

真实世界数据:评估电子健康记录 和医疗索赔数据以支持药品和 生物制品的监管决策

行业指南

指南草案

本指南文件仅供评论使用

关于本草案文件的意见和建议应在《联邦公报》发布关于指导原则草案有效性的通知后60天内提交。将电子意见提交至 <https://www.regulations.gov>。向食品药品监督管理局（地址:5630 Fishers Lane, RM）的档案管理人员（HFA-305）提交书面意见。1061, Rockville, MD 20852。所有评论均应标明联邦公报发布的可用性通知中列出的文档编号。

关于本草案文件或真实世界证据计划的问题，请发送电子邮件至 CDERMedicalPolicy-RealWorldevidence@fda.hhs.gov

真实世界数据:评估电子健康记录 和医疗索赔数据, 以支持药品和 生物制品的监管决策

行业指南

其他副本可从以下机构获得:
通讯办公室, 药物信息部药物评价
与研究中心
美国食品药品监督管理局
0001 New Hampshire Ave.,
Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
电话:855-543-3784或301-796-3400; 传真:301-431-6353
邮箱:druginfo@fda.hhs.gov
<https://www.fda.gov/Drugs/Guidance-Compliance-Regulatory-Information/Guidances-Drugs>
和/或
对外信息、外联和发展办公室生物
制品评价和研究中心
美国食品药品监督管理局
10903 New Hampshire Ave.,
Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
电话:800-835-4709或240-402-8010
电子邮箱:ocod@fda.hhs.gov
<https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics/biologics-guidances>

目录

真实世界数据:评估电子健康记录和医疗索赔数据, 以支持药品和生物制品的 监管决策.....	4
I. 引言和范围.....	4
II. 背景.....	6
III. 一般考虑.....	6
IV. 数据来源.....	7
A. 数据源的相关性.....	8
B. 数据采集:一般讨论	8
C. 缺失数据:一般考虑	13
D. 验证:一般注意事项	13
表1:二进制变量的灵敏度、特异性、阳性预测值(PPV)和阴性预测值(NPV) 的计算示意图	15
V. 研究设计要素.....	15
A. 时间段定义.....	15
B. 研究人群的选择.....	16
C. 暴露确定和确认.....	16
D. 结果确定和验证.....	20
E. 协变量确定和验证.....	25
VI. 数据累积、管理和转换为最终研究特定数据集期间的数据质量	26
图1:EHR数据生命周期的说明性示例 ¹⁶	27
.....	27
A. 数据特征.....	28
B. QA/QC计划文件.....	31
C. 数据管理过程文件	31
VII. 术语表.....	32
VIII. 参考文献	36

包含非约束性建议草案——不用于实施

真实世界数据:评估电子健康记录和医疗索赔数据, 以支持药品和生物制品的监管决策

行业指南¹

本指南草案一经定稿,将代表美国食品药品监督管理局(FDA或机构)关于该主题的最新见解。它不为任何人建立任何权利,也不对FDA或公众具有约束力。您可以使用替代方法,如果它满足适用法规和法规的要求。如需讨论替代方法,请联系标题页所列负责本指南的FDA工作人员。

I. 引言和范围

2016年12月13日签署的《21世纪治愈法案》(Cures Act)²旨在加速医疗产品的开发,并为需要的患者更快、更高效地带来创新。除其他条款外,《治愈法案》在《联邦食品、药品和化妆品法案》(FD&C法案)(21 U.S.C.355g)中增加了第505F条。根据本章节,FDA创建了一个评价真实世界证据(RWE)潜在用途的计划框架,以帮助支持已根据FD&C法案第505(c)节批准的药物的新适应症的批准,或帮助支持或满足批准后研究要求(RWE计划)⁴。

FDA发布本指南作为其RWE计划的一部分,并在一定程度上满足FD&C法案第505F节发布关于在监管决策中使用RWE的指南的要求。⁵ RWE计划将涵盖使用真实世界数据(RWD)来源(例如来自常规临床实践的信息)来得出RWE的临床研究。

1本指导原则由食品药品监督管理局药品评价与研究中心(CDER)与生物制品评价与研究中心(CBER)和肿瘤学卓越中心(OCE)合作编写。

2公法114-255

3出于本指导原则的目的,所有提及的药物均包括人用药品和生物制品。本指南不适用于医疗器械。有关医疗器械的信息,请参见标题为“使用真实世界证据支持医疗器械监管决策”的指南,网址为<https://www.fda.gov/Regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices>。

4请参见FDA真实世界证据计划框架,网址为<https://www.fda.gov/media/120060/Download>。该框架和RWE计划还涵盖了根据《公共卫生服务法》许可的生物制品。

5见FD&C法案第505F(e)节。

本指导原则旨在为申办方、研究人员和其他感兴趣的利益相关者提供建议，以便在临床研究中使用电子健康记录⁶(EHR)或医疗索赔数据⁷以支持有效性或安全性的监管决策。

出于本指南的目的，FDA对RWD和RWE的定义如下：⁸

- RWD是与患者健康状况或常规从各种来源收集的医疗保健提供有关的数据。
- RWE是从RWD分析中得出的关于医疗产品的使用和潜在受益或风险的临床证据。

RWD的示例包括源自EHR的数据、医疗索赔数据、来自产品和疾病登记的数据、患者生成的数据（包括在家使用的数据）以及从其他来源收集的可以告知健康状况的数据，例如数字健康技术。本指南侧重于提供者记录的健康相关数据，这些数据可以从两个来源中提取：EHR和医疗索赔数据。EHR和医疗索赔数据是包含患者健康信息的电子医疗保健数据的类型，这些数据广泛用于安全性研究，并且越来越多地建议用于有效性研究。EHR和医疗索赔数据可被视为各种临床研究设计中的数据源。

本指南讨论了以下与临床研究中EHR和医疗索赔的潜在用途相关主题，以支持监管决策：

1. 选择应适合解决研究问题并充分描述研究人群、暴露、感兴趣的结局和关键协变量的数据源。
2. 开发和验证研究设计要素的定义（例如，暴露、结局、协变量）
3. 在数据收集、数据管理以及最终研究特定数据集期间的数据来源和质量

本指南不提供关于研究设计选择或统计分析类型的建议，也不认可任何类型的数据源或研究方法。对于所有研究设计，确保用于支持监管决策的数据的可靠性和相关性非常重要。

6关于在本指南中首次提及的粗斜体的单词和短语的定义，请参见术语表（第VII节）。

7对于本指导原则的目的，术语*临床研究*是指所有研究设计，包括但不限于干预性研究，其中治疗是通过方案分配的（例如，随机或单臂试验，包括那些使用RWD作为外部对照组的试验）和非干预性研究，其中治疗是在常规临床治疗过程中确定的，即观察性研究（例如，病例对照或队列研究）。在整个指南中，FDA可互换使用术语clinical studies, studies, 和 study。

8见FDA真实世界证据计划框架，网址为<https://www.fda.gov/media/120060/Download>。

出于本指南的目的，术语 *可靠性* (*reliability*) 包括数据准确性、完整性、出处和可追溯性。术语 *相关性* (*relevance*) 包括关键数据要素（暴露、结局、协变量）的可用性和研究中足够数量的代表性患者。

本文件的内容不具有法律的效力和效力，也不以任何方式约束公众，除非特别纳入合同。本文件仅旨在向公众提供有关法律现有要求的明确信息。FDA指南文件（包括本指南）应仅视为建议，除非引用了特定的法规或法律要求。FDA指南中使用的“应该” (*should*) 一词表示建议或推荐，而不是要求。

II. 背景

FDA行业和FDA工作人员使用《电子医疗保健数据进行和报告药物流行病学安全性研究的最佳实践指南》（2013年5月）侧重于在药物流行病学安全性研究中使用电子医疗保健数据。⁹2013年的该指南包括记录药物流行病学安全性研究的设计、分析和结果的建议，以优化FDA对提交给FDA的方案和研究报告的审查。

本指南通过扩展与数据源选择相关的指南的某些方面来补充2013年指南，并为评估临床研究中使用的EHR和医疗索赔数据的相关性和可靠性提供了额外的指南。本指南还更广泛地概述了与在临床研究中使用EHR和医疗索赔数据有关的考虑因素，包括旨在告知FDA对产品有效性进行评估的研究。

III. 一般考虑

对于使用将提交给FDA以支持监管决策的EHR或医疗索赔数据的所有研究，申办方应在进行研究之前提交方案和统计分析计划。申办方在开展研究前寻求FDA意见，应要求与相关FDA审评部门进行意见或会议讨论研究。应预先定义研究设计、分析、实施和报告的所有基本要素，对于每个研究要素，方案和最终研究报告应描述如何从选定的RWD来源确定该要素，包括适用的验证研究。关于研究要素的更多信息见第V节，研究设计要素。

本指南提供了关于选择数据源的建议，以最大化从临床研究的EHR和医疗索赔中获得的数据的完整性和准确性。使用某些研究设计特征或特定分析来解决错误分类或缺失信息，以及实现协变量平衡的方法，将在其他

⁹我们定期更新指导原则。关于指南的最新版本，请查看FDA指南网页 <https://www.FDA.gov/regulatory-information/search-fda-guidance-documents>。

侧重于研究设计和分析的FDA RWE指南中讨论。本指南阐述了对于确定数据可靠性和相关性至关重要的问题，并应在方案中解决，包括：

1. 研究问题和支持关键研究要素的数据源的适当性和潜在局限性。
2. 确定研究设计要素的时间段。
3. 研究设计要素（例如，研究人群的入选/排除标准、暴露、结局、协变量）和验证研究结果的概念定义和操作定义。有关研究设计要素的概念和操作定义示例，请参见第V节，研究设计要素。
4. 质量保证和质量控制（QA/QC）程序，用于数据累积、管理和转化为最终研究特定数据集。

IV. 数据来源

提交给FDA的方案应确定研究提出的所有数据源，以及其他相关的描述性信息（讨论如下）。FDA不会认可一个数据源而不是另一个数据源，也不会试图限制可能与回答研究问题相关的可能数据源。

应对每个数据源进行评估，以确定可用信息是否适用于解决特定的研究假设。由于现有的电子医疗保健数据不是为了支持向FDA提交的监管文件而开发的，因此了解其用于该目的时的潜在局限性非常重要。潜在限制的示例包括：

