整，样
447 本量是否够大，被重点关注的安全性信息是否记录，实施中
448 的质量保证措施是否得力，数据的使用有无障碍，数据是否
449 可追溯等。

450 （二）疾病登记建设的规范性

451 规范进行疾病登记建设是对研究的科学性和高效性的
452 有力保障。规范性应体现在首先有周密的计划和设计，然后

453 有严格的执行和质控。具体还要看，是否采用了行业的通用
454 标准，是否在队列中保持统一的规范/标准，是否有专门的变
455 量词典、操作手册、SOP等，执行环节的记录是否完整，数
456 据管理与使用的角色和权限是否清晰，组织和沟通的机制是
457 否顺畅，等等。

458 (三) 研究人群代表性

459 根据研究方案所制定的病例入选/排除标准，评估研究人
460 群对于目标人群的代表性。如果研究人群与目标人群有较大
461 差异，应考虑研究结论的偏倚和适用范围。对于来自不同数
462 据源的融合，应评估研究人群的一致性，如进行基线特征的
463 比较。

464 (四) 随访期、随访问隔和时间窗

465 疾病登记的重要结局变量的随访期一定要足够长，否则
466 无法做出准确可靠的估计，例如，当研究的主要终点为生存
467 时间，如果随访期太短不足以估计事件的中位时间，则难以
468 对疗效做出客观判断。但从另一方面看，如果感兴趣的结局
469 事件需要特别长的随访期，例如，有关乳腺癌生存时间的研
470 究，随访期可能需要十余年甚至更长时间，在不具备开展该
471 项目研究的条件（如经费、研究团队等）下，选择这样的研
472 究问题是欠妥的。

473 随访问隔的设计应合理，间隔太长不利于刻画疾病的变化规律，会降低时间依赖型事件的效应估计精度；间隔太短
474 会增加研究的负担和实操的难度，从而有可能影响研究的质量。
475 虽然疾病登记的时间窗相对随机对照试验可以放宽，但
476 要适度，否则同样会影响研究的质量。
477

478 **（五）数据质量及其控制**

479 数据质量的评估在《用于产生真实世界证据的真实世界
480 数据指导原则（试行）》的指南中有详尽阐述。由于疾病登记
481 通常持续时间较长，其日常的运行和维护显得非常重要，因
482 此，应特别关注其质量控制计划和具体执行情况。例如，是
483 否有详细的质量控制计划，是否安排专职质控人员，是否制
484 定相应的 SOP，是否有系统的培训计划以及相应的活动记录
485 等。此外，疾病登记数据的可追溯性尤为被关注，因为疾病
486 登记数据有可能不是直接来自源数据，例如融合的外部数据，
487 如果在数据传输或互操作过程中出现失误，有可能导致数据
488 的不可追溯，应避免此类情况的发生。

489 **（六）数据使用的合规性**

490 疾病登记数据的使用应没有安全和法律方面的风险，且
491 符合监管方面的要求。基于 EDC 系统建立的疾病登记数据
492 库在研究中使用的的应是匿名化的数据。疾病登记收集和使

493 用还应符合伦理要求，在研究使用患者隐私为匿名化且无需
494 继续观察等情形下可以考虑豁免知情同意。

495 **六、疾病登记数据的递交及与监管机构的沟通交流**

496 **（一）疾病登记数据的递交**

497 用于支持注册申请的疾病登记数据需要按照递交标准进
498 行转换，然后才能递交，转换和递交可参见《用于产生真实
499 世界证据的真实世界数据指导原则（试行）》、《药物临床试验
500 数据递交指导原则（试行）》。

501 **（二）与审评机构的沟通交流**

502 为保证疾病登记数据的质量符合监管要求，使用登记数
503 据开展真实世界研究作为关键证据支持监管注册的，需与药
504 审中心及时沟通交流并达成共识。在真实世界研究正式开始
505 前，应就研究目的、适用性、疾病登记数据能否满足监管对
506 RWD 的适用性要求进行交流，包括疾病登记的计划与设计、
507 运行与维护、病例的规模和队列的时间跨度、重要的变量、
508 数据的完整度、拟开展的数据治理及其计划等。沟通交流的
509 具体时间节点及注意事项参见《真实世界证据支持药物注册
510 申请的沟通交流指导原则（试行）》。

