**真实世界证据支持药物研发的基本考虑**

**（征求意见稿）**

**国家药品监督管理局药品审评中心**

**2019年5月**

**目 录**

[**一、** **引言** 1](#_Toc8307579)

[1. 背景与目的 1](#_Toc8307580)

[2. 国内外监管机构在法规或指南制定方面的进展 2](#_Toc8307581)

[**二、** **真实世界研究的相关定义** 4](#_Toc8307582)

[1. 真实世界数据 5](#_Toc8307583)

[2. 真实世界证据 9](#_Toc8307584)

[**三、** **真实世界证据支持药物研发和监管决策的几种情形** 9](#_Toc8307585)

[1. 罕见病治疗药物 9](#_Toc8307586)

[2. 修订适应症或联合用药范围 9](#_Toc8307587)

[3. 上市后药物的再评价 11](#_Toc8307588)

[4. 中药医院制剂的临床研发 11](#_Toc8307589)

[5. 指导临床研究设计 13](#_Toc8307590)

[6. 精准定位目标人群 14](#_Toc8307591)

[**四、** **真实世界研究的基本设计** 14](#_Toc8307592)

[1. 实用临床试验 14](#_Toc8307593)

[2. 使用真实世界数据作为对照的单臂试验 15](#_Toc8307594)

[3. 观察性研究 16](#_Toc8307595)

[**五、** **真实世界证据的评价** 16](#_Toc8307596)

[1. 真实世界证据和其所支持的科学问题 16](#_Toc8307597)

[2. 如何从真实世界数据到真实世界证据 17](#_Toc8307598)

[**参考文献** 18](#_Toc8307599)

[**附录1：词汇表** 20](#_Toc8307600)

[**附录2：真实世界研究常用统计分析方法** 24](#_Toc8307601)

[**附录3：中英文词汇对照** 40](#_Toc8307602)

**真实世界证据支持****药物研发的基本考虑**

1. **引言**
2. 背景与目的

随机对照临床试验(Randomized Controlled Trial，RCT)被认为是评价药物有效性的“金标准”，并为药物临床试验普遍采用。RCT严格控制试验入组与排除标准和其它条件，并进行随机化分组，因此能够最大限度地减少影响因果推断的因素，使得研究结论较为确定，所形成的证据可靠性也较高。但RCT亦有其局限性：严苛的入排标准可能会使试验人群对目标人群的代表性变差；所采用的标准干预与临床实践用药不完全一致；有限的样本量和较短的随访时间导致对罕见不良事件探测不足等。这些局限性使得RCT的研究结论外推于实际临床应用时面临挑战。此外，对于某些缺乏有效治疗措施的罕见病和危及生命的重大疾病，常规RCT或难以实施，或需高昂的时间成本，或可能引发伦理问题。因此，在药物研发领域如何利用真实世界证据（Real World Evidence，RWE），或者将其作为RCT的辅助证据，用以评价药物的有效性和安全性，已成为全球相关监管机构、制药工业界和学术界共同关注且极具挑战性的问题。

首先，我们需要从概念上厘清真实世界证据的定义和内涵。

其次，作为真实世界证据基础的真实世界数据（Real World Data，RWD），其能否、或如何起到充分的支撑作用，涉及诸多亟待商榷的问题，包括数据来源、数据标准、数据质量、数据共享、数据的基础建设等。

其三，监管法规的真空。目前国际上尚无成熟的相关法规出台，在没有成熟经验的情况下，如何制定适合我国制药行业现实的指南需要积极的探索研究和创新。

其四，评价真实世界证据的方法学有待规范。真实世界证据源于对真实世界数据的正确和充分分析，所采用的分析方法主要是因果推断方法，涉及较复杂的模型与模型假设、相应的协变量筛选、混杂因素识别、中间变量及工具变量定义等，这对统计分析人员提出了更高的要求，也对法规的制定提出了迫切需求。

其五，真实世界证据的适用范围有待明确。真实世界证据的主要作用是与传统临床试验提供的证据互为补充，综合形成完整而严谨的证据链，提高药物研发的科学性和效率，而非替代之。因此，需要根据药物研发的现实情况明确真实世界证据的适用范围，并能够随现实情况变化进行调整。

鉴于上述，本指南旨在厘清药物研发中真实世界研究的相关定义，明确真实世界证据在药物研发中的地位和适用范围，探究真实世界证据的评价原则，以为工业界利用真实世界证据支持药物研发提供科学可行的指导意见。

1. 国内外监管机构在法规或指南制定方面的进展

2009年2月美国的经济复苏刺激法案（The American Recovery and Reinvestment Act）对实效比较研究（Comparative Effectiveness Research，CER，见词汇表）起到了巨大推动作用。基于CER的真实世界环境的背景，真实世界研究（Real World Research/Study，RWR/RWS）的概念被提出。

美国于2016年12月通过《21世纪治愈法案》（21st Century Cures Act），旨在鼓励美国食品药品监督管理局（the Food and Drug Administration，FDA）开展研究并使用真实世界证据以支持药物和其它医疗产品的监管决策，加快医药产品的开发。在该法案的推动下，2017-2018年FDA先后发布了《使用真实世界证据以支持医疗器械监管决策》、《临床研究中使用电子健康档案数据指南》和《真实世界证据计划的框架》。

欧盟药物管理局（[European Medicines Agency](https://www.baidu.com/link?url=fo7gjGGXO9sqmga69KE5ucZtsqLVYz0geo9RyhsoroulVRj7XH1dEHCwKlBD_nm3&wd=&eqid=c37e94d80006bf9a000000065c299f60)，EMA）在2013年发布了《阿尔茨海默病疾病进展和临床试验评估的数据驱动模型新方法的意见书》，讨论了利用真实世界中观察性数据建立疾病进展模型的技术细节。EMA于2014年启动了适应性许可试点项目（Adaptive Licensing Pilot），用以评估利用观察性研究数据辅助决策的可行性。2016年11月发布了《药物上市后有效性评价科学指南》。

日本医药品医疗器械综合机构（PMDA）在人用药品技术要求国际协调理事会（International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, ICH）层面提出更高效利用真实世界数据开展上市后药物流行病学研究技术要求新议题的国际协调。

我国系统性开展使用真实世界证据支持药物研发和监管决策的工作尚处于起步阶段。国家药监机构在审评实践中已经开始应用真实世界证据，例如2018年批准的扩展贝伐珠单抗（Bevacizumab）联合以铂类为基础的化疗方案，采用三项回顾性研究结果作为真实世界证据支持最终决策。另一个案例是某药上市后开展前瞻性、观察性真实世界研究，以提供更充足的有效性和安全性证据。

1. **真实世界研究的相关定义**

广义的真实世界研究既包括以自然人群为对象的研究，也包括以临床人群为对象的研究；后者所得到的真实世界证据既可用于支持医疗产品研发与监管决策，也可用于其它科学目的。本指南仅限于用于支持医疗产品研发与监管决策的真实世界研究（见下图）。

图1从RWD到支持医疗产品监管决策RWE的路径

我们将真实世界研究定义为：在真实世界环境下收集与患者有关的数据（真实世界数据），通过分析，获得医疗产品的使用价值及潜在获益或风险的临床证据（真实世界证据），其主要研究类型是观察性研究，也可以是实用临床试验。