1. 医疗索赔数据的目的是支持对护理的支付；索赔可能无法准确反映特定疾病，或者患者可能患有索赔数据中未反映的特定疾病或状况。
2. 生成的EHR数据可用于临床护理，也可作为计费 and 审核实践质量措施的基础。EHR系统中记录的数据取决于每个医疗保健系统的患者护理实践及其提供者的临床实践。此外，数据收集仅限于在EHR系统或网络中捕获的数据，并且可能不代表全面护理（例如，在卫生保健系统之外获得的护理）。
3. 对于建议使用EHR的前瞻性临床研究，可以修改EHR系统，以通过EHR系统的附加模块在常规护理期间收集额外的患者数据。但是，由于添加能力有限，为了收集大量额外信息的模块，基于EHR的数据收集可能仍不全面。

应在方案中描述选择的数据源用于研究目的的历史经验和使用情况。该描述应包括所选择的数据源在捕获研究要素（例如，入选和排除标准、暴露、结局、关键协变量）方面的表现如何，以及如何对特定研究实践验证数据。

A. 数据源的相关性

世界各地的医学实践和卫生保健系统之间存在差异，这可能会影响数据源与研究问题的相关性。不同类型的商业或政府医疗保健支付计划中的患者可能在一系列特征上有所不同，例如年龄，社会经济状况，健康状况，风险因素和其他潜在的混杂因素。卫生保健系统和保险计划中的各种因素，例如药物分层（例如，一线，二线），处方决策和患者覆盖范围，可能会影响在一个卫生保健系统中接受给定治疗的患者在疾病严重程度或其他疾病特征方面可能与在另一个卫生保健系统中接受相同治疗的患者不同的程度。如果数据来源用于检验研究假设，那么确定数据来源是否涵盖与研究相关的所有人群也很重要。

FDA建议提供：

1. 选择特定数据源以解决特定假设的原因。
2. 关于卫生保健系统的背景信息，包括（如果可用）所关注疾病的任何特定诊断方法和首选治疗，以及在拟定数据源中收集和验证这些信息的程度。
3. 对医疗保健系统（如适用）中处方和使用实践的描述，包括已批准的适应症、制剂和剂量。

非美国 数据来源，FDA建议提供所有这些因素如何影响研究结果在美国人群中的普遍性的解释。

B. 数据采集:一般讨论

EHR系统或医疗索赔数据库中的记录仅在患者与卫生保健系统存在交互时生成。由于EHR和医疗索赔数据是在常规护理期间收集的，而不是根据预先指定的研究方案收集的，因此在EHR和医疗索赔数据源中可能不存在解决拟定研究中某些问题所需的信息。申办方应证明每个数据源包含采集与研究问题和设计相关的人群、暴露、关键协变量、关注的结局和其他重要参数（例如暴露时间、结局时间）所需的细节和完整性。

1. 登记和全面获取护理

当使用EHR和医疗索赔数据源时，应说明覆盖范围的连续性（登记和取消登记），因为当患者经历就业或其他生活环境变化时，他们通常会在不同的健康计划中登记和取消登记。使用这些数据的研究结果的有效性部分取决于患者迁入和迁出卫生计划和卫生保健提供组织的文件。此类文件允许定义入组期（在此期间可获得有关患者的数据）和退出期（当无法获得有关患者的数据时）。应制定入组或连续覆盖范围的定义，并记录在方案中。

此外，FDA建议在获取与研究问题相关的护理和结局方面解决数据源的全面性。该信息将有助于评估收集所有关注的暴露和结局用于监管决策的可能性。例如，不包括住院数据的门诊数据源通常不适合研究可能导致住院的结局。第二个例子是一项研究，其中结果取决于实验室检测的特定频率，而临床医生通常不会以这种频率安排这些检测。

FDA建议说明在研究期间如何采集所有相关人群、暴露量、结局和协变量，特别是在数据可用性随时间变化很大的情况下。数据源应包含足够数量的患者和足够长的随访时间，以根据生物学上合理的时间框架确定感兴趣的结局，如果结局与暴露相关，则预期可能会发生。应提供关于数据源中患者随访时长分布的信息，因为随访时长可能会告知所选数据源是否充分，或者是否需要额外的支持性数据来确定长潜伏期结果。

通常，EHR和医疗索赔数据不能系统地捕获非处方药或根据健康计划无法报销的药物的使用，或工作场所提供的免疫接种。如果这些暴露与研究问题特别相关，则数据来源可能不适用，或者方案应描述如何解决该信息缺口。

获得具有某些类型隐私问题（例如，性传播感染、药物滥用、精神健康状况）的患者的全面药物覆盖和医疗护理数据可能具有挑战性，如果不这样做可能导致信息不完整或错误。

患有这些疾病的患者可以在联邦认可的医疗中心或私人诊所接受治疗，如果使用自付费用，则可能无法产生保险索赔。此外，某些人群更常参加实验性临床试验，例如，患有某些癌症的患者或根据制药公司援助计划接受药物治疗的患者。在这种情况下，大多数电子医疗保健数据源可能无法完全捕获患者的健康数据。如果这些问题与所关注的研究问题相关，则方案应描述如何解决这些问题。

2. 数据链接与整合

数据链接可用于随时间增加个体患者数据的广度和深度，并为验证目的提供额外的数据。如果研究涉及在内部数据源（例如母婴链接）或外部数据源（例如生命记录、疾病和产品注册、生物库数据）之间建立新的数据链接，则方案应描述每个数据源、将获得的信息、链接方法以及随时间推移数据链接的准确性和完整性。如果研究涉及生成额外的数据

（例如，访谈、邮件调查、计算机或移动应用问卷、通过数字健康技术进行的测量），则方案应描述数据收集的方法以及将收集的数据与电子健康护理数据整合的方法。数据链接的概率和确定性方法可能会导致不同的链接质量，尽管这两种方法可能会根据情况而有价值。用于数据链接的确定性方法使用与唯一或一组公共标识符完全匹配的记录，并且可以使用单个或多个步骤过程确定匹配状态。用于数据链接的概率方法使用较少的限制性步骤，其中比较的标识符由较少的变量或其中的一部分组成（Carreras et al.2018）。当使用概率方法时，分析计划应包括检验结果的匹配程度和稳健性的影响。见第VI节，数据收集、管理和转换为最终研究特定数据集期间的数据质量。

对于需要合并来自多个数据源或研究中心的数据的研究，鉴于不同数据源的人群特征、临床实践和编码的潜在异质性，FDA建议证明是否以及如何获得不同来源的数据并以可接受的质量进行整合。

由于患者通常会访问多个医疗保健站点，尤其是在地理位置连续的区域，因此包含来自许多站点的去识别数据会导致单个人可能会有来自不同医疗保健站点的多个记录。在不同站点中存在同一个人的多个记录可能会导致特定数据度量的计数过多，或者，如果某些站点记录不可用，则可能导致收集的患者历史仅反映患者总数的一小部分医疗保健历史。特别注意数据管理，包括个人级别和人口级别的联系以及对多对一和1:1联系的理解，对于评估新数据链接的适当性至关重要。对于在所有站点（例如，多站点医院网络）中共享唯一患者标识符的数据源而言，这种情况不是问题，并且仅在患者在网络外寻求护理时才会发生。FDA建议考虑并记录为解决重复或碎片问题而执行的管理类型，并记录为解决无法通过管理完全纠正的问题而采取的方法。见第VI节，数据收集、管理和转换为最终研究特定数据集期间的数据质量。

3. 分布式数据网络

EHR和医疗索赔数据系统的分布式数据网络（或系统）通常与使用通用数据模型(CDM)结合使用，已越来越多地用于医疗产品安全性监督和研究目的。使用将来自多个站点的数据转换为单个CDM的分布式网络的主要好处是能够对多个数据集执行相同的查询（无需任何或实质性修改）。在一些

分布式数据网络中，查询可以同时在所有网络站点或在每个站点异步运行，并在协调中心合并结果以返回最终用户。分布式数据网络有许多常用的操作模型。一些网络由单个业务实体使用一致的EHR系统或医疗索赔数据库结构进行管理，尽管在许多地点维护数据，但它们的结构和管理方式一致（例如，美国退伍军人事务部退伍军人健康管理局）。另一种方法是混合分布式模型，其中将来自许多远程站点的数据子集发送到集中存储库，以允许对组合数据集进行直接研究（例如，美国疾病控制与预防中心的国家综合征监测系统，以前称为BioSense 2.0）。第三种常用方法是在具有多个所有者和数据库结构的数据系统网络中，不同地点（例如，FDA哨兵系统的成员地点）的数据结构和管理方式不同。在此模型中，将研究问题发送到各个网络成员站点，并将答案返回到中心位置以进行整理和报告。

后一种类型的网络由不同的数据系统（例如哨兵系统）组成，通过使用CDM来实现。使用CDM的网络通常还提供用于分析的工具和方法，一致的数据管理水平以及对数据模型的定期修订，以根据研究人员需要合并新的数据概念。此外，已经开发了允许将数据从一个CDM转换到另一个CDM的能力的方法，但是这些方法涉及额外的数据转换，这提出了额外的质量考虑。数据收集和转换为CDM以及一般QA/QC流程的讨论见第VI节，数据收集、管理和转换为最终研究特定数据集期间的数据质量。

分布式数据网络通常由EHR、医疗索赔或注册数据组成。但是，将许多数据源结合起来，尤其是将数据转换添加到CDM中，会增加一层应该考虑的复杂性。由于分布式健康数据网络有许多不同的配置，因此本指南中讨论的配置不应被认为是全面的。

将不同的数据库结构转换为具有CDM的公共卫生网络，可以跨卫生保健站点进行研究，否则这些站点将更加复杂和昂贵。但是，CDMs可能为研究人员带来额外的挑战。为了满足特定的研究目的，创建了许多CDM，包括为FDA的前哨系统，生物制剂有效性和安全性倡议以及国家以患者为中心的临床研究网络开发的CDM。根据预期用途所需的数据测量类型和详细信息，对CDM中采集的数据的选择进行了优化（例如，用于为监管决策提供信息的上市后安全性监测的哨兵系统，用于以患者为中心的结局研究的国家以患者为中心的临床研究网络）。因此，CDM驱动的数据网络很少包含各个医疗保健站点上存在的所有源信息，并且为给定的CDM网络选择的数据元素可能不足以满足所有研究目的或问题。此外，CDM通常在模型中通常具有许多可选的数据元素，也就是说，尽管模型具有可用于填充数据的此类数据元素，但各个站点可以选择是否将其原始数据放入可选字段。