511

参考文献

- 512 [1] 国家药品监督管理局. 真实世界证据支持药物研发与审
513 评的指导原则（试行）. 2020.1
- 514 [2] 国家药品监督管理局. 用于产生真实世界证据的真实世
515 界数据指导原则（试行）. 2021.4
- 516 [3] 国家药品监督管理局. 患者报告结局在药物临床研发中应
517 用的指导原则（试行）. 2021.12
- 518 [4] AHRQ. Registries for Evaluating Patient Outcomes A User's
519 Guide. 4th Edition. 2020.9
- 520 [5] EMA. Guideline on Registry-based Studies. 2021.12
- 521 [6] FDA. Real-World Data: Assessing Registries to Support
522 Regulatory Decision-Making for Drug and Biological Products
523 (Draft). 2021.11
- 524 [7] Charrow J, Andersson HC, Kaplan P, et al. The Gaucher
525 registry: demographics and disease characteristics of 1698
526 patients with Gaucher disease. Arch Intern Med. 2000; 160 (18):
527 2835-2843
- 528 [8] Pida Björkgren, Uppsala Clinical Research Center (UCR).
529 SWEDHEART Annual report 2022. www.swedeheart.se; 2023
- 530 [9] 国家药品监督管理局. 真实世界研究支持儿童药物研发与
531 审评的技术指导原则（试行） 2020.8

- 532 [10] 国家药品监督管理局. 罕见疾病药物临床研究技术指导
533 原则（试行）2022.1
- 534 [11] 国家药品监督管理局. 罕见病药物临床研究统计学指导
535 原则（试行）2022.6
- 536 [12] 国家药品监督管理局. 单臂临床试验设计用于支持抗肿瘤
537 药上市申请的适用性技术指导原则（试行）2023.3
- 538 [13] Belen O, Concato J and Kraus S. FDA Approval
539 Demonstrates the Role of Real-World Evidence in Regulatory
540 Decision-Making on Drug Effectiveness.
541 [https://www.fda.gov/drugs/
542 approval-demonstrates-role-real-world-evidence-regulatory-
543 decision-making-drug-effectiveness](https://www.fda.gov/drugs/news-events-human-drugs/fda-approval-demonstrates-role-real-world-evidence-regulatory-decision-making-drug-effectiveness); 2017
- 544 [14] Erdman J, Wolfram J, Nimke D, et al. Lung Transplant
545 Outcomes in Adults in the United States: Retrospective Cohort
546 Study Using Real-world Evidence from the SRTR.
547 Transplantation 2022;106: 1233-1242
- 548 [15] FDA. Considerations for the Design and Conduct of
549 Externally Controlled Trials for Drug and Biological Products
550 (Draft). 2023.2
- 551 [16] 国家药品监督管理局. 药物真实世界研究设计与方案框
552 架指导原则（试行）. 2023.2

- 553 [17] FDA. FDA approves first treatment for a form of Batten
554 disease.
555 <https://www.fda.gov/news-events/press-announcements/fda->
556 [approves-first-treatment-form-batten-disease;](https://www.fda.gov/news-events/press-announcements/fda-) 2017
- 557 [18] Markham A. Cerliponase Alfa: First Global Approval. *Drugs*
558 2017;77(11): 1247-1249
- 559 [19] Schulz A, Ajayi T, Specchio N, et al. Study of
560 intraventricular cerliponase alfa for CLN2 disease. *N Engl J*
561 *Med.* 2018; 378: 1898- 1907.
- 562 [20] Nyberg K and Hedman P. Swedish guidelines for registry-
563 based randomized clinical trials. *Ups J Med Sci.* 2019;124(1):33-
564 36
- 565 [21] Gliklich RE, Leavy MB, Karl J, et al. A framework for
566 creating standardized outcome measures for patient registries. *J*
567 *Comp Eff Res.* 2014;3(5):473-480
- 568 [22] 国家药品监督管理局. 临床试验的电子数据采集技术指
569 导原则. 2016.7
- 570 [23] 国家药品监督管理局. 药物临床试验数据递交指导原则
571 (试行). 2020.7
- 572 [24] 国家药品监督管理局. 药物临床试验数据管理工作技术
573 指南. 2016.7