1. 真实世界数据
2. 定义

美国联邦食品、药品和化妆品法案（FD&C Act）第505F(b)条将真实世界数据定义为“来源于传统临床试验以外的，关于药物使用或潜在获益、风险的数据”。FDA在《真实世界证据方案的框架》和《使用真实世界证据以支持医疗器械监管决策》中将真实世界数据定义为“与患者健康状况有关的和/或日常医疗过程中收集的各种来源的数据”。例如，电子健康档案（Electronic Health Record， EHR）数据、电子病例（Electronic Medical Record，EMR）数据、医保数据（Medical Claims Data）、产品和疾病登记中心的数据、患者报告数据（包括居家环境）、其它健康检测（如移动设备）的数据等。

我们将真实世界数据定义为：与患者使用药物以及健康状况有关的和/或来源于各种日常医疗过程所收集的数据。

1. 真实世界数据的来源

国内真实世界数据的常见来源包括：

1）卫生信息系统（Health Information System，HIS）：类似于EMR/HER，包括结构化和非结构化数据字段的数字化患者记录，如患者的人口学特征、临床特征、诊断、治疗、实验室检查、安全性和临床结局等。

2）医保系统：包含有关患者基本信息、医疗服务利用、处方、结算、医疗索赔和计划保健等结构化字段的数据。

3）疾病登记系统：特定疾病（通常是慢性病）患者的数据库，通常来源于医院的疾病人群队列登记。

4）国家药品不良反应监测哨点联盟（CASSA）：利用医疗机构电子数据建立药品及医疗器械安全性的主动监测与评价系统。

5）自然人群队列数据库：国内已经建立或正在建立的自然人群队列和专病队列数据库，可成为潜在的真实世界数据。

6）组学相关数据库：采集患者的生理学、生物学、健康、行为和可能的环境相互作用的组学相关信息，如药物基因组学、代谢组学和蛋白质组学的数据库。

7）死亡登记数据库：由医院、疾病预防控制中心，和户籍部门联合确认的死亡登记所形成的数据库。

8）来自移动设备端的数据：应用移动设备，如可穿戴设备，检测受试者获得的相关数据。

9）其他特殊数据源：为特殊目的创建的数据库，如国家免疫规划数据库等。

1. 数据质量评价

真实世界数据的质量主要通过其相关性和可靠性进行评估。

1）相关性：评估真实世界数据是否适合监管用途的重要相关因素包括但不限于：

①是否包含与临床结局相关的重要变量和信息，如药物使用、患者人口学和临床特征、协变量、结局变量、随访时间、样本量等；

②临床结局定义是否准确，相应的临床意义是否明确；

③目标人群定义是否准确、具有代表性；

④研究方案和统计分析计划是否满足对研究假设的验证。

2）可靠性：真实世界数据的可靠性主要从数据的完整性（Completeness）、准确性（Accuracy）、质量保证（Quality Assurance）和质量控制（Quality Control）方面进行评价。

①完整性：真实世界数据无法避免数据缺失问题，但缺失比例应该有一定的限度。对于不同研究，数据的缺失程度可能会有不同，当特定项目的数据缺失比例超过一定限度时，会使研究结论有很大的不确定性，此时，需要慎重考虑该数据能否作为支持产生真实世界证据的数据。

②准确性：数据的准确性极为重要，需要依据较权威的参照来源进行识别和验证。例如，血压的测量需要使用校准过的血压计，且测量过程需遵循操作规范；终点事件是否经独立的终点事件委员会做出判断等等。

③质量保证：质量保证是指预防、探测和纠正研究过程中出现的数据错误或问题的措施。质量保证与监管合规性（Regulatory Compliance）密切相关，应贯穿于数据管理的每一个环节，作为最基本的要求，数据管理的每一环节必须制定相应的标准操作规程（Standard Operation Procedures, SOPs）。

④质量控制：数据的采集、修改、传输、保存和归档，以及数据的处理、分析和提交等每一个环节均需进行质量控制，以保证真实世界数据的准确性和可靠性。对此有必要制定一个完整、规范、可靠的数据管理流程或方案。

1. 数据标准

数据标准使递交的资料具有可预测性和一致性，并具有信息技术系统或科学工具可使用的形式。为了使多个来源的真实世界数据协同工作，需将数据转换成通用格式，具有通用表述形式（如术语、词汇表、编码方案等）。

此外，真实世界数据的质量是否可以支持药物研发需考虑的重要因素还包括（但不限于）：数据采集是否有明确流程和合格人员；是否使用了共同的定义框架，即数据字典；是否遵守采集关键数据点的共同时间框架；是否建立与收集真实世界数据有关的研究计划、协议和/或分析计划的时间安排；用于数据元素捕获的技术方法是否充分，包括各种来源数据的集成、药物使用的数据记录、与索赔数据的链接等；患者的选择是否将偏倚最小化以体现真正的目标人群；数据输入、传输是否具有可用性和及时性；是否有充分和必要的患者保护措施，如患者隐私保护和符合监管规定的知情同意。

1. 真实世界证据

真实世界证据是通过对真实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益或风险的临床证据。该定义在概念上不限于通过回顾性观察研究获得证据，还允许前瞻性地获取更广泛的数据以形成证据，特别是包括实用临床试验（Pragmatic Clinical Trial, PCT）一类的研究设计。

1. **真实世界证据支持药物研发和监管决策的几种情形**

真实世界证据可以经多种形式支持药物研发，涵盖上市前临床研发以及上市后再评价等多个环节。任何以药品注册为目的的对于真实世界证据的使用都需要提前和监管部门进行充分的沟通交流，以确保双方对于研究目标和方法学的认识达成一致。

1. 罕见病治疗药物

罕见病治疗药物临床试验除病例稀少、招募困难外，最大的挑战是对照的选择，因为罕见病通常没有或很少有可选治疗。因此，以自然疾病队列形成的真实世界数据就可以作为外部对照。

外部对照主要用于非随机单臂试验，可以是历史的也可以是平行的。历史外部对照以早先获得的真实世界数据作为对照；平行外部对照则是将与单臂试验同期开展的疾病登记数据作为对照。采用外部对照需注意考虑目标人群的异质性和可比性对真实世界证据的影响。

1. 修订适应症或联合用药范围

对于已经上市的药物，经过长期临床实践可能会发现有必要扩大适应症，通常情况下采用RCT支持扩大适应症。但当RCT不可行或生成的证据非最优时，PCT则是一个很好的选择。例如，临床实践发现一种用于治疗糖尿病的新药可能会使心血管疾病（如心力衰竭）患者潜在获益，对此若采用RCT设计，会使研究对象招募极为困难，甚至面临伦理问题，而采用PCT设计可能更具有可行性。