在使用CDM驱动的网络之前，应考虑CDM收集的数据元素-包括模型中是否存在所需的数据元素，如果存在，它们是必需的还是可选的元素-以确

定研究的适用性，以及是否可以通过补充定制的研究特定数据元素、收集额外数据或使用数据集中存在的其他数据元素来解决已识别的缺陷，这些数据元素是缺失信息的合理替代。应当指出的是，此类变通方案将涉及申办者的其他考虑因素，例如涉及验证替代终点的工作或涉及其他数据的任何人类受试者研究考虑因素。使用更灵活的CDM也可以提高适用性，这些CDM经常扩展以用于新用途。有关代理变量（proxy variables）的信息，请参见第IV.C节，缺失数据：一般注意事项。

4. 可计算表型

标准化的可计算表型可以促进识别相似的患者人群，并能够有效地选择人群以进行跨多个医疗保健系统的大规模临床研究。可计算的表型定义应包括元数据和有关定义、其预期用途、定义的临床原理或研究依据以及在各种医疗保健环境中评估验证的数据的支持性信息（Richesson et al.2016）。由数据元素和表型算法组成的可计算表型定义应在方案和研究报告中描述，并且还应以计算机可处理的格式提供。可计算表型定义的临床验证应在方案和研究报告中描述。有关验证的其他信息，请参见第IV.D节，验证：一般注意事项。

5. 非结构化数据

大量的关键临床数据是EHR内的非结构化数据，既可以作为自由文本数据字段（例如医师笔记），也可以作为计算机文件中的其他非标准化信息（例如基于PDF的放射学报告）。为了提高数据抽象（data abstraction）的效率，越来越多地使用包括现有技术和新兴技术在内的一系列方法来将非结构化数据转换为可计算格式。最近的创新包括技术支持的抽象，其中软件为人类数据抽象者（例如肿瘤注册者）提供了一种机制，以一致且可扩展的方式进行其工作。

人工智能(AI)领域的技术进步可能允许更快速地处理非结构化电子健康护理数据。进步包括自然语言处理，机器学习，尤其是深度学习，以：（1）除了EHR中的结构化字段外，还从非结构化文本中提取数据元素；

（二）开发计算机算法，确定结局；或（3）评估图像或实验室结果。FDA不认可任何特定的AI技术。

所有这些方法都是计算机辅助的各个级别，但目前需要大量的人工辅助管理和决策，从而将额外级别的数据可变性和质量考虑因素注入最终的特定于研究的数据集中。如果方案建议使用AI或其他推导方法，则方案应指定所用计算机算法的假设和参数、信息用于构建算法的数据来源、算法是否受到监督（即，使用专家的输入和审查）或无监督，以及与方法验证相关的指标。对数据质量的相关影响应记录在方案和分析计划中。

C. 缺失数据:一般考虑

在两种广泛的情况下，数据源可能缺少信息。第一种情况是打算收集信息（例如，EHR中存在的结构化字段），但数据源中不存在。这是传统缺失数据的一个例子。第二种情况是，该信息不打算在EHR和医疗索赔数据中收集，因此不存在。重要的是要区分这两种情况，并了解EHR和医疗索赔中存在或不存在信息的原因。例如，缺乏有关实验室检测结果的信息可能是由不同的情况引起的：（1）医疗保健提供者可能没有开具该检验；

（2）可能已开具但未进行的检验；（3）可能已经进行了检验，但结果未存储或捕获在数据源中；或（4）可能已经进行了检测并将结果存储在数据源中，但是数据不是可访问的格式，或者在生成最终研究特定数据集时在转换和管理过程中丢失。由于提供者可能会根据患者的特征开具实验室检测，因此不开具检测的决定或患者放弃检测的决定可能会影响数据在拟定研究中的适用性。

如上所述，数据链接是解决丢失数据的一种方法。也可以识别作为缺失数据代理的变量。潜在的代理变量的一个例子包括医疗保险D部分处方药计划下的低收入补贴，作为患者社会经济地位的代理。

应根据对存在和不存在信息的原因的理解制定方案和统计分析计划。应包括描述性分析，以描述缺失数据的特征。应支持关于研究终点和重要协变量的统计分析基础的缺失数据（例如，随机缺失，非随机缺失）的假设，并考虑缺失数据的含义。

D. 验证:一般注意事项

使用EHR和医疗索赔数据源的研究应包括重要研究变量的概念定义，包括研究人群入选和排除标准、暴露、结局和协变量。概念定义应反映当前的医学和科学关于感兴趣变量的思考，例如：（1）定义人群选择条件或作为感兴趣结果或协变量的临床标准；或（2）测量药物摄入量以确定感兴趣的暴露。

应基于概念定义制定操作定义，从数据源中提取最完整和准确的数据。在许多使用EHR或医疗索赔数据的研究中，操作定义将是使用结构化数据元素的基于代码的电子算法。在其他研究中，操作定义可以从非结构化数据中提取相关信息或构建结合结构化和非结构化数据元素的算法中得出。操作定义也可以指定额外的数据收集，例如患者调查（在适当时）。

由于操作定义通常是不完善的，并且无法对每个主题的目标变量进行准确分类，因此导致的错误分类可能导致假阳性和假阴性（表1），并可能在一定方向和程度上使暴露与结果之间的关联产生偏差。尽管对感兴趣变量的完全验证¹⁰可以最大程度地减少错误分类并最大程度地提高研究内部有效性，但了解潜在错误分类对研究内部有效性和研究推断的影响，是确定

感兴趣变量可能需要验证以及验证程度的关键步骤。例如，在量化药物作用的研究中，应优化内部有效性，并应最小化关键变量的错误分类，以准确测量关联。当预期错误分类的存在不会改变结果的解释时（例如，对于信号检测，或者当假设的效应量很大并且错误分类对关联度量的影响被认为是最小的时），某些错误分类在某些研究中可能是可以容忍的。）。

为了理解所关注变量（例如，暴露、结局、协变量）的潜在错误分类如何影响关联性测量和结果解释，申办方应考虑：（1）错误分类的程度；（2）差异错误分类与非差异错误分类（例如，按暴露对结果进行差异错误分类）；（3）依赖与独立的错误分类（例如，当在同一调查中自我报告两者时，暴露和结果的相关错误分类）；（4）暴露与结果之间的关联可能偏向的方向。

尽管研究变量的完全验证被认为是最严格的方法，但在某些情况下，为每个受试者验证变量可能不可行（例如，非常大的研究人群，缺乏所有研究受试者的参考标准¹⁰数据），并且评估变量的操作定义的性能可能就足够了。基于表1中描述的性能指标，申办方应考虑是否需要在更大程度上验证变量（例如，按操作定义分类的所有阳性结果），并与相关审评部门讨论。

由于操作定义的性能取决于各种因素，例如数据来源、研究人群、研究时间框架和参考标准的选择，FDA建议使用合理的抽样方法（例如，随机抽样、分层抽样）在足够大的研究人群样本中评估操作定义的性能，作为拟定研究的一部分。如果申办方建议使用在先前研究中评估过的操作定义，理想情况下应选择在相同数据源和相似研究人群中评估过的操作定义。此外，疾病、诊断和编码的长期趋势可能需要使用最近的数据评估操作定义。应始终评估用于建立灵敏度、特异性和预测值的先前研究的质量。

方案应包括计划验证的详细描述，包括参比标准品的选择依据、验证方法、方法、工艺和取样策略（如适用）。如果提出了先前评估的操作定义，则应提供额外的信息，包括哪些数据来源和研究人群以及在什么时间范围内进行评估，评估的性能指标的价值以及关于性能指标是否适用于拟定研究的讨论。FDA还建议在方案中纳入预先规定的敏感性分析，以证明偏倚（如果存在）是否以及如何影响基于验证数据的研究结果。

关于研究设计元素验证的进一步讨论，见第V.C.5节，暴露验证；第V.D.3节，结局确认；和第V.E.3节，混杂因素和效应修饰因子的验证。

10出于本指南的目的，完整的验证涉及根据所选择的参考标准为每个研究受试者的目标变量分配一个准确的值。例如，病历审查可与概念性定义结合使用，以确定受试者是否符合关键入选标准或是否经历了结局事件。（在不同程度上，该过程可能涉及裁决。）

11在本指南中，参考标准是最佳可用基准，也称为“金标准”。

表1:二进制变量的灵敏度、特异性、阳性预测值(PPV)和阴性预测值(NPV)的计算示意图

基于拟定操作定义的条件	基于参比标准品的条件		合计	
	Yes	No		
Yes	a (真阳性)	b (假阳性)	a+b	PPV= $a/(a+b)$
No	c (假阴性)	d (真阴性)	c+d	NPV= $d/(c+d)$
合计	a+c	b+d	N	
	灵敏度= $a/(a+c)$	特异性= $d/(b+d)$		

V. 研究设计要素

关键研究设计要素的确定和验证在下文中进行了详细讨论。应首先确定感兴趣的研究问题，然后确定最适合解决这些问题的数据来源和研究设计。研究不应设计为适合特定的数据源，因为特定数据源的局限性可能会限制研究设计的选择，并限制可以得出的推论。关于使用RWD来源时研究设计和分析的考虑因素将在其他RWE指南文件中讨论。

A. 时间段定义

FDA建议在方案中明确定义与研究设计相关的不同时间段（例如，确定研究人群、定义入选和排除标准、评估暴露、评估结局、评估协变量、随访患者的时间段）。应明确描述时间尺度的重点（例如日历时间、年龄、暴露后时间），并充分详细说明回答研究问题所需的时间单位（例如年、月、日、小时、分钟）的数据可用性。

方案应说明拟定的时间段和对研究有效性的潜在影响的合理性。例如，应提供关于暴露前的时间段是否适合确定研究人群和重要的基线协变量，随访时间是否足以观察研究结果的发生，以及更新时间依赖性协变量信息的时间段是否适合捕获这些变量的变化的理由。此外，在考虑结局定义时，疾病发作（例如，早期症状）可能需要与已确诊的诊断区分开，视研究问题而定。在定义结局评估的随访时间的开始和结束时，应考虑预期结局（如果与暴露相关）可能发生的生物学上合理的时间范围。