574 [25] Ma X, Long L, Moon S, Adamson BJS, et al. Comparison of
575 population characteristics in real-world clinical oncology
576 databases in the US: Flatiron Health, SEER, and NPCR. medRxiv.
577 Preprint posted online May 30, 2020. doi:10.1101/2020.03.16.
578 20037143

579 [26] 国家药品监督管理局.真实世界证据支持药物注册申请的
580 沟通交流指导原则（试行）. 2023.2

581 [27] <https://www.nicor.org.uk/myocardial-ischaemia-minap>
582 heart-attack-audit/The Myocardial Ischaemia National Audit
583 Project (MINAP)

584 [28] <https://clinicaltrials.gov/study/NCT04564742> (DAPA-MI)
585 A Registry-based, Randomised, Double-blind, Placebo-
586 Controlled Cardiovascular Outcomes Trial to Evaluate the Effect
587 of Dapagliflozin on Cardiometabolic Outcomes in Patients
588 Without Diabetes With Acute Myocardial Infarction at Increased
589 Risk for Subsequent Development of Heart Failure.

590

591 **附录 1：词汇表**

592 **产品登记 (Product Registries)**：针对某一种产品或多种产
593 品所做的登记。

594 **单臂临床试验 (Single-arm/One-arm Trial)**：是一种只设置
595 试验组的非随机临床试验，通常采用外部对照，如历史对照
596 或平行对照。

597 **登记 (Registries)**：又称为病例登记，是指根据事先确定的
598 登记目的，有组织的、系统的、根据所定义的随访/观察时限
599 和时点收集与人口学、疾病、暴露、诊疗、结局等相关的患
600 者个人层面的纵向数据所形成的数据库，也称为登记系统，
601 登记数据库是登记形成的最终成果。

602 **登记研究 (Registry-based Study)**：基于登记所产生的数据
603 开展的研究。

604 **电子病历 (Electronic Medical Record, EMR)**：由医疗机构
605 中授权的临床专业人员创建、收集、管理和访问的个体患者
606 的健康相关信息电子记录。

607 **观察性研究 (Observational Study)**：根据特定研究问题，
608 不施加主动干预的、以自然人群或临床人群为对象的、探索
609 暴露/治疗与结局因果关系的研究。

610 **患者报告结局 (Patient-reported Outcome, PRO)**：为任何
611 来自患者直接报告且不被他人修改或解读的对自身疾病和

612 相应治疗感受的评估结局。

613 **患者为中心的药物治疗 (Patient-focused Drug Development,**
614 **PFDD)**：指一套系统的方法，在药物全生命周期中，该法有
615 助于确保患者的体验、观点、需求和优先顺序能够被捕获并
616 有效地融入到药物的研发和评估中。

617 **患者体验数据 (Patient Experience Data/ Patient Input)**：任
618 何人自愿提供的有关患者对于疾病和状态的体验的信息，此
619 类信息包括患者的体验、观点、需求和喜好，但不限于症状
620 及其疾病自然史，状态对其功能和生存质量的影响，治疗体
621 验，结局对患者的重要性，患者对结局和治疗的偏好，以及
622 其它患者关注的重要信息。

623 **回顾性观察性研究 (Retrospective Observational Study)**：
624 在研究开始时确定目标人群、并根据历史数据（研究开始前
625 生成的数据）开展的观察性研究。

626 **临床试验 (Clinical Trial)**：属于干预性临床研究，是将一种
627 或多种干预（可能包括安慰剂或其它对照）前瞻性地分配给
628 人类受试者，以评估这些干预对健康相关的生物学或行为
629 结局的影响。