在儿童用药领域，国内临床实践中常有超说明书用药的情况，利用真实世界证据支持适应症人群的扩大也是药物研发的一种策略。

利用真实世界证据支持扩大联合用药的一个典型案例是贝伐珠单抗（Bevacizumab），该药是一种血管内皮生长因子（VEGF）人源化单克隆抗体制剂，于2015年在中国获批联合化疗（卡铂与紫杉醇）用于不可切除的晚期、转移性或复发性非鳞状非小细胞肺癌患者的一线治疗。而真实世界中患者所联合的化疗方案并不局限于卡铂与紫杉醇，还包括培美曲塞联合铂类、吉西他滨联合顺铂等。2018年10月该药获批将治疗方案扩展为联合以铂类为基础的化疗方案，其中三项真实世界研究结果提供了强有力的支持证据。这三项研究回顾性分析了三家医院的患者数据，均显示在含铂双药化疗基础上联合贝伐珠单抗较单纯化疗显著延长PFS和OS，与全球人群数据具有一致性，并且未发现新的安全性问题。此外，相关真实世界研究还提供了EGFR突变和脑转移等不同患者亚组中的疗效数据，从多角度证实了贝伐珠单抗联合疗法的有效性和安全性。

1. 上市后药物的再评价

基于RCT证据获批的药物，通常由于病例数较少、研究时间较短、试验对象入组条件严格、干预标准化等原因，存在安全性信息有限、疗效结论外推不确定、用药方案未必最优、经济学效益缺乏等不足，需要利用真实世界数据对药物在自然人群中的有效性、安全性、用药方案，以及[经济学](https://wiki.mbalib.com/wiki/%E7%BB%8F%E6%B5%8E%E5%AD%A6)效益等方面进行更全面的评估，并不断根据真实世界证据做出决策调整。

例如，某药是目前已经在全球50多个国家/地区获得批准的心血管药物，对其开展的国际多中心临床试验中，由于中国亚组人群样本量较小、心血管事件发生数有限，并且药物暴露时间较短，导致中国人群的疗效结果变异性较大。作为境外已上市且临床急需药物，为进一步明确该药在中国患者中的疗效，申请人计划上市后开展一项前瞻性、观察性、真实世界研究，以评价该药联合标准治疗对比单独标准治疗对患有心血管疾病的中国患者主要心血管事件的预防效果。

1. 中药医院制剂的临床研发

在临床上被广泛长期使用却未获批准上市的中药医院制剂是我国特有的现象。对于此类药物的临床研发，若能将真实世界研究与随机对照临床试验相结合，将为中药医院制剂探索出科学可行的临床研发路径和监管决策依据。

探索应用真实世界证据支持中药医院制剂的研发策略有多种，图2和图3是可能路径中的两种。观察性研究与RCT研究相结合的路径如图2所示，第一阶段先开展回顾性观察性研究，此阶段应尽可能地收集既往的与使用该药品有关的真实世界数据，包括所有可能的协变量；制定数据清理规则；选择可能的对照；对数据质量进行评估；采用恰当的统计方法进行全面详细的分析。如果通过回顾性观察性研究得出该药品在临床应用中对患者具有潜在获益，可以进入下一研究阶段，否则研究终止。第二阶段开展前瞻性观察性研究。由于有了第一阶段的研究基础，该阶段可以将前瞻性观察性研究设计的更加周密，包括数据的采集及其系统、数据的质量控制、数据清理的规则、明确定义对照等。在前瞻性观察性研究进展到某一时期，如果数据分析结果与回顾性观察性研究结果一致，且继续显现出该药品在临床应用中对患者具有明显获益，可适时平行开展第三阶段的RCT研究。RCT研究可以先进行探索性RCT研究，但如果前期的观察性研究证据较充分，也可以直接进行确证性RCT研究。从时间上看，RCT研究的周期被前瞻性观察性研究所覆盖，后者可以与RCT研究同时结束，也可在RCT研究结束后继续延展一段时间，应视真实世界证据是否充分而定。



图2 中药医院制剂研发的路径策略之一

观察性研究与PCT研究相结合的路径如图3所示，第一阶段先开展回顾性观察性研究，如果得出该药品在临床应用中对患者具有潜在获益，可以进入下一研究阶段，否则研究终止。第二阶段开展PCT研究， PCT研究提供的证据可以用于支持其临床有效性和安全性的评价。



图3 中药医院制剂研发的路径策略之二

1. 指导临床研究设计

利用真实世界证据指导临床研究设计有着更现实的用途。例如，在上一节列举的两种中药医院制剂研发的路径，都采用了回顾性观察性研究所产生的真实世界证据，包括疾病的自然史、疾病在目标人群的流行率、标准化治疗的疗效和有效性（Effectiveness）、以及与疗效和有效性有关的关键协变量在目标人群中的分布和变化等等，为下一阶段的研究设计提供了依据。更为普遍的，真实世界证据可为入选和排除标准、样本量估计的参数、非劣效界值的确定等提供有效的参考依据。

1. 精准定位目标人群

精准医疗旨在更好地预测药物对特定人群（亚组）的治疗获益和风险，基于真实世界数据的真实世界证据为精准医疗提供了这种可能。例如，传统临床试验因样本量有限，往往在研究计划中忽略或无暇顾及亚组效应，使得潜在的治疗应答者或具有严重副作用的高风险人群的重要信息不能充分体现，从而导致目标人群失准。通过对真实世界数据的详尽分析，可以充分考察不同亚组的治疗获益和风险，进而得到真实世界证据以支持更精准的目标人群定位。

对于靶向治疗药物的临床前和早期临床研究，生物标记物的识别甚为关键。利用人群队列中的组学数据、公共基因库信息，以及相关的临床资料等真实世界数据，通过多种机器学习类的目标靶向分析技术得到真实世界证据，可以支持靶向治疗药物的精确人群定位。

1. **真实世界研究的基本设计**
2. 实用临床试验

实用临床试验（Pragmatic Clinical Trial, PCT）又称实操临床试验，是指尽可能接近临床真实世界环境的临床试验，是介于RCT和观察性研究之间的一种研究类型。与RCT不同的是：PCT的干预既可以是标准化的，也可以是非标准化的；既可以采用随机分组方式，也可以自然选择入组；受试病例的入选标准较宽泛，对目标人群更具代表性；对干预结局的评价不局限于临床有效性和安全性等。与观察性研究所不同的是，PCT是干预性研究，尽管其干预的设计具有相当的灵活性。

由于PCT需要考虑所有可能的潜在因素的影响，包括各种偏倚和混杂因素的影响，故其研究设计和统计分析较为复杂，所需的样本量通常远超RCT设计。PCT如果采用随机化方法将减小混杂因素等偏倚的影响从而提供稳健的因果推断。此外，PCT在大多数情况下不采用盲法，对于如何估计和纠正由此产生的检测偏倚，需给予足够的重视。由于是在更接近真实临床实践环境下开展的研究，与其它研究类型相比，PCT所获得的证据被视为最好的且可行的真实世界证据。

1. 使用真实世界数据作为对照的单臂试验

使用外部对照具有局限性，主要包括医疗环境不同，医疗技术随时间变化，诊断标准不同，结局测量不同，患者的基线水平不同，干预多样化，数据质量难以保证等等。这些局限使得研究对象的可比性、研究结果的精确性、研究结论的可靠性和外推性等均面临挑战。