方案还应说明标准治疗的潜在时间变化、其他治疗的可用性、诊断标准以及与研究问题和设计相关的任何其他相关因素。其他相关因素可能包括保险处方变更（如果已知）、分步治疗和实验室分析变更。在开发研究方法之前，申办方应与相关的FDA审评部门讨论数据捕获这种潜在时间变化的能力以及潜在时间变化的影响。

B. 研究人群的选择

方案应包括确定如何实施入选和排除标准（例如，人口统计学因素、医学状况、疾病状态、严重程度、生物标志物）的方法的详细描述，以从数据源中识别出符合这些标准的适当患者。方案应说明在拟定数据源中收集的信息的完整性和准确性，以满足入选和排除标准。

应验证用于选择研究人群的关键变量。例如，为了评估在免疫性血小板减少性紫癜患者中的药物作用，应根据疾病的概念定义（包括体征和症状、血小板水平以及排除其他可能的血小板减少症原因）验证由操作定义国际疾病分类，第九版，临床修订（ICD-9-CM）诊断代码287.31确定的疾病。

在某些情况下，满足入选和排除标准所需的关键变量（例如，妊娠研究的胎龄）可以由医疗保健提供者使用护理点可用的信息生成。例如，医疗保健提供者可以根据患者自我报告的最后一次月经、超声检测和其他相关信息，在EHR中输入计算出的胎龄。如果使用此类数据，则方案应描述信息来源和医疗保健提供者用于生成数据的方法（如果已知）。

C. 暴露确定和确认

本节中讨论的关于医疗索赔数据或EHR中暴露确定的考虑因素主要适用于非干预性研究，因为在干预性研究中记录了暴露分配。

1. 暴露定义

出于本指南的目的，术语“暴露”（**exposure**）适用于在拟定研究中评估的关注的医疗产品或治疗方案。感兴趣的产品称为治疗，可以与无治疗、安慰剂、标准治疗、另一种治疗或上述组合进行比较。将可能影响研究结果的其他变量视为协变量，并在第V.E节“协变量确定和验证”中进行讨论。暴露定义应包括有关药物剂量、制剂、规格、给药途径、给药时间、给药频率和所研究产品持续时间（如果相关）的信息。可能还需要描述产品的特定生产商（例如，当不同生产商使用疫苗的专有名称时）。

暴露的描述应包括产品的预期用途或规定用途（例如，数量、频率和特定剂量）、暴露开始与可能合理预期观察到效应的最早时间之间的时间，以及预期的效应持续时间。这通常将需要了解药物的药理学特性--例如，预防骨质疏松症的一次性输注可能会产生数月的作用。见第V.C.3节，暴露的确定:持续时间和第V.C.4节，暴露的确定:剂量。

2. 暴露的确定:数据来源

申办方应能够证明在拟定数据源中识别特定目标产品的能力，证明该数据源包含允许识别特定目标产品的数据字段和代码（例如，通过特定编码）。例如，并非总是能够单独从计费或诊断代码推断出特定的疫苗配方，例如在单个计费代码用于多种疫苗的系统中。方案应描述使用的编码系统、表示的粒度水平（例如，使用映射到国家药品代码[NDC]标识符的RxNorm）以及编码系统达到的特异性。

当依赖编码数据时，操作暴露定义应基于所选数据源的编码系统，并反映对该数据源中药物的处方、交付和报销特征（如果适用）的理解。例如，在美国，操作定义应包括适当的药房代码（NDC或医疗保健通用程序编码系统），以捕获药物在各种环境中的使用情况。这种方法在非口服药物可能根据获取方式分配不同代码的情况下尤其重要。例如，使用注射药物的患者可以从药房购买，在这种情况下，将记录NDC代码，或者可以由提供者为患者给药，并使用HCPCS J代码记录药物及其给药。¹²

在合并来自非结构化和结构化数据的信息时，报告操作定义和方法也是必不可少的。新兴的方法可能涉及审查病历中的非结构化信息，并结合药房配药和医生处方数据和注释，以评估一个人是否开了处方并接受了感兴趣的药物，以及患者继续用药是否存在问题。在对接受常规护理结肠镜检查的退伍军人进行的回顾性队列研究中确定阿司匹林暴露时发现了此类方法的一个示例（Bustamente et al.2019）。

当使用医疗索赔数据源时，重要的是要考虑，如果这些未获取的处方与研究的暴露相关，则可能会分发与保险索赔无关的处方。未捕获的处方可能包括低成本的仿制药，通过折扣计划获得的药物，制药公司提供并由医疗保健提供者分发的样品，以及通过互联网或患者自付费用购买的药物。此外，非处方药和膳食补充剂通常不会在电子医疗保健数据库中捕获。重要的是解决不完全暴露确定的可能性及其对研究有效性的影响，见第V.C.5节，暴露验证。

¹²药物的J代码是医疗索赔中使用的医疗保健通用程序编码系统II级代码，用于报告通常不能自我给药的注射药物；化学疗法，免疫抑制药物和吸入溶液；还有一些口服药物。

3. 暴露的确定:持续时间

数据源应记录相关的暴露持续时间（产品随时间的预期使用）。鉴于某些医疗产品被设计为一次性暴露（例如疫苗），而其他产品可能预期在较长时间内使用，数据源的适用性将随所研究的特定医疗产品而变化。FDA建议描述暴露的持续时间以及暴露对关注结果产生影响的时间段。

持续时间可指连续暴露或累积暴露，取决于研究问题。对于某些产品，预期会产生即时或近乎即时的效果；对于其他产品，预期在一定时间间隔后会产生效果（例如，促进骨骼强度的药物）。FDA建议考虑治疗中止后持续药效的持续时间，以包括可能发生药效的整个时期。例如，疫苗效果可能会在疫苗接种后持续数年，并且可能会认为在此期间暴露于疫苗。另一方面，抗凝剂的作用不会超过几个小时或几天。FDA还建议证明选择用于估计暴露持续时间的单位（例如，小时，天）的合理性，并确保以这些单位获得数据。

由于患者可能无法完全按时补充处方，或者可能会提前补充处方，因此治疗中可能存在缺口或库存，并且可能会在数据中反映出来。¹³FDA建议在方案中描述和证明研究人员如何测量使用情况，解决数据源中治疗中的潜在缺口，并在有早期补充的情况下处理补充库存。间歇性治疗（例如，按需治疗疼痛的药物）和经常向患者提供样本的治疗（例如，昂贵的药物，新上市的药物）在准确评估实际暴露量和暴露持续时间方面存在挑战，见第V.C.5节，暴露量的验证。

4. 暴露的确定:剂量

暴露数据应包括剂量信息。根据研究中的暴露和关注的问题，描述每次给药剂量或每日剂量以及估计的累积剂量可能是有用的。

合理的做法是从数据源中提供的剂量信息开始，然后在方案或研究报告中讨论在估计目标暴露剂量时做出的具体假设，尤其是对于儿科患者。见第V.C.6节，特殊人群中的给药。同样重要的是，如果有多种剂型，报告不同剂型（例如，肠胃外与口服）将如何纳入剂量计算。

¹³本指南不涉及与药物依从性相关的问题。

¹⁴在某些情况下，RWD来源可能是唯一的参考。例如，如果通过患者是否支付处方费用来定义暴露，则可以使用医疗索赔数据，并且该信息将作为参考来源。

5. 暴露验证

除了在医院环境或输注环境中给予的药物以外，电子医疗保健数据捕获药物处方和向患者分发药物，但通常不捕获实际的患者药物暴露，因为这取决于患者获得和使用规定的治疗。

理想情况下，验证涉及将拟定数据源中的暴露分类与参考数据源¹⁴进行比较，并产生可用于敏感性分析的错误分类估计值。验证可以从定义概念和操作定义开始。

例如，为了定义药物X在特定研究中的新用途，概念定义可以是“开始使用药物X，并且在过去365天内没有暴露于药物X”，操作定义将是“至少一次针对药物X的门诊处方索赔（通过NDC代码xxx识别），并且在处方分发日期之前的365天内没有针对药物X的索赔”。对于在门诊环境中使用的处方药物，通过记录处方已填写，配药或计费数据在反映药物暴露方面往往比大多数EHR更准确。在这种情况下，可能需要通过检查医疗索赔数据来验证EHR处方数据。对于在医疗保健环境中给药的药物（例如，疫苗、注射剂、血液制品），EHR中记录的给药可能提供比医疗索赔记录中更完整的信息。在这些情况下，通过检查EHR来验证医疗索赔数据可能是有用的。在某些情况下，当参考数据来源不可用时，在相同人群中进行的其他研究或在文献中发表的其他研究可以提供暴露状态潜在错误分类的估计值（例如，对研究参与者进行调查以评估药物的摄入量，已发表的通过药房/工作场所/学校获得疫苗接种的人数报告）。

FDA建议记录用于计算和验证持续时间、剂量、转换和其他暴露特征的方法。应在适当的研究文件中解决验证和错误分类问题。

6. 特殊人群给药

除了报告关于处方、分发或给药剂量的经验证的信息外，可能还需要额外的信息来评估给药是否适用于特殊人群（例如，如果存在明显的剂量不足或过量）。例如，在评估服用具有大量肾脏清除率药物的患者的剂量时，可能需要获得血清肌酐、肌酐清除率或估计的肾小球滤过率的测量值，以评估给药的适当性。另一个例子是估计儿科人群的暴露量时，可能需要获得患者的体重并描述体重类别中的剂量。对于索赔数据，可能更频繁地需要额外的数据以允许对剂量进行适当的评估，但如果没有必要的信息，则在使用EHR时也可能发生。

7. 其他注意事项

选择合适的对照药物是临床研究的重要组成部分。应在方案中明确定义提供对照数据的信息，并提供充分的详细信息。方案应讨论选择以下内

容的理由：（1）对照数据的来源；和（2）时间段（如果对照组与治疗组不同步）。由于对照药物可能与目标产品在疾病、禁忌症、安全性特征或使用者的疾病严重程度或合并症以及其他患者特征中的特定适应症不同，因此确保有足够的数据可供FDA评估暴露人群和对照人群的可比性非常重要。

应从数据源中描述和确定相关的合并用药使用情况。应详细描述研究中伴随用药的定义。定义伴随用药可能包括在同一天分发药物，药物供应时间重叠或患者在研究期间已填写两种或多种药物的处方的情况。还应描述确定合并用药（例如非处方药）的局限性。