630 **疾病登记 (Disease Registry)**：针对某一种疾病，或某一类
631 疾病，或多类疾病所做的登记。

632 **前瞻性观察性研究 (Prospective Observational Study)**：在

633 研究开始时确定目标人群、并在研究开始前确定收集暴露/治
634 疗和结果数据的观察性研究。

635 **数据标准 (Data Standard)**：是关于如何在计算机系统之间
636 构建、定义、格式化或交换特定类型数据的一系列规则。数
637 据标准可使递交的资料具有可预测性和一致性，且具有信息
638 技术系统或科学工具可以使用的形式。

639 **数据融合 (Data Linkage)**：将多来源的数据和信息加以合并、
640 关联及组合，形成统一的数据集。

641 **数据元素 (Data Element)**：临床研究中记录的受试者的单
642 一观察值，例如，出生日期，白细胞计数，疼痛严重程度，
643 以及其它临床观察值。

644 **数据治理 (Data Curation)**：针对特定临床研究问题，为达到
645 适用于统计分析而对原始数据所进行的治理，其内容至少包
646 括数据提取（含多个数据源）、数据安全性处理、数据清洗
647 （逻辑核查及异常数据处理、数据完整性处理）、数据转化
648 （通用数据模型、归一化、自然语言处理、医学编码、衍生
649 变量计算）、数据质量控制、数据传输和存储等若干环节。

650 **通用数据模型 (Common Data Model, CDM)**：是多学科合
651 作模式下对多源异构数据进行快速集中和标准化处理的数

652 据系统，其主要功能是将不同数据标准的源数据转换为统一
653 的结构、格式和术语，以便跨数据库/数据集进行数据整合。

654 **外部对照 (External Control)**：在临床试验中，以试验对象
655 以外的真实世界数据为对照，以评价所研究的干预效果。外
656 部对照可以是历史数据，也可以是平行观测所获得的数据。

657 **源数据 (Source Data)**：临床研究中记录的临床症状、观测
658 值和用于重建和评估该研究的其他活动的原始记录和核证
659 副本上的所有信息。源数据包含在源文件中（包括原始记录
660 或其有效副本）。

661 **真实世界数据 (Real-World Data, RWD)**：与患者健康状
662 况有关的和/或来源于各种日常医疗过程所收集的、经过分析
663 具有潜在形成真实世界证据可能的数据。

664 **真实世界研究 (Real-World Research/Study, RWR/RWS)**：
665 为 CER 所涵盖，系指收集真实世界环境下与患者有关的数据
666 据（真实世界数据），通过分析，获得医疗产品的使用价值
667 及潜在获益或风险的临床证据（真实世界证据），其主要研
668 究类型是观察性研究，也可是实用临床试验等。

669 **真实世界证据 (Real-World Evidence, RWE)**：通过对真实
670 世界数据的分析获得的关于医疗产品的使用情况和潜在获
671 益或风险的临床证据。

672

附录 2: 中英文词汇对照

中文	英文
标准操作规程	Standard Operation Procedure, SOP
病例登记	Patient Registry
单臂临床试验	Single-arm/One-arm Trial
登记	Registries
登记研究	Registry-based Study
电子病历	Electronic Medical Record, EMR
观察性研究	Observational Study
患者报告结局	Patient Reported Outcome, PRO
患者为中心的药物治疗	Patient-focused Drug Development , PFDD
患者体验数据	Patient Experience Data/ Patient Input)
回顾性观察性研究	Retrospective Observational Study
疾病登记	Disease Registries
基于登记的随机对照临床 试验	Registry-based/Register-based RCT
前瞻性观察性研究	Prospective Observational Study
数据标准	Data Standard
数据融合	Data Fusion
数据元素	Data Element

中文	英文
数据治理	Data Curation
通用数据模型	Common Data Model, CDM
外部对照	External Control
源数据	Source Data
卫生信息系统	Health Information System, HIS
医疗服务登记	Health Service Registries
真实世界数据	Real World Data, RWD
真实世界研究	Real World Research/Study, RWR/RWS
真实世界证据	Real World Evidence, RWE
质量控制	Quality Control, QC

674