为克服或减少这些局限，首先要确保所采集的数据符合真实世界数据的相关质量要求。其次，在设计方面，采用平行外部对照设计要优于历史对照，前瞻性平行外部对照可采用疾病登记模式，保障数据记录尽可能完整、准确。其三，在统计分析方面采用恰当的方法，如合理利用倾向评分（Propensity Scores，PS）方法，虚拟配对对照（Virtual Matched Control）方法等。

1. 观察性研究

观察性研究所采集的数据无疑最接近真实世界，但其最主要的局限在于存在各种偏倚、数据质量难以保证、观测和未观测的混杂因素较难识别等，使得研究结论具有很大的不确定性。

观察性研究所收集的数据是否适合产生真实世界证据，以支持监管决策，关注要点包括：①数据特征是什么？（例如，相关终点的数据采集、记录的一致性、是否有缺失数据的描述等）②研究设计和分析的特点是什么？（例如，有无合适的阳性对照？考虑到潜在未测混杂因素以及潜在测量变异性，非劣效设计是否适用？）③预先确定了何种敏感性分析和统计诊断方法对真实世界数据进行分析？

分析观察性研究真实世界数据的关键技术是因果推断。真实世界研究常用统计分析方法见附录2。

1. **真实世界证据的评价**

评价真实世界证据应依从两个主要原则：真实世界证据是否可以支持需要回答的科学问题；已有的真实世界数据是否可以通过科学的分析得到所需的真实世界证据。

1. 真实世界证据和其所支持的科学问题

在决定使用包括真实世界证据在内的任何证据之前，首先应明确需要回答的科学问题。例如，药品上市后和其它药品联合使用的安全性考虑；已获批产品的扩大适应症研究；为某单臂临床试验建立稳健可靠的历史对照等。此处应考虑使用真实世界证据的初衷，是由于对应的科学问题面向真实世界，还是由于虽然面向临床研究但传统临床试验无法有效实施。如果是后者，真实世界证据是否可以代替传统临床试验，回答相同的问题并得到稳健的结论，将是衡量真实世界证据应用的重要准则。

1. 如何从真实世界数据到真实世界证据

回答这一问题一般至少应包括以下几个显著特征：①研究环境和数据采集接近真实世界，如更有代表性的目标人群，符合临床实践的干预多样化，干预的自然选择等；②合适的对照；③更全面的效果评价；④有效的偏倚控制，如随机化的使用，测量和评价方法的统一等；⑤恰当的统计分析，如因果推断方法的正确使用、合理的缺失数据处理、充分的敏感性分析等；⑥合理的结果解释；⑦各利益相关方达成共识。

最后需要特别强调的是，所有与产生真实世界证据相关的研究设计、假设、以及具体定义，均应事先在研究方案中明确阐述。同时，任何以药品注册为最终目的的对于真实世界数据和证据的使用，都需要提前与监管部门进行充分的沟通交流，以确保双方对于研究目标和方法等的共识。事后补救的数据引用、定义、分析以及解释，通常不被监管决策所接纳。

**参考文献**

[1] 孙宇昕, 魏芬芳, 杨悦. 真实世界证据用于药械监管与卫生决策的机遇与挑战[J]. 中国药物警戒, 2017(06):353-358.

 [2] 杨薇, 谢雁鸣. 美国AHRQ《评估患者结局的注册登记指南(第2版)》解读[J]. 中国中药杂志, 2013,38(18):2958-2962.

 [3] CFDA. 关于深化审评审批制度改革鼓励药品医疗器械创新的意见[S]. 2017.

 [4] Beckie TM, Mendonca MA, Fletcher GF, et al. Examining the Challenges of Recruiting Women Into a Cardiac Rehabilitation Clinical Trial[J]. Journal of Cardiopulmonary Rehabilitation & Prevention, 2009,29(1):13-21, 22-23.

 [5] Berger M, Daniel G, Frank K, et al. A FRAMEWORK FOR REGULATORY USE OF REAL-WORLD EVIDENCE[M]. 2017.

 [6] Center Watch. The value of real world evidence in rare disease research[EB/OL]. https://www.iqvia.com/-/media/library/experts/the-value-of-real-world-evidence-in-rare-disease-research.pdf?vs=1&hash=8E3A6A07E4ABA0E05F72B64D4789281134EF1A72.

 [7] Chalkidou K, Tunis S, Lopert R, et al. Comparative effectiveness research and evidence-based health policy: experience from four countries[J]. Milbank Q, 2009,87(2):339-367.

 [8] Daniels M. Streptomycin treatment of pulmonary tuberculosis[J]. Med Times, 1949,77(6):251.

 [9] Dreyer NA. Advancing a Framework for Regulatory Use of Real-World Evidence[J]. Therapeutic Innovation & Regulatory Science, 2018,52(3):362-368.

[10] EMA. Draft scientific guidance on post-authorization efficacy studies[EB/OL]. (2015-11-30)https://www.ema.europa.eu/en/news/supporting-better-use-medicines.

[11] EMA. New cell-based therapy to support stem cell transplantation in patients with high-risk blood cancer[Press Release][EB/OL]. https://www.ema.europa.eu/en/news/new-cell-based-therapy-support-stem-cell-transplantation-patients-high-risk-blood-cancer.

[12] FDA. Use of Real World Evidence to Support Regulatory Decision Making for Medical Devices. [S]. 2017.

[13] FDA. Framework For FDA's Real-world Evidence Program[EB/OL]. https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/RealWorldEvidence/UCM627769.pdf.

[14] Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process[J]. Am J Prev Med, 2001,20(3 Suppl):21-35.

[15] Institute OM. Initial National Priorities for Comparative Effectiveness Research[M]. Washington, DC: The National Academies Press, 2009.

[16] IMI GetReal. http://www.imi-getreal.eu/About-GetReal/Overall-objectives.

[17] James S. Importance of post-approval real-word evidence[J]. European Heart Journal Cardiovascular Pharmacotherapy, 2018,4(1):10.

[18] Jelínek T, Maisnar V, Pour L, et al. Adjusted comparison of daratumumab monotherapy versus real-world historical control data from the Czech Republic in heavily pretreated and highly refractory multiple myeloma patients[J]. Current Medical Research and Opinion, 2018,34(5):775-783.

[19] Makady A, de Boer A, Hillege H, et al. What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews[J]. Value in Health, 2017,20(7):858-865.

[20] Olariu E, Papageorgakopoulou C, Bovens SM, et al. Real World Evidence in Europe: A Snapshot of its Current Status[J]. Value in Health, 2016,19(7):A498.

[21] Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence - What Is It and What Can It Tell Us?[J]. The New England journal of medicine, 2016,375(23):2293-2297.

[22] Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. JAMA 2003; 290 (12): 1624-1632

[23] USA. National Institutes of Health Health Care Systems Research Collaboratory[EB/OL]. https:/ / www .nihcollaboratory .org/about-us/ Pages/ default .aspx.

[24] USA CONGRESS. H.R.34—21st Century Cures Act[EB/OL]. https://www.congress.gov/bill/114th-congress/house-bill/34/text?q=%EF%BC%857B%EF%BC%8522search%EF%BC%8522%EF%BC%853A%EF%BC%855B%EF%BC%852221st+Century+Cures+Act%EF%BC%8522%EF%BC%855D%EF%BC%857D&r=3.