D. 结果确定和验证

选择数据源的关键步骤是确定它是否捕获了感兴趣的临床结果。由于电子健康护理数据通常捕获引起医疗保健专业人员注意并记录在医疗记录中的结局，因此代表在医疗护理之外发生的轻度症状或事件（例如，院外死亡）的结局通常不会被很好地捕获。相反，离散的结局或急性事件（例如中风，心肌梗塞，新感染）比现有问题（例如抑郁症，牛皮癣，关节炎）的恶化更有可能被捕获，而现有问题（例如抑郁症，牛皮癣，关节炎）的恶化不会导致可识别的事件。与传统的临床试验不同，仅使用电子医疗保健数据来确定结果的研究可能没有方案定义的随访访视，并且可能没有在确定结果所需的时间间隔对事件进行监测。此外，在传统的临床试验中，对感兴趣的结果的评估可能更加标准化和全面。因此，关于所关注结果的数据的可用性、准确性和完整性以及外部数据链接的需求应谨慎考虑。应在研究开始之前评估数据源是否以及在何种程度上捕获关注的结果，并且与关注的暴露无关。

1. 关注结局的定义

许多结果涉及医生记录的诊断，作为常规护理的一部分。为了尽量减少不同医生在实践中和随时间变化的影响（例如，使用不同的诊断和分类标准，以不同的方式对同一事件进行编码），FDA建议根据疾病的临床、生物学、心理学和功能概念（如适用）定义感兴趣的结局。关注结局的概念定义（也称为病例定义）应反映对疾病的医学和科学理解，并可能因研究而异。例如，对于过敏反应，概念定义（或病例定义）可能包括以下临床标准：突然发作、体征和症状快速进展、 ≥ 1 项主要皮肤病学标准和 ≥ 1 项主要心血管或呼吸标准。方案应包括概念定义的详细描述，包括确认结局的体征、症状以及实验室和放射学结果。

概念性定义应该能够在RWD源中操作。例如，肿瘤学中的随机对照试验通常在设定随访评估的特定时间和频率时使用基于肿瘤的关注结局，并且通常包括在临床实践中可能不是标准治疗的分子或生物标志物检测。由于达

到客观缓解（肿瘤缩小）或基于标准化临床试验标准（例如，RECIST 1.1）的肿瘤进展日期通常不会在RWD来源中记录，因此可能需要探索替代指标或多组分定义，并证明其使用的合理性。一般来说，与在性质上更具主观性或标度性的结果（例如类风湿性关节炎的关节痛恶化或抑郁症状恶化）相比，可能更容易获得具有明确诊断标准且可能在RWD中可能始终采用的标准（例如中风，心肌梗塞或肺栓塞）。申办方应与FDA审评部门讨论拟定的结局定义。

2. 结局的确定

为了帮助识别所选数据源和研究人群中的潜在病例，应基于所关注结局的概念定义，制定使用诊断和程序代码（例如，ICD-9-CM、ICD-10）、实验室检查（例如，LOINC）和数值或非结构化数据（例如，医生的就诊记录、放射学和病理学报告）的操作定义。如果操作定义包括从EHR或其他数据源中的非结构化数据中提取的信息（例如，在出生证明记录中提及用于识别婴儿神经管缺陷的脊柱裂），则方案应提供用于处理非结构化数据的方法和工具以及这些方法的验证的详细描述和原理。有关非结构化数据的其他信息，请参见第IV.B.5节，非结构化数据。当建议患者或医生生成的数据（例如，主观终点所需的数据）用于评估感兴趣的结局或补充操作定义时，方案应规定如何定义、构建和验证结局指标（例如，体征评分、严重程度指数）（如适用），以及如何收集数据。

当结果分类错误时，操作定义的敏感性和特异性是不完善的。鉴于通常不可能使敏感性和特异性达到完美（即100%），因此结果错误分类可能会导致假阳性和假阴性。FDA建议在制定或选择拟定研究的操作定义时，考虑结局错误分类对研究有效性的潜在影响。例如，当在队列研究中研究不经常发生的结局时，鉴于结局事件的患病率较低，重要的是实现高特异性以最小化假阳性病例和高灵敏度，从而可以捕获更多真实病例。

由于数据库特定的敏感性和特异性以及可变的疾病患病率，为一个数据源或研究人群开发的操作定义在其他来源或人群中可能表现不佳。阳性预测值(PPV)和阴性预测值(NPV)与敏感性和特异性相关，并且是测量预测值的人群中结果患病率的直接函数。因此，PPV和NPV因数据来源和研究人群特征（例如人口统计学因素、基础疾病、合并症、临床环境）而异。

方案应包括操作定义、编码系统、选择用于构建操作定义的信息的原理和相关局限性的详细描述（例如，选择顺序可能与它们的医学重要性不对应的主要或次要诊断代码），以及对结果错误分类的潜在影响。如果在先前的研究中已经评估了操作定义的性能，则应讨论对拟议研究的适用性。此外，由于先前研究中用于确定敏感性，特异性和预测值的病例定义可能会包括与为拟定研究开发的概念定义不同的诊断标准，应仔细考虑先前研究中评估的性能指标的正确使用。

3. 结果验证

FDA期望对结果变量进行验证，以尽量减少结果错误分类。尽管对结果变量的完全验证被认为是最严格的方法，但在某些情况下，验证每个受试者的结果可能不可行，并且评估结果的操作定义的性能可能就足够了。

结果验证涉及使用临床上适当的概念性结果定义来确定患者的状态（按操作定义分类）是否真正代表了感兴趣的结果，通常是通过查看以电子或纸质格式记录在患者病历中的临床细节来进行的。

FDA建议使用标准化病历审查流程，包括使用标准化工具、流程文件和人员培训。标准和可重复的过程对于最小化评估者内和评估者间的变异性至关重要，特别是对于通常无法在系统之间共享病历且无法进行集中式病历审查的多中心研究。即使采用集中式病历审查，标准化过程也有助于确保不同裁定员或单个裁定员随时间应用相同的标准。报告比较指标（例如，kappa统计量）有助于确保可复制性。应在方案中说明估计的病历检索率的合理性，并应讨论对内部和外部有效性的影响。此外，由于对患者暴露状态的了解可能会影响观察者并导致不同的错误分类，因此应通过掩盖研究问题或编辑暴露信息来考虑对暴露状态设盲的抽象人和裁决人，尤其是当抽象人（abstractor）和裁决人（adjudicator）可能将暴露与关注的结果相关联时。方案应提供如何处理观察者偏倚的描述。

理想情况下，通过对结果变量的完整验证，为每个受试者分配结果变量的准确值，以最大程度地减少结果错误分类并提高研究内部有效性。在实践中，更常用的方法是评估验证研究中操作定义的性能。诸如敏感性、特异性和预测值之类的性能指标无法准确地对案例和非案例进行分类；相反，它们可以告知结果错误分类的程度，并在存在错误分类的情况下促进结果的解释。

通常在验证研究中评估PPV。PPV是通过操作定义确定为真阳性病例的潜在病例的比例。因此，PPV告知在确定的病例中包括假阳性病例的程度。当对假阴性病例的关注可以忽略不计时（例如，当灵敏度被认为足够高以使假阴性病例的数量最少时），高PPV可能足以提供对结果变量有效性的置信度，而中等至低PPV可能需要对所有潜在病例的结果变量进行完整验证。当范围对于假阳性病例和假阴性病例的程度是值得关注的，申办方应考虑评估定量偏倚分析所需的所有性能指标，以评估结局错误分类对关联性指标的影响，或采取更严格的方法，通过验证所有潜在病例和非病例的结局变量，对每个受试者的结局变量进行准确分类。总之，所要求的验证程度应由必要的确定性水平和潜在的错误分类对研究推断的影响来确定。

一般而言，申办方在选择操作定义时应考虑假阳性和假阴性病例之间的权衡，并确定适当的结果验证方法以支持内部有效性。例如，为了识别婴儿的神经管缺陷，包括一系列住院和门诊诊断代码的操作定义可能具有高灵敏度，低特异性和低PPV。将操作定义仅限于住院患者诊断代码或诊断和程序（例如，手术修复）代码的组合可能会增加PPV，但会错过相当大比例的真实病例（灵敏度低）。由于缺少真实病例对于不经常报告的结果尤其令人担忧，因此一种方法是选择高灵敏度的操作定义，并对所有潜在病例进行结果变量的完整验证，以最大程度地识别出所有真实病例并导致假阳性的可能性通过验证将阳性病例降至最低。与罕见疾病结局不同，当关注的结局涉及更常见的事件（例如，疾病特异性住院）或病情改善或恶化时，常见诊断的操作定义可能在相当程度上产生假阳性和假阴性病例，因为真实病例和真实非病例均普遍存在。因此，可能难以获得准确和完整的信息（例如，实验室测试结果，功能性测量）以用于操作定义以准确地分类病例和非病例。对于此类结局，仅测量PPV将是不足以告知结果错误分类。

在无法完全验证每个研究对象的结果变量的情况下，应使用合理的抽样策略在拟定的研究人群中评估操作结果定义的性能。如前所述，在先前研究中评估的操作定义的使用在理想情况下应在相同的数据源和相似的研究人群中使用，因为操作定义的性能可能因数据源和研究场景而有很大差异，如果疾病、诊断和编码存在长期趋势，则可能需要更多的最新数据。应评估用于建立灵敏度、特异性和预测值的先前研究的质量。特别是，在先前研究中用于建立这些指标的病例定义应与为拟议研究开发的概念性结果定义兼容。应证明这些指标对拟定研究的适用性，并可考虑进行敏感性分析。

在没有完整的患者信息和结局变量的完整验证的情况下，结局错误分类仍然是研究内部有效性的威胁，并且对暴露与结局之间关联测量的影响取决于暴露组之间错误分类的程度是否不同。差异错误分类涉及暴露组之间敏感性，特异性和疾病患病率差异的复杂相互作用，因此可能使关联偏向或远离零值。因为很难预测偏倚的方向，差异错误分类是安全性和有效性研究的一个问题。与微分错误分类不同，非微分错误分类倾向于使关联偏向零值。因此，在安全性研究中可能会遗漏真正的风险，而在有效性研究中可能需要更大的研究人群来证明药物的作用。