[25] Velentgas P, Dreyer NA, Nourjah P, et al. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide.[J]. Methods, 2013.

[26] Victor L Serebruany, Moo Hyun Kim, Thomas A Marciniak, Worldwide reporting of fatal outcomes after ticagrelor to the US Food and Drug Administration, *European Heart Journal - Cardiovascular Pharmacotherapy*, Volume 4, Issue 1, January 2018, Pages 6–9, <https://doi.org/10.1093/ehjcvp/pvx024>

[27]von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies[J]. Journal of Clinical Epidemiology, 2008,61(4):344-349.

[28] White R. Building trust in real-world evidence and comparative effectiveness research: the need for transparency[J]. Journal of Comparative Effectiveness Research, 2017,6(1):5-7.

**附录1：词汇表**

**病例登记**（**Patient Registry**）：根据一个或多个预定的科学、临床，或政策目的，使用观察性研究方法收集统一的临床和其它数据的系统，以评价特定疾病、病症，或暴露人群的特定结局。

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

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

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

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

**前瞻性观察性研究（Prospective Observational Study）**：在研究开始时确定目标人群、并在研究开始前确定收集暴露/治疗和结果数据的观察性研究。

**实效比较研究（Comparative Effectiveness Study）:**是一种适合大多数研究类型的研究方法，系指在尽可能接近真实世界的环境下，考虑个体和群体两个层面，通过比较，从临床有效性和安全性、社会人文效应及经济效益三个方面评价其利弊，帮助患者、医生、决策者和服务购买者等利益相关方做出改善医疗服务的决策，以使最恰当的干预或策略在最适宜的目标人群和最佳的时机获得最好的效果。

实效比较研究基于真实世界背景，其应用范围较广，注重于面向自然人群的决策依据，因此需要尽可能全面地考虑众多因素对结局的影响，所采用的设计通常较为复杂，所需样本量巨大，同时对因果推断部分的统计分析有很高要求。

**实用临床试验（Pragmatic Clinical Trial，PCT）**：又称实操临床试验，是指尽可能接近临床真实世界环境的临床试验，它是介于RCT和观察性研究之间的一种研究类型。与RCT所不同的是，PCT的干预既可以是标准化的也可以是非标准化的；可以采用随机分组也可以自然选择入组；受试病例的入选标准较宽，对目标人群更具代表性；评价干预的结局不局限与临床有效性和安全性等。与观察性研究所不同的是，PCT是干预性研究，尽管干预的设计具有相当的灵活性。

由于PCT要考虑所有可能的潜在因素的影响，包括各种偏倚和混杂因素的影响，故其研究设计和统计分析较为复杂，所需样本量通常远超RCT设计。PCT在大多数情况下不采用盲法的，对如何估计和纠正由此产生的检测偏倚需给予足够的重视。尽管如此，但由于PCT是在更接近真实临床实践环境下开展的研究，与其它研究类型相比，它所获得的证据被视为最好的且可行的真实世界证据。

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

**随机对照临床试验（Randomized Controlled trial，RCT）**：是一种采用随机化分组方法并选择合适对照设计的临床试验。

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

**医保数据（Medical Claims Data）**：医疗保健提供方向保险公司提交的用以获得治疗和其他干预措施赔款的医疗索赔信息汇编。

**因果推断（Causal inference）**：基于真实世界数据，刻画干预或暴露与临床结局或健康结局的因果关系路径，充分考虑各种协变量和已测或未测混杂因素的影响，并控制可能的各种偏倚，采用恰当的统计模型和分析方法，做出干预或暴露与临床结局或健康结局的因果关系的推断结论。

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

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

**真实世界证据（Real-World Evidence，RWE）**：通过对真实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益或风险的临床证据。

**附录2：真实世界研究常用统计分析方法**

相较于RCT研究，真实世界研究中的因果推断需要特别注意对混杂效应的调整，因此会用到一些相对较复杂的统计模型和分析方法。这些方法既包括经典的统计方法，如传统多变量回归方法，也包括一些相对来说更加前沿和复杂的统计方法，如倾向性评分匹配方法和工具变量方法。本文件仅对这些统计方法做概括性说明，更具体的方法和应用细节见所附参考文献，也不排除未介绍方法的合理应用。

1. 描述性分析和非调整分析

对于描述性分析，根据不同的数据类型选择合适的描述统计量及统计图即可，主要包括：描述连续型变量取值范围、离散程度和集中趋势的指标，描述分类变量的计数和百分比指标，以及描述数据分布情况的统计图等。对于真实世界研究，正确有效的描述性统计分析可以发挥较为重要的作用。例如，在疾病登记队列研究中，按暴露因素的不同水平对相关协变量进行分层描述统计可帮助考察其均衡性；在倾向性评分匹配数据集中，按暴露因素分组汇总统计相关协变量可帮助发现残余不均衡等。

单变量或非调整性假设检验，如两样本*t*检验，可用于协助识别与暴露因素和/或研究结局相关的协变量。对于真实世界研究，通常需要从大量协变量中考虑可能的混杂效应，利用描述性统计分析对受试者的相关特征进行广泛和全面的探索性分析是非常必要的。

1. 调整分析
2. 协变量的选择
* 采用调整协变量的因果推断方法时，首先面临协变量选择问题。协变量选择方法大致分为两类，一类是基于暴露至结局相关路径构成的因果关系网络，识别出风险因子（Risk Factor）、混杂因素（Confounder）、中介变量（Intermediate Variable）、时变型混杂因素（Time-varying Confounder）、碰撞节点变量（Collider Variable）及工具变量（Instrumental Variable），将风险因子和混杂因素作为协变量纳入模型，同时避免纳入中间变量、碰撞节点变量和工具变量。这具体涉及到以下几个重要概念：
* **风险因子（Risk Factor）：**对结局变量有预测作用，但对暴露因素水平没有影响的基线协变量。在因果关系中如附图1所示，其中表示风险因子，表示处理或暴露因素，表示结局变量，对进行调整不影响总效应的估计，即既不引入偏倚也不消除偏倚，但是可以提高估计量的准确度和模型效率。

附图1. 风险因子（）与结局变量（）因果关系图示

*R*

*A*

*Y*

* **混杂因素**（Confounder）：既影响处理或暴露因素水平，又对结局变量有预测能力的因素，分为已测得混杂因素和未测得混杂因素。在因果关系中如附图2所示，其中表示处理或暴露因素，表示结局变量，和表示为两个未测得的混杂因素，表示一个已测得混杂因素，可以作为的代理变量，对进行调整可以消除对结局的混杂效应。

*A*

*U*2

*Y*

*U*1

*C*

附图2. 已测得混杂变量（）和未测得混杂变量（，）的因果关系图示

**- 中介变量**（Intermediate Variable）：指受治疗或暴露因素影响的治疗后变量，可能在或不在暴露与结局的因果路径上，其图示分别见附图3a和3b，其中表示处理或暴露因素，表示测量时刻的结局变量，表示中介变量，表示和之间的未测混杂因素。若要估计总效应，对于附图3a的情况，的总效应分直接效应和间接效应，对进行调整会屏蔽掉间接效应，导致总效应估计偏倚；而对于附图3b的情况，对进行调整会使原本独立的与之间产生相关性，继而使成为的混杂因素，反而引入了混杂偏倚，此时如果未对进行调整，就会导致总效应估计偏倚。在真实世界研究中，如果所调整的协变量取值不是基线时刻测得，就有可能导致过度调整偏倚。

*M*

*Y*

*A*

附图3a. 在处理（）与结局（）因果路径上的中介变量（）示意图

*M*

*Y*

*A*

*U*

附图3b. 不在处理（）与结局（）因果路径上，但与结局间存在未测混杂因素（）的中介变量（）示意图

**- 碰撞节点变量**（Collider Variable）：在因果图中，如果某一变量有两个独立的父母节点，那么在该变量处就形成一个碰撞节点，对碰撞节点处的变量进行调整会使原本独立的父母节点变得相关，有可能引入暴露与结局间的混杂效应，导致因果效应估计偏倚。在因果关系中如附图4所示，其中表示变量*L*和结局*Y*之间的未测混杂因素，表示变量*L*和暴露因素*A*之间的未测混杂因素，*L*即为一个碰撞节点变量，同时有和两个独立的父母节点。对*L*进行调整会使原本独立的和变得相关，从而在*A*和*Y*间引入混杂效应，导致效应估计偏倚。附图3b中的中介变量*M*也是一个碰撞节点变量。

*U2*

*A*

*Y*

*L*

*U1*

附图4. 碰撞节点变量()示意图

**- 工具变量**（Instrumental Variable）**：**指对处理或暴露因素水平有因果作用，但与结局变量间除了通过影响暴露因素间接影响结局变量外，没有其它因果关联的治疗前变量。工具变量必须与暴露和结局的混杂因素无关。在因果关系中如附图5所示，其中表示暴露因素和结局之间的混杂因素，即为一个工具变量。如果直接将工具变量纳入统计模型进行调整，会放大的混杂效应。对于工具变量另有其它分析策略，可用来消除混杂效应（见工具变量估计）。

*Z*

*A*

*Y*

*U*

附图5. 工具变量()示意图

但在通常情况下，真实的完整网络结构是未知的。在实际应用中，当部分因果结构已知时，协变量的选择方法可以基于相关专业背景知识，对所有观测到的、可能与结局相关的基线变量，已知的结局相关危险因素，以及治疗或结局的所有直接起因变量，都进行调整。另一类协变量选择方法是基于高维自动变量选择的方法，原理是利用反应变量间关联程度的统计方法，从数据中经验地学习变量间的相关关系，筛选出与处理因素和/或结局变量相关的变量作为协变量，主要包括前进法、后退法、机器学习类方法（如Boosting法，随机森林方法，LASSO法等）和自动高维代理调整方法。上述两类方法也可以结合使用，即首先利用专业经验知识，确定一个变量集合，然后使用适宜的经验学习方法，从中筛选出纳入最终分析模型的协变量。这样做的优点是限制了对经验学习的依赖性，在减小混杂效应的同时也减小了过度调整的风险。

1. 传统多变量回归

回归分析方法是调整潜在混杂变量影响、估计处理效应的一种常用策略，一般调整的变量是同时与研究的处理因素和结局指标相关，且在因果路径上位于处理因素之前的变量。如果是位于处理至结局因果路径上的中间变量，对其进行调整就会屏蔽掉部分的处理效应，导致过度调整偏倚。利用传统多变量回归对潜在混杂和效应修饰因素进行直接调整的方法，在观察性研究中已经有长期、大量的应用，这些方法在真实世界研究中也是同样适用的。使用回归分析方法需要注意相应的模型假设是否成立。例如，线性回归模型假设结局变量的均数是关于协变量的线性函数，因此，选择线性回归方法前需要先考虑这一假设能否成立。此外是否选择回归模型方法或其它方法，还与数据的特征有关。例如，如果研究的阳性事件数相对于纳入模型的协变量数足够多（如协变量个数的8倍或10倍以上），且处理因素非罕见时，传统的logistic回归方法就是一个较为合理的选择，可以考虑作为主要分析。而如果数据不是呈现出上述特征，就应该选择其它适宜的方法。另外，回归分析方法可能存在外推（Extrapolation）风险，即拟合的模型实际上超出了样本数据的表达范围。对于外推风险，可以利用倾向性评分等统计方法进行诊断。

在协变量数目较多的情况下，使用逐步回归等自变量选择方法能得到较为精简的模型，但对于自变量选择方法的类型（如基于*p*值或基于参数估计值变化程度），以及自变量的纳入排除标准（如*p*=0.05，0.10，或0.20），可能存在一定的主观性。对于一些真实影响疾病风险，但影响程度相对较温和的协变量，利用自变量选择方法（如逐步回归）确定的最终模型可能会漏掉这些重要协变量。此外，使用逐步回归方法可能导致低估处理因素参数估计的标准误。另一个策略是利用汇总的协变量，如PS或疾病风险评分（Disease Risk Score, DRS），进行回归分析。在结局事件相对较罕见情况下（如少于协变量个数的8倍或10倍），倾向性评分方法优于传统logistic回归方法；但对于罕见处理（或暴露）和常见结局的情况，传统logistic回归方法通常优于倾向性评分方法。

1. 倾向性评分

由Rosenbaum和Rubin提出的倾向性评分方法，是一种针对较多协变量情况的混杂效应调整方法。若用表示所有已观测协变量，表示研究感兴趣的处理或暴露因素（表示暴露），则倾向性评分定义为在观察到的协变量条件下，观察对象接受某种处理（或暴露）的概率。倾向性评分可以综合概括特征变量的作用，反映所有已观测协变量在两组间的均衡性。Rosenbaum和Rubin已证明，如果对原始协变量进行调整能够有效控制混杂效应，那么，仅对基于这些协变量的倾向性评分进行调整，也足以控制混杂效应。倾向性评分通常可以通过回归模型估计得到，例如常用的以观测协变量作为自变量，处理情况作为因变量的logistic回归模型：



对于处理（或暴露）因素常见但是结局事件罕见，或可能存在多重结局的情况，倾向性评分方法尤其适用。利用倾向性评分进行因果效应估计，通常可采用的方法包括倾向性评分匹配法（Propensity-Score Matching），倾向性评分分层法（Stratification/ Subclassification），逆概率加权法（Inverse Probability of Treatment Weighting，IPTW），以及将倾向性评分作为唯一协变量纳入统计模型进行调整分析的方法。

利用倾向性评分进行因果效应估计时，一定要先判断倾向性评分接近的患者在不同处理组间的协变量分布是否均衡。判断方法包括但不限于：按倾向性评分调整后，可视化地考察不同处理组间的倾向性评分分布情况，或者对各处理组间研究对象的协变量作统计检验。如果不同组间倾向性评分分布的重合性不高，则利用倾向性评分进行调整分析得到的效应估计值仍存在偏倚风险。对于重合性不好的情况可以考虑补救方案，如限制研究对象范围为各组倾向性评分分布的重叠区域。