当结局定义未充分完善且包括与关注暴露不统一相关的情况时，可能会发生非差异结局错误分类。例如，神经管缺陷包括原发性神经形成缺陷和神经形成后缺陷。原发性神经缺陷直接归因于原发性神经衰竭（即神经管闭合），其发生在受精后约18至28天之间。对神经后缺陷的病理生理学了解较少。因此，在妊娠初期的关键时期内的药物暴露可能不会以相同的方式影响后期的神经。当结局定义包括初级和后神经期时，可能无法检测到初

级神经缺陷的风险（如果有的话）。

在暴露状态设盲的研究中，差异结果错误分类可能会最小化。但是，即使当数据收集方法似乎排除了差异结果错误分类的可能性时，在特定研究的实际数据中也不能保证非差异结果错误分类。例如，观察、诊断和记录结局是否发生的医生可能是决定哪些患者接受了旨在预防结局的治疗的同一医生，或者该医生可以在了解他们接受了哪种治疗的情况下，以不同的方式监测疾病进展或治疗副作用。如果数据包括在公示日期之后发生的事件，则特定药物安全性问题的公示也可能导致偏倚错误分类。因此，当使用真实世界的的数据时，结果错误分类偏差的方向可能仍然无法预测。此外，当一项研究中存在一个以上的错误分类时，申办方应考虑它们之间的关系。例如，尽管非差异暴露错误分类和非差异结果错误分类均可能使关联偏向零值，但当两个错误分类是相关的时，总体而言，它可能产生远离零值的偏差（Lash et al.2009）。因此，当评价潜在错误分类对研究推断的影响时，申办方应避免过度依赖偏向零值的非差异错误分类。在这种情况下，根据拟定研究人群中的暴露状态评估操作结果定义的性能可能是必要的。

关于结局验证，申办方应证明拟定的验证方法的合理性，例如验证所有潜在病例或非病例的结局变量，而不是评估拟定操作定义的性能；如果将进行后者，则说明将评估哪些性能指标的合理性。方案应包括结局验证设计、方法和过程以及抽样策略（如适用）的详细描述。如果提出了先前评估的操作定义，则应提供额外的信息，包括：（1）数据来源和研究人群；（2）在什么时间范围内进行验证；（3）性能特征；（4）评估性能的参考标准；（5）讨论先前的验证数据是否适用于拟议的研究。

FDA建议在方案中纳入定量偏倚分析作为敏感性分析，以证明结局错误分类是否以及如何影响研究结果。方案应预先规定将用于定量偏倚分析的指标（例如，灵敏度、特异性、PPV、NPV），并描述如何在结果验证中测量所选指标。

4. 死亡率作为结局

在美国，死亡和死亡原因通常不包括在电子健康护理数据中，但在患者接受医疗护理时发生的死亡除外。

可以通过与公共或商业生命统计数据源联系来确定死亡（死亡事实和死亡原因），以增加死亡变量的完整性和新近性。然而，外部死亡率数据的使用受到此类数据和数据链接方法的所有限制（Haynes 2019; Navar et al.2019; Curtis 2018）。死亡率数据质量及其影响的仔细记录应包括在方案中。

如果在电子医疗保健数据系统中未记录死亡，则在暴露于研究药物后死亡

的患者可能会在电子医疗保健数据中被观察为未提出任何进一步的医疗索赔或在特定日期后未接受任何额外的护理。对于所关注的结局或结局（例如，心肌梗死或卒中）包括致死性结局的研究，排除在暴露于研究药物后任何时间似乎失访的患者可能会产生偏倚。应将这些患者纳入生命统计系统的检索中，以查看他们是否因死亡而退出系统（退出），并且可能有必要在缺乏相反数据的情况下将其死亡分类为感兴趣的结果。

E. 协变量确定和验证

出于本指南的目的，特定研究中的协变量可以包括两种类型的元素：混杂因素和效应修饰因素。

1. 混杂因素

在一项非随机研究中收集了有关潜在混杂因素的信息，以支持在分析中平衡治疗组和对照组的适当努力。用于识别和处理研究中混杂因素的流行病学和统计学方法将在RWE研究设计的未来指南文件中讨论。

在研究中识别出潜在的混杂因素后，应对拟定的数据源进行评估，以确定其是否足以捕获可能导致混杂的重要因素的信息。这些包括在拟定数据源中被充分捕获的混杂因素（测量的混杂因素）和未被充分捕获的混杂因素（未测量的混杂因素或测量不完善的混杂因素）。在电子医疗保健数据（尤其是索赔数据）中可能无法测量或不完全测量的混杂因素示例包括人种/种族、疾病家族史、生活方式因素（例如吸烟、饮酒、营养摄入、体力活动）、某些体格测量（例如体重指数）、在没有保险的情况下获得的药物和药物使用适应症。FDA建议考虑与其他数据源的潜在联系或额外的数据收集，以扩大对原始数据源中未测量或测量不完善的重要混杂因素的捕获。

2. 效应修饰剂

药物有效性或安全性的研究通常报告平均治疗效果，即使相同的治疗在不同人群中可能具有不同的效果。关于潜在效应修饰因子的信息用于更好地理解治疗效应的异质性，即人群中个体治疗效应方向和程度的非随机、可解释的变异性（Velentgas et al.2013）。在研究中应检查人口统计学变量（例如年龄、性别、人种、种族）或相关合并症对效应进行修改的可能性，并且在选定的数据源中应提供相关的效应修改器。

3. 混杂因素和效应修饰因素的验证

对于所有关键协变量，包括混杂因素和效应修饰因素，FDA建议在方案和研究报告中提供操作定义的有效性并证明其合理性。如果测量的协变量在

患者的随访期间可能发生变化（时变协变量）并且对分析很重要，则方案应描述是否可以采集时变协变量的信息以及采集时变协变量的频率，特别是因为RWD中时变协变量的采集可以根据疾病的严重程度进行区分（例如，在病情较重的患者中进行更多检测）。

在评估协变量操作定义的有效性时，FDA建议根据协变量的性质确定最佳参考数据源。在验证医疗事件或手术应用（例如，合并症、既往病史）等协变量的操作定义时，适用与第V.D.3节“结局验证”中相同的原则。

关于验证与药物使用相关的协变量的操作定义的讨论，如合并用药或既往药物使用，见第V.C.5节，暴露验证。当评估其他协变量操作定义的有效性时，例如疾病家族史、生活方式因素或药物使用指征，适当的参考可能包括患者或提供者调查或适当的数据关联。

当需要补充信息来捕获重要的协变量或用于协变量验证时，FDA建议描述获得总体研究人群补充信息的可能性。如果该补充信息仅适用于部分研究人群，FDA建议在相关研究文件中讨论对内部有效性的潜在影响。

VI. 数据累积、管理和转换为最终研究特定数据集期间的数据质量

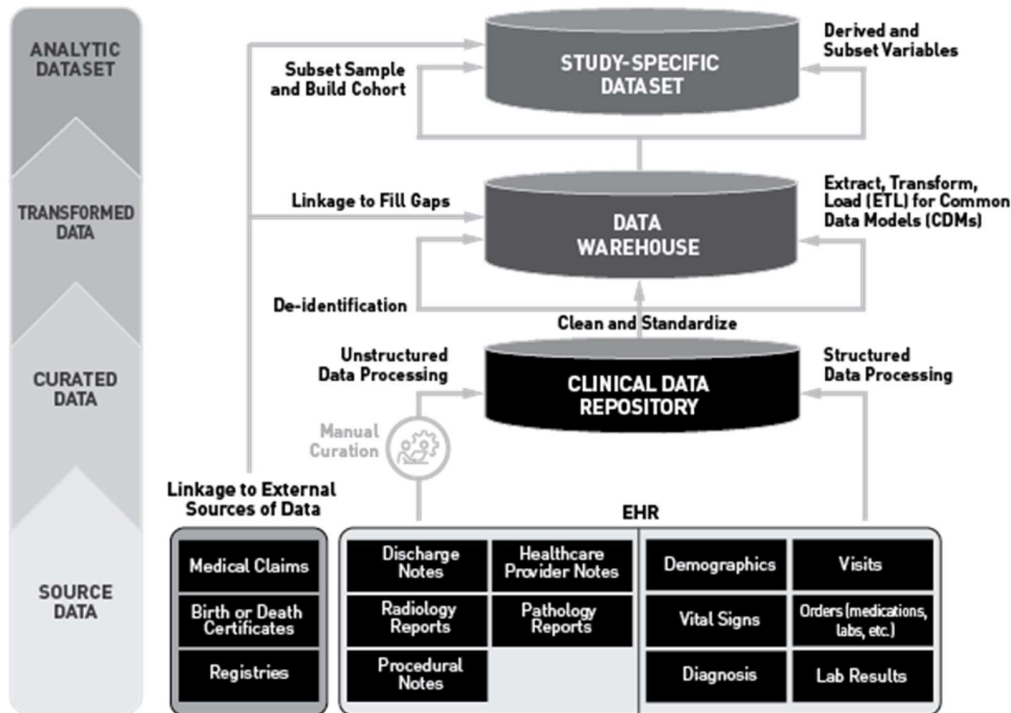
本节讨论了在数据生命周期过程中检查数据质量时需要考虑的要点。尽管数据生命周期可能会根据数据类型和设置（即，医疗保健设置，例如药房，诊所，急诊科和医院）而有所不同，但通常，生命周期涉及多个阶段：从原始源数据中应计的数据；管理临床数据存储库的数据；必要时对数据进行转换和去识别，创建数据仓库；并生成用于分析的研究特定数据集（见图1）。

数据生命周期的概念说明了检查数据质量的过程的迭代性质。过程不是一次性考核；相反，这是一个持续进行的过程，在该过程中，在周期的每个阶段进行数据质量检查、清理¹⁵和监测，并且可以重复进行一些检查（即，在周期的多个阶段进行）。

¹⁵数据清理（有时称为数据清理）是从数据库中纠正或删除不准确数据（或格式不正确，重复的数据或记录）的过程。需要纠正/删除的数据有时被称为“脏数据”。数据清理是保持数据质量的一项重要任务。

¹⁶该图说明了应用于EHR数据以产生可能适用于研究用途的数据集的一些过程（即，从原始源数据到最终分析数据集的步骤）。该图显示了EHR数据的过程；索赔数据的过程可能有所不同。本节描述了每个工艺步骤的质量检查。