在可能的条件下，匹配是倾向性评分较好的一种应用方式，再联合上述限制研究对象范围的方法，可以进一步提高各组间倾向性评分分布的重叠性。此外，如果提供匹配后所有研究协变量的组间均衡性的汇总结果，如绘制统计图或计算调整前后各协变量的标准化差（Standardized Differences，通常要求调整后不超过20%），并与随机化临床试验的协变量均衡性结果进行比较，将有助于评价匹配的效果。但是，倾向性评分匹配方法只能对已知的观察到的协变量进行控制，对未知或未观察到的混杂因素的影响，需要借助其它方法衡量协变量控制的效果以及分析结果的稳健性（例如，敏感性分析）。需注意，采用匹配设计所得因果效应估计的标准误与未匹配情况有所不同。

纳入倾向性评分模型的协变量应为混杂变量或与结局变量相关的变量，如果纳入仅与暴露因素有关的协变量会导致估计量的方差增大。

传统回归调整方法与倾向性评分匹配方法各有利弊，前者不能保证研究协变量一定均衡，后者可能会导致样本量减少，因此，进一步的敏感性分析是非常必要的。

1. 疾病风险评分

疾病风险评分与倾向性评分作用相似，是一个基于所有协变量的综合指标，若用表示所有已观测协变量，表示研究感兴趣的处理或暴露因素（表示暴露），则疾病风险评分定义为假定无暴露和特定协变量条件下，发生结局事件的概率。

估计DRS的方法可以分为两类，第一类方法利用研究样本的所有观测进行拟合，将处理因素与协变量作为自变量，研究结局作为因变量，拟合回归模型，以logistic回归模型为例，即：

，

然后将各研究样本的协变量取值回代入上述模型，同时将处理因素变量取值设置为对照组取值，计算所有研究样本的预测值，得到相应的DRS。对研究样本按照DRS分层，使用分层估计方法，即可以估计出处理因素对研究结局的因果效应。第二类估计DRS的方法仅利用对照（非暴露）组的研究样本（或处理因素发生前的历史数据，以及没有处理因素或处理因素发生率非常低的样本数据）拟合DRS模型，仍以logistic 回归模型为例，即



然后将所有研究样本的协变量取值回代入DRS模型，对所有研究样本计算相应的DRS预测值。

对于结局事件常见但处理（暴露）因素罕见、或者可能存在多重暴露的研究，DRS方法是一种较好的选择，能够平衡不同组间样本的基线疾病风险。对于处理（暴露）因素多水平，且部分水平较罕见的情况，建议选择DRS方法而不是PS方法。

1. 工具变量

采用上述传统多元回归方法、倾向性评分方法和疾病风险评分方法控制混杂的局限性在于：只能控制已知已测的混杂因素，但对于未知未测的混杂因素无法调整。而采用工具变量的因果效应分析方法不涉及具体地对混杂因素/协变量的调整，故能够控制未知的混杂因素，进而估计出处理与结局的因果效应。如果某变量与处理因素水平相关，并且对结局变量的影响只能通过影响处理因素实现，同时与暴露和结局的混杂因素不相关，那么该变量可以称为一个工具变量。确定工具变量后，即使存在未知未测的混杂因素，通过分别估计工具变量对处理和工具变量对结局的影响效果，并将两者对比，即可以估计出暴露对结局的因果效应。利用工具变量估计因果效应的方法，最大的难点在于找到合适的工具变量。

首先，工具变量必须与暴露和结局的所有观测到或未观测到的混杂因素不相关，否则工具变量方法会导致因果效应估计偏倚。其次，工具变量对结局不能有直接影响，除非通过处理至结局的通路间接作用于结局，否则也可能导致效应估计偏倚。最后，工具变量必须与研究的处理因素相关，而且相关性越高越好，如果相关性太弱，称为弱工具变量，利用弱工具变量得到的效应估计量在有限样本情况下性能较差，估计值波动较大且精确性差，任何可能的偏倚都会被过度放大。满足上述三个条件的变量，就可以作为工具变量来估计暴露对结局的因果效应。实际情况中有时可能难以甚至无法找到满足上述条件的变量，并且，对于如何评价找到的变量是否满足上述条件，尚缺乏特别适宜的统计方法。

工具变量法估计因果效应通常利用二阶段最小二乘估计方法，每个阶段均利用最小二乘原理:

第一阶段:拟合处理因素()和工具变量()的回归模型，得到处理因素的预测值；

第二阶段:用处理因素预测值对结局变量进行回归预测，回归系数即为处理因素致结局的因果效应的无偏估计。

工具变量的选择对于因果效应的估计尤为重要。工具变量对处理因素的影响方向及强度在研究总体中需要具备较好的同质性，否则，所估计的因果效应并非研究总体人群中的平均因果效应，而是在工具变量对其处理因素水平确有特定影响的子人群中的效应估计值，故称为局部平均处理效应（Local Average Treatment Effect，LATE）。需要注意的是，当暴露因素或处理因素为非连续型变量时，二阶段最小二乘方法得到的因果效应估计及误差估计可能存在潜在的统计学偏倚。

1. 缺失数据考虑

缺失数据问题在真实世界研究中通常难以避免，不仅结局变量可能缺失，协变量也有可能缺失，这就使得处理组间的可比性较难评估，进而可能导致处理效应估计偏倚。研究者和申办方应优化试验设计，尽可能地将缺失比例降到最低。

在进行主要分析前，应先尝试分析是否是真正的数据缺失以及缺失的原因。首先，没有数据并不等于数据缺失，例如病人没有做某项检查、或者医生根本就没有要求病人做某项检查，这些数据本就不应该存在、也不应该按缺失数据处理，这在真实世界数据中很常见。如果属于缺失数据，应尝试分析缺失机制，通常可以分为三种情况：完全随机缺失（Missing Completely At Random，MCAR）、随机缺失（Missing At Random，MAR），和非随机缺失（Missing Not At Random，MNAR）。完全随机缺失指数据缺失的概率与所有已测或未测的协变量及结局变量均无关。如果用表示结局变量（该变量缺失时表示为，未缺失表示为）；用表示暴露因素及相关的基线协变量；用表示缺失指示变量（表示缺失，表示未缺失），则完全随机缺失机制可表示为：。随机缺失指在给定的已测协变量取值和结局变量条件下，数据是否缺失是随机的，与潜在结局无关，即。而非随机缺失指数据的缺失概率与缺失值本身有关，同时也可能与已测协变量及结局数据有关。

对于缺失数据，选择正确的方法进行填补或分析是避免偏倚和信息损失的有效手段。恰当的填补方法应根据缺失机制和临床问题建立相应的假设来确定。一般来说，对于完全随机缺失的情况，利用样本均值或广义估计方程预测值进行填补皆可，也可以只对数据完整的样本进行分析。对于随机缺失的情况，可以构建统计模型，利用协变量取值对进行预测填补。一般建议采用多重填补（Multiple Imputation，MI）方法，如传统回归模型方法、马尔科夫链蒙特卡洛（Markov Chain Monte Carlo，MCMC）方法、全条件定义法（Fully Conditional Specification，FCS）等。此外对于纵向研究（Longitudinal Study）的随机缺失机制，还可以选择重复测量混合效应模型（Mixed Model for Repeated Measures，MMRM）进行分析，需注意MMRM方法虽然是一种推荐的处理缺失数据的方法，但其并不对缺失数据进行填补。对于非随机缺失机制的情况，可利用模式混合模型（Pattern Mixture Models，PMM）方法，分别对缺失数据和非缺失数据构建不同的统计模型进行分析。