图1:EHR数据生命周期的说明性示例¹⁶



评价EHR和医疗索赔数据质量的指南主要关注分布式数据网络，在该网络中，对不同的数据源进行汇总、链接和处理，以创建综合数据仓库（Miksad和Abernethy 2018；Girman et al.2018；Daniel et al.2018；Kahn et al.2016；Wang et al.2017；Mahendraratnam et al.2019）。尽管FDA不认可任何特定的指南或检查表集，但研究人员应评估数据的完整性、准确性和合理性，包括根据其原始来源（例如出院记录、病理报告、登记记录）验证数据，并在适用的情况下符合基于共识的数据标准。研究人员应提供选择这些标准的科学依据，并应阐明这些标准如何足以确保相关数据源的完整性、准确性和合理性。

研究方案和分析计划应详细说明数据来源（整个数据生命周期中使用的管理和转换程序），并描述这些程序如何影响数据完整性和研究的总体有效性。以下是在数据生命周期的每个步骤检查数据时需要考虑的要点，包括(A)关于数据值的完整性、一致性和合理性的数据表征，(B)记录包括转换过程的QA/QC计划；(C)定义一套确保数据完整性的程序。

A. 数据特征

EHR和医疗索赔数据的格式和来源在医疗保健实体（例如，保险公司，实践，提供者，数据供应商）之间可能存在很大差异。通常，研究人员应解决用于确保数据完整性和准确性的程序，以及在数据生命周期内进行数据应计，管理和转换的过程。FDA建议自动化数据质量报告，在适用于所选数据源时，以标准化方式包括以下特征和过程：

- 数据累积
 1. 数据检索和处理方法，以尽量减少缺失数据提取、不可信值和在治疗点（例如，在手动或自动医疗保健数据收集过程的临床实践期间）采集的数据中的数据质量检查，以确保核心数据元素的准确性和完整性。
 2. 核心数据要素的来源，以便能够将这些要素追溯到各自的来源点，并明确记录可能发生的修改。
 3. 数据可用性的及时性、跨越的数据年数和覆盖范围的连续性（例如，患者入组的中位持续时间）。
 4. 处理数据差异和重复记录。RWD可能源于跨不同设置和平台的多个数据流，这可能会导致相同变量的数据差异（例如，当相同变量的信息元素在不同的数据源中以不同的方式输入），甚至在同一数据源中为同一患者输入重复的记录。
 5. 数据持有人在相关数据收集期间实施数据错误纠正的原因和时间。
 6. 在数据收集的相关期间，数据持有人实施的可能影响数据累积和/或数据质量检查的流程变更的原因和时间。
 7. 在整个研究期间，与关注变量相关的编码实践和版本（例如，国际疾病分类[ICD]诊断代码，医疗保健通用程序编码系统代码）的任何更新或变更。
 8. 研究期间数据的任何其他变更（例如，收集、报告、定义）及其对研究结果的潜在影响。

- 数据管理

1. 不同来源的数据随时间的例行迁移。
2. 各研究中心采用的质量保证(QA)测试和数据质量检查，以及用于确定数据质量技术是否适用于数据的预期目的的标准。
3. 明确定义的核心数据要素，具有一致和已知的临床意义，并了解数据来源，以及所用临床定义的文件。
4. 评估数据要素的完整性和随时间变化的趋势。
5. 非结构化和结构化数据处理（例如，非结构化数据的抽象和转换为结构化数据），包括手动与自动技术。
6. 跨系统的结构化数据的协调。
7. 适用时，遵守公开的，基于共识的数据管理标准。
8. 映射的准确性（例如，在存在不同编码系统的情况下，例如医学的系统化命名法-临床术语[SNOMED CT]与ICD-10-CM）。
9. 其他协调和映射注意事项（如适用）（如果数据跨越多个国家，例如，除美国数据外，还使用英国数据）。

- 数据转换

1. 作为数据仓库创建的一部分，实现应用于整个存储库总体的提取、转换和加载过程。
2. 对患者记录进行去识别，并能够在不丢失可追溯性的情况下重新识别原始源数据中的唯一患者。
3. 用于转换和清理数据的算法，以及标准操作程序的可用性，包括验证数据的程序。
4. 数据标准化（例如，数据类型、大小、格式），用于数据元素和语义的内部一致性，包括本地代码与目标术语的语义（例如，实验室数据）。

5. 当将多个数据源转换为CDM时，用于将数据转换为CDM的过程（例如，通用术语和结构），CDM的全面性（例如，CDM是否包含关键数据元素），用于识别和处理数据源内和数据源间重复记录的方法（例如，算法/方法），以及仅限于CDM对样本量和患者随访持续时间或药物暴露持续时间的潜在影响。见第IV.B.3节，分布式数据网络。
 6. 实施与数据模型一致性错误有关的数据检查。
 7. 用于准备数据链接的数据转换过程。见第IV.B.2节，数据链接和综合。
 8. 记录链接（即链接来自多个数据集的记录）和去重（即在数据集中找到重复的记录）过程的质量，其质量可能会有所不同，具体取决于用于执行匹配的数据的准确性和链接算法的准确性。
 9. 可能导致研究结果偏倚的错误量化（例如，错误匹配，错过匹配）。这些在评估连锁质量时很重要（Harron et al.2017）。重要的是要报告链接算法的详细信息和适当的指标（例如，链接错误率，匹配率，链接和未链接数据的特征比较）。其他考虑因素包括错误是随机的还是非随机的，潜在的偏倚以及对风险估计和研究结果的影响。
 10. 裁定链接数据中差异的程序以及处理链接差异的计划（例如，调整链接错误的风险估计）。
- 研究特定分析数据集
 1. 编写分析数据集时，遵守研究方案和统计分析计划中概述的数据规范。
 2. 其他研究特定的数据转换，例如仅针对感兴趣的患者子集进行的数据转换，而不适用于数据仓库中的所有患者记录（例如，从非结构化文本病理报告中手动提取数据）。
 3. 对最终分析数据集实施数据检查，以确定数据元素（例如身高、体重、血压）的不合理值，如何处理这些值，以及关键分析变量的数据完整性。
 4. 缺失和不可信数据的程度、百分比和模式。根据分析计划

提出的处理缺失数据的方法，可以进行插补，并将其包括在最终分析数据集中以及所描述的插补类型中。

B. QA/QC计划文件

应在研究方案和分析计划中描述用于构建分析数据的QA/QC计划、分析过程中处理质量控制问题的计划方法以及根据数据要素考虑不同水平的数据质量（以及对研究结果的潜在影响）。一般来说，在研究设计期间开展数据相关活动之前确保数据质量的活动，这些活动，包括如何收集数据的标准化程序，可被视为QA（Szklo和Nieto 2006）。质量控制包括从数据收集到最终分析数据集汇编所采取的决策和步骤，以确保其符合预先规定的标准并确保所用工艺可重现。包括临床输入的多学科方法对于确保充分捕获和处理数据是必要的，特别是对于电子医疗保健系统而言，电子医疗保健系统固有地包含了医疗保健交付的细微差别和复杂性。

C. 数据管理过程文件

从数据收集到撰写最终研究报告，应对所有手动和自动数据检索和转换过程进行全面评估，以确保数据的完整性。研究人员应确保管理和转换过程不会改变数据的含义或导致重要上下文信息的丢失。过程描述应包括保障措施或检查，以确保患者数据不会被重复或过度代表。此外，用于挖掘和评估非结构化数据的过程文件应描述用于提取非结构化数据（例如，临床医生注释）和补充结构化数据（例如，诊断代码）的技术（例如，自然语言处理）。

应在研究方案或分析计划中描述用于管理和准备最终研究特定分析数据集的过程。分析人员应接受适当的培训或具有用于编译分析数据集的数据和软件的经验。为了便于FDA审查，所有提交的程序（例如，由分析人员编写的程序）都应该用注释进行彻底注释，以描述程序中编写的每个数据管理和分析步骤的意图或目的（例如，在统计分析程序中注释每个数据步骤）。

VII. 术语表

准确度 (Accuracy): 测量值与预期测量值的真实值之间的接近程度。

17

人工智能 (Artificial Intelligence, AI): 制造智能机器，特别是智能计算机程序的科学和工程 (McCarthy 2007)。

通用数据模型 (Common Data Model, CDM): 将各种电子医疗保健数据源标准化为通用格式，以确保所有提供数据的研究中心之间的互操作性¹⁸。

完整性 (Completeness): “存在必要的数据” (美国国立卫生研究院协作2014)。

可计算表型 (Computable Phenotype): 可以通过使用一组定义的数据元素和逻辑表达式，通过对EHR系统或临床数据存储库 (包括疾病注册，索赔数据) 进行计算机查询来确定的临床状况或特征。可计算的表型定义提供了识别具有关注条件的患者人群的规范¹⁹。

概念性定义 (Conceptual Definition): 用一般或定性术语解释研究结构 (例如，暴露、结局、协变量) 或特征。

伴随用药 (Concomitant Medication): 与目标产品或比较剂同时使用的处方药或非处方药或补品。

一致性 (Conformance): “数据与标准化类型、尺寸和格式一致” (Daniel et al.2018)。

混杂 (Confounder Factor, 混杂因素): 在分析中适当调整后可用于减少混杂偏倚的变量。混杂是由于暴露与影响结果发生的其他因素相关联而导致的暴露对结果影响的测量的失真。当暴露与结局之间的全部或部分明显关联实际上由影响结局且本身不受暴露影响的其他变量解释时，就会发生混杂 (Porta 2014)。

17改编自计量指南联合委员会国际计量词汇-基本和一般概念及相关术语，2012年第3版。

18改编自Sentinel系统原则和政策 (2019年7月)，网址为

<https://www.sentinelinitiative.org/sites/default/files/about/Sentinel-system-principles-and-policies.pdf>

19见NIH合作实用临床试验生活教科书“基于电子健康记录的表型分析”章节，网址为

<https://rethinkingclinicaltrials.org/resources/ehr-phenotyping/>。

覆盖范围的连续性 (Continuity of Coverage):患者被纳入医疗保健系统的时间段,在此期间任何医疗服务或药物处方都将被记录在该医疗保健系统的电子记录系统中。²⁰

协变量 (Covariate):既不是关注的暴露也不是结局的变量,但测量的目的是描述人群,或者因为它可能是研究设计或分析中要考虑的混杂因素或效应修饰因素。

累积剂量 (Cumulative Dose):在特定时间内给予患者的目标药物(暴露量)的总量。²¹

数据累积 (Data Accrual):收集数据的过程。

数据管理 (Data Curation):对源数据应用标准(如Health Level 7、ICD-10-CM);例如,将代码应用于EHR中的不良事件、疾病分期、疾病进展和其他医学和临床概念。