此外，还有单一值填补方法。单一值填补方法的优点是原理简单、易于操作，缺点是即使在随机缺失条件下也不能保证结果正确有效，并且没有考虑缺失值的变异性，一般不建议用于主要分析。

在可能有协变量缺失的观察性研究中，对不同缺失模式可考虑使用一些现有的统计方法，包括完整数据分析法、多重填补法（MI）和倾向评分法（PS）。

完整数据分析法是将任何存在协变量缺失（或队列研究中有随访缺失）的患者排除后进行统计分析，这样做会降低检验效能。注意，仅当缺失数据与研究设计以及处理因素均不相关时，该法可以提供效应量的无偏估计。

多重填补法（MI）考虑到缺失值的不确定性，用可能的取值对缺失数据进行多次填补。如前所述，通常所用的MI是在随机缺失的假设下进行的，这意味着缺失数据可能与观察到的协变量有潜在关联，但与未观察到的变量无关。因为MI会产生多个数据集，有两种估计倾向评分的方法，即基于每个填补数据，或利用所有填补数据进行估计。这两种方法产生的多个处理效应会同时考虑填补数据内及填补数据间的变异，可采用Rubin法进行合并。

需要明确，三种数据缺失机制假设（MCAR、MAR和MNAR）通常均无法直接检测，只能通过对数据收集过程的描述和理解来说明其合理性。

需注意，最佳或唯一处理缺失数据的方法是几乎不可能存在的，也没有任何方法可以得到与原始完整数据一样的稳健无偏估计。应对缺失数据的最佳策略关键不是如何分析数据，而是体现在研究的精巧设计和理想实施上。

4.敏感性分析

上述各种因果推断方法均有各自的适用条件和模型假设。例如，倾向性评分匹配方法不需要满足工具变量方法的模型假设，而工具变量方法能够处理倾向性评分方法不适用的、可能存在未测混杂因素的情况。因此，针对因果效应统计分析方法的选择，可以进行敏感性分析，通过采用不同统计模型进行因果推断，对分析结果的稳健性进行评价，从而优先考虑具有良好稳健性的统计模型。除此之外，更全面的敏感性分析可参见《制定观察性实效比较研究计划指南》。

最后需要特别说明的是，对于分析结果的解释，真实世界研究与其它确证性研究一样，应尽可能全面、客观、准确、充分，不仅强调统计意义（如P值和置信区间），更要注重临床实际意义；不仅要看最终的结论，还要看形成该结论的整个证据链的逻辑性和完整性；不仅要看整体结论，也要关注亚组效应；此外，对各种可能偏倚和混杂的控制和影响需要给予尽可能详尽的阐述。

**附录3：中英文词汇对照**

|  |  |
| --- | --- |
| 中文 | 英文 |
| 21世纪治愈法案 | 21st Century Cures Act |
| FDA不良事件报告系统 | FDA Adverse Event Reporting System, FAERS |
| 阿尔茨海默病疾病进展和临床试验评估的数据驱动模型新方法的意见书 | Qualification opinion of a novel data driven model of disease progression and trial evaluation in mild and moderate Alzheimer’s disease |
| 标准操作规程 | Standard Operation Procedure, SOP |
| 标准化差 | Standardized Differences |
| 病例登记 | Patient Registry |
| 单臂临床试验 | Single-arm/One-arm Trial |
| 电子病历 | Electronic Medical Record, EMR |
| 电子健康档案 | Electronic Health Record, EHR |
| 多重填补 | Multiple Imputation, MI |
| 非随机缺失 | Missing Not At Random, MNAR |
| 分层法 | Stratification/Subclassification |
| 风险因子 | Risk Factor |
| 工具变量 | Instrumental Variable |
| 观察性研究 | Observational Study |
| 国家药监局药品审评中心 | Center for Drug Evaluation, CDE |
| 国家药品不良反应监测哨点联盟 | CASSA |
| 患者报告结局 | Patient Reported Outcome, PRO |
| 回顾性观察性研究 | Retrospective Observational Study |
| 混杂因素 | Confounder |
| 基线观测值结转 | Baseline Observation Carried Forward, BOCF |
| 疾病风险评分 | Disease Risk Score, DRS |
| 监管合规性 | Regulatory Compliance |
| 局部平均处理效应 | Local Average Treatment Effect, LATE |
| 临床试验 | Clinical Trial |
| 马尔科夫链蒙特卡洛模拟 | Markov Chain Monte Carlo, MCMC |
| 美国经济复苏刺激法案 | The American Recovery and Reinvestment Act |
| 美国联邦食品，药品和化妆品法 | Federal Food, Drug, and Cosmetic Act, FD&C |
| 美国食品药品监督管理局 | Food and Drug Administration, FDA |
| 模式混合模型 | Pattern Mixture Models, PMM |
| 末次观测值结转 | Last Observation Carried Forward, LOCF |
| 逆概率加权方法 | Inverse Probability of Treatment Weighting, IPTW |
| 欧盟药物管理局 | European Medicines Agency, EMA |
| 碰撞节点变量 | Collider Variable |
| 前瞻性观察性研究 | Prospective Observational Study |
| 倾向性评分 | Propensity Scores, PS |
| 倾向性评分匹配法 | Propensity-Score Matching |
| 热卡填补 | Hot-Deck Imputation |
| 人用药品注册技术要求国际协调会 | International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, ICH |
| 日本医药品医疗器械综合机构 | Pharmaceutical and Medical Devices Agency, PMDA |
| 时变型混杂因素 | Time-varying Confounder |
| 实效比较研究 | Comparative Effectiveness Research, CER |
| 实用/实操临床试验 | Pragmatic Clinical Trial, PCT |
| 适应性许可试点项目 | Adaptive Licensing Pilot |
| 数据标准 | Data Standard |
| 随机对照临床试验 | Randomized Controlled Trials, RCT |
| 随机缺失 | Missing At Random, MAR |
| 条件均值插补 | Conditional Mean Imputation |
| 外部对照 | External Control |
| 外推 | Extrapolation |
| 完全随机缺失 | Missing Completely At Random, MCAR |
| 完整性 | Completeness |
| 卫生信息系统 | Health Information System, HIS |
| 血管内皮生长因子 | Vascular Endothelial Growth Factor, VEGF |
| 医保数据 | Medical Claims Data |
| 因果推断 | Causal Inference |
| 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 |
| 质量保证 | Quality Assurance |
| 质量控制 | Quality Control |
| 中介变量 | Intermediate Variable |
| 重复测量混合效应模型 | Mixed Model for Repeated Measures, MMRM |
| 准确性 | Accuracy |
| 最差观测值结转 | Worst Observation Carried Forward, WOCF |