数据元素 (Data Element):数据字段中对应于一名患者的数据(摘自Daniel, et al.2018)。

数据完整性 (Data Integrity):数据的完整性、一致性和准确性。²²

数据存储库 (Data Repository):整合来自不同临床来源的数据的数据库,例如EHR系统中的数据,以提供患者接受的护理的更广泛的情况²³。

数据转换 (Data Transformation):包括数据提取、清理和集成(例如,到CDM中)。

数据仓库 (Data Warehouse):由经过数据转换和去识别的数据存储库中的数据组成。

²⁰参见FDA行业指南《使用电子医疗保健数据进行和报告药物流行病学安全性研究的最佳实践》(2013年5月)。

²¹改编自“NCI癌症术语词典”,网址为<https://www.Cancer.gov/publications/dictionaries/cancer-terms/def/cumulative-dose>。

²²参见FDA行业指南数据完整性和药品CGMP的合规性问答(2018年12月)。

²³改编自Shortliffe, EH和JJ Cimino, 2014,《生物医学信息学:医疗保健和生物医学中的计算机应用》,第4版,纽约(NY):Springer。

去识别化 (De-Identification) :从个人健康信息中删除个人标识符的过程。²⁴

分布式数据网络 (Distributed Data Network) :由多个分散的健康护理数据站点组成的网络, 提供查询或分析来自任何或所有站点的数据的能力。

效应修饰因子 (Effect Modifier) :生物学、临床、社会学或以其他方式改变研究中另一因子效应的因子 (Porta 2014) 。

电子医疗保健数据 (Electronic Health Care Data) :分析数据, 是从计算机或其他电子技术平台获得的自动化健康数据的有组织的集合。²⁵

电子健康记录 (Electronic Health Record, EHR) :EHR系统中包含的个体患者记录。典型的个体EHR可能包括患者的病史, 诊断, 治疗计划, 免疫接种日期, 过敏, 放射学图像, 药房记录以及实验室和测试结果。²⁶

医疗索赔数据 (Medical Claims Data) :医疗服务提供者提交给保险公司以接受治疗和其他干预措施付款的医疗索赔信息汇编。医疗索赔数据使用标准化的医疗代码, 例如世界卫生组织的国际疾病分类编码 (ICD-CM) 诊断代码, 以识别诊断和治疗。²⁷

错误分类 (Misclassification) :将个体、值或属性错误分类为应当分配的类别以外的类别 (Porta 2014) 。

缺失数据 (Missing Data) :本将用于研究分析, 但未观察到、未收集或无法获得的数据。这是指预期收集但不存在的信息和预期不收集并因此不存在的信息。

阴性预测值 (Negative Predictive Value, NPV) :分类结果为阴性时受试者未患疾病的概率。

操作定义 (Operational Definition) :研究人员在特定研究中测量结构所遵循的特定于数据的操作或程序。

²⁴参见卫生和公共服务部关于根据健康保险携带和责任法案(HIPAA)隐私规则对受保护的健康信息进行去识别的方法的指南, 可在 https://www.hhs.gov/sites/default/files/ocr/Privacy/HIPAA/understanding/coveredentities/de-identification/hhs_deid_guidance.pdf上获取。

²⁵改编自Hartzema, A, HH Tilson和KA Chan, 2008, 药物流行病学和治疗风险管理, 辛辛那提(OH):Harvey Whitney书籍。

²⁶参见FDA行业指南《临床研究中电子健康记录数据的使用》(2018年7月)

²⁷见FDA真实世界证据计划框架(2018年12月)

合理性 (Plausibility) :数据值的可信度或真实性 (Kahn et al.2016) 。

阳性预测值 (Positive Predictive Value, PPV) :分类结果为阳性时, 受试者发生疾病的概率。

来源 (Provenance) :“说明数据 (在数据库、文件或存储库中) 的来源以及如何以及为什么到达当前位置的解释”的审计跟踪。²⁸

敏感性 (Sensitivity) :当受试者患有疾病时, 分类结果为阳性的概率。

源数据 (Source Data) :临床研究中重建和评价研究所需的临床发现、观察结果或其他活动的原始记录和原始记录认证副本中的所有信息。源数据包含在源文件 (原始记录或认证副本) 中。²⁹

特异性 (Specificity) :在受试者没有疾病的情况下, 分类结果为阴性的概率。

研究周期 (Study Period) :用于研究的数据日历时间范围 (Wang et al.2017) 。

可溯源性 (Traceability) :允许理解分析结果 (研究报告中的表、列表和图)、分析数据集、列表数据集和源数据之间的关系。³⁰

验证 (Validation) :通常根据参考标准, 确定方法合理或数据正确测量的过程。³¹

28数据库系统百科全书数据来源的定义, 请访问
https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-39940-9_1305。

29参见FDA行业指南《临床研究中电子健康记录数据的使用》(2018年7月)。

30见FDA技术规范文件研究数据技术一致性指南 (2019年10月)。

31改编自Porta, M, 2014, 流行病学词典, 第六版, 纽约(NY):牛津大学出版社

I. 参考文献

- Bustamante, R, A Earles, JD Murphy, AK Bryant, OV Patterson, AJ Gawron, T Kaltenbach, MA Whooley, DA Fisher, SD Saini, S Gupta, and L Liu, 2019, Ascertainment of Aspirin Exposure Using Structured and Unstructured Large-scale Electronic Health Record Data, *Med Care*, 57:e60–e64.
- Curtis, M, SD Griffith, M Tucker, MD Taylor, WB Capra, G Carrigan, B Holzman, AZ Torres, P You, B Arneri, and AP Abernethy, 2018, Development and Validation of a High-Quality Composite Real-World Mortality Endpoint, *Health Services Research*, 53(6)Part I:4460-4476.
- Carreras, G, M Simonetti, C Cricelli, and F Lapi, 2018, Deterministic and Probabilistic Record Linkage: an Application to Primary Care Data, *J Med Sys*, 42(5):82.
- Daniel, G, C Silcox, J Bryan, M McClellan, M Romine, and K Frank, 2018, Characterizing RWD Quality and Relevancy for Regulatory Purposes, Duke Margolis Center for Health Policy, accessed January 9, 2019, https://healthpolicy.duke.edu/sites/default/files/atoms/files/characterizing_rwd.pdf.
- Girman, CJ, ME Ritchey, W Zhou, and NA Dreyer, 2019, Considerations in Characterizing Real-World Data Relevance and Quality for Regulatory Purposes: A Commentary, *Pharmacoepidemiol Drug Saf*, 28(4):439–442.
- Harron, KL, JC Doidge, HE Knight, RE Gilbert, H Goldstein, DA Cromwell, and JH van der Meulen, 2017, A Guide to Evaluating Linkage Quality for the Analysis of Linked Data, *Int J Epidemiol*, 46(5):1699–1710.
- Haynes, K, 2019, Mortality: The Final Outcome, *Pharmacoepidemiol Drug Saf*, epub ahead of print Jan 31, 2019, doi: 10.1002/pds.4715.
- Kahn, MG, TJ Callahan, J Barnard, AE Bauck, J Brown, BN Davidson, H Estiri, C Goerg, E Holve, SG Johnson, ST Liaw, M Hamilton-Lopez, D Meeker, TC Ong, P Ryan, N Shang, NG Weiskopf, C Weng, MN Zozus, and L Schilling, 2016, A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, *EGEMS*, 4(1):1244.
- Lash, TL, MP Fox, and AK Fink, 2009, *Applying Quantitative Bias Analysis to Epidemiologic Data*, New York (NY): Springer.
- Mahendraratnam, N, C Silcox, K Mercon, A Kroetsch, M Romine, N Harrison, A Aten, R Sherman, G Daniel and M McClellan, 2019, Determining Real-World

Data's Fitness for Use and the Role of Reliability, Duke Margolis Center for Health Policy, accessed July 24, 2020
https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd_reliability.pdf.

McCarthy, J, 2007, What Is Artificial Intelligence?, John McCarthy's Home Page, updated November 12, 2007, <http://www-formal.stanford.edu/jmc/index.html>.

Miksad, RA, and AP Abernethy, 2018, Harnessing the Power of Real-World Evidence (RWE): A Checklist to Ensure Regulatory-Grade Data Quality, *Clin Pharmacol Ther*, 103(2):202–205.

National Institutes of Health Collaboratory, 2014, Assessing Data Quality for Healthcare Systems Data Used in Clinical Research, accessed August 27, 2019,
https://dcricollab.dcri.duke.edu/sites/NIHKKR/KR/Assessing-data-quality_V1%200.pdf#search=Assessing%20data%20quality.

Navar, AM, ED Peterson, DL Steen, DM Wojdyla, RJ Sanchez, I Khan, X Song, ME Gold, and MJ Pencina, 2019, Evaluation of Mortality Data from the Social Security Administration Death Master File for Clinical Research, *JAMA Cardiol*, epub ahead of print Mar 6, 2019, doi: 10.1001/jamacardio.2019.0198.

Porta, M, 2014, *A Dictionary of Epidemiology*, Sixth Edition, New York (NY): Oxford University Press.

Richesson, RL, MM Smerek, and CC Blake, 2016, A Framework to Support the Sharing and Reuse of Computable Phenotype Definitions Across Health Care Delivery and Clinical Research Applications, *EGEMS*, 4(3):1232.

Szklo, M, and FJ Nieto, 2006, *Epidemiology: Beyond the Basics*, 2nd Edition, Burlington (MA): Jones & Bartlett Learning.

Velentgas, P, NA Dreyer, P Nourjah, SR Smith, and MM Torchia, editors, 2013, *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*, AHRQ Publication No. 12(13)–EHC099, accessed January 9, 2019,
https://effectivehealthcare.ahrq.gov/sites/default/files/related_files/user-guide-observational-cer-130113.pdf.

Wang, SV, S Schneeweiss, ML Berger, J Brown, F de Vries, I Douglas, JJ Gagne, R Gini, O Klungel, CD Mullins, MD Nguyen, JA Rassen, L Smeeth, and M Sturkenboom, 2017, Reporting to Improve Reproducibility and Facilitate Validity Assessment in Healthcare Database Studies V1.0, *Pharmacoepidemiol Drug Saf*, 26(9):1018–1032.