

真实世界研究指南

2018 年版

发布方：吴阶平医学基金会

中国胸部肿瘤研究协作组

真实世界研究指南

2018 年版

组长：

吴一龙 陈晓媛 杨志敏

执笔专家（按汉语拼音）：

程颖 陆舜 宁毅 唐智柳 田志钢 王洁 吴婷 徐晋小 夏素琴 余秋琼 周清

指南编委会专家成员（按汉语拼音）：

方维佳 高培 谷成明 马飞 毛磊 王学兴 阎昭 张博恒 朱昊 周脉耕 钟文昭

鸣谢：非常感谢各位行业专家给予的大力支持，没有大家的齐心协力，我们的指南无法如此顺利的完成，也特别感谢新屿信息科技（上海）有限公司在指南撰写过程中给予的大力支持！

目录

一. 背景	3
二. 几个重要概念的关系.....	4
三. 真实世界研究的思路与流程.....	5
四. 临床问题的确定和研究设计常见类型.....	6
五. 基于不同数据源的研究要素	7
1. 基于现有数据.....	7
2. 基于前瞻性数据.....	11
六. 数据管理流程及数据质量控制	15
1.数据管理流程.....	15
2. 数据质量控制.....	15
七. 偏倚和混杂的控制	16
1. 选择偏倚 (selection bias) 的控制.....	16
2. 信息偏倚 (information bias) 的控制.....	17
3. 混杂	18
八. 证据等级评价	20
九. 案例分析.....	22
十. 新技术方案及未来展望.....	24
附录 I 国外指南参考.....	25
附录 II 参考文献	26

一. 背景

近年来，国内外对真实世界研究（Real World Study, RWS）的关注度日益增加。首先，从政策层面看，美国食品药品监督管理局（Food and Drug Administration, FDA）于 2016 年 7 月 27 日发布《采用真实世界证据支持医疗器械的法规决策》草案，几经修订到 2017 年 8 月 31 日发布最新版^[1]，该指南表明了 FDA 对真实世界证据用于医疗器械法规决策中的态度。国家食品药品监督管理总局（China Food and Drug Administration, CFDA）也于 2017 年 10 月 8 日出台了《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^[2]，提到为满足临床急需药品医疗器械使用需求，加快审评审批，允许可附带条件批准上市，上市后按要求开展补充研究，此类补充研究也可部分归属于 RWS 范畴。

其次，从医疗大环境看，医疗大数据的构建给 RWS 提供了前所未有的便利。各级医疗机构、医保部门、医药监管部门积累了大量的医疗数据，各级数据库的电子化，以及各种电子设备的普及，各级数据库平台的建立，极大地增加了利用高质量数据进行真实世界研究的可能性。另外，患者、医生、医疗保险提供方、监管者以及政策制定者都在努力寻求各种来自于特定人群药品使用的准确信息。而 RWS 则将成为药品临床应用、医保制定、决策制定等各方参考的重要依据。

虽然 RWS 越来越多受到各方的重视，但因目前大多医疗数据分布零散，没有进行系统性的收集和结构化处理，RWS 所需样本量相对较大，数据异质性强，混杂和干扰因素多，对研究设计和统计方法的要求比传统研究更高。更重要的是，我国目前尚没有系统性的指南对 RWS 提供建议和指导，故如何帮助中国研究者更好地开展 RWS 研究，提高 RWS 质量，是编写本指南的初衷。本指南为国家重点研发计划项目（项目编号：2016YFC1303800）和国家公益性行业专项（项目编号：201402031）。

二. 几个重要概念的关系

1. 真实世界证据与真实世界数据

真实世界数据不等同于真实世界证据。真实世界数据通过严格的数据收集、系统的处理、正确的统计分析以及多维度的结果解读，才能产生真实世界证据。美国 FDA 在评估真实世界数据能否成为真实世界证据时需要看其数据的质量，包括真实世界数据与其结果的相关性以及可靠性等^[1]。

2. 真实世界研究与随机对照临床试验

RWS 是对临床常规产生的真实世界数据进行系统性收集并进行分析的研究，与随机对照临床试验（Randomized Controlled Trial, RCT）是互补的关系，并不对立。RWS 和 RCT 一样，都需要科学合理的研究设计，研究方案以及统计计划。另外判断 RWS 和 RCT 的标准不是试验设计和研究方法，而是研究实施的场景^[3]。RWS 数据源自医疗机构、家庭和社区等，而非存在诸多严格限制的理想环境。RWS 与 RCT 相比，各具特点，详见表 1：

表 1：随机对照临床试验与真实世界研究区别对照表

特点	随机对照临床试验	真实世界研究
研究目的	以效力研究为主（Efficacy）	研究目的多样，包括效果研究（Effectiveness）
研究人群	理想世界人群，严格的入排标准	真实世界人群，较为宽泛的入排标准
样本量	根据统计学公式推算获得，样本量较少	根据真实数据环境或统计学公式推算获得，样本量可大可小
研究时间	较短（多以评估结局指标为终点）	短期或者长期（以获得所有治疗以及长期临床结局为终点）
研究结果	内部有效性高	外部可推性强
研究设计	随机对照； 前瞻性研究	随机或非随机抽样，也可观察； 可前瞻，也可回顾
研究实施场景	理想世界：高度标准化的环境	真实世界：医疗机构、社区、家庭
数据	标准化，收集过程较严格规范	来源多样，异质性高

三. 真实世界研究的思路与流程

RWS 的开展须从临床问题的确定、现有数据情况的评估切入（采用既往回顾性数据或是前瞻性采集数据，进一步到研究设计的选择以及统计分析方法的确定、数据的管理、统计分析、结果解读和评价、以及根据需求判断是否加入事后分析（Adhoc analysis）等步骤（见图 1）。由于 RWS 可能存在一些内在的偏倚（Bias），这些偏倚可能限制了真实世界数据在因果关系上的推理和解读。因此，为了减少潜在的偏倚，需要谨慎而周密的研究设计，并且应该确定研究问题后尽早开始制定研究方案和统计分析计划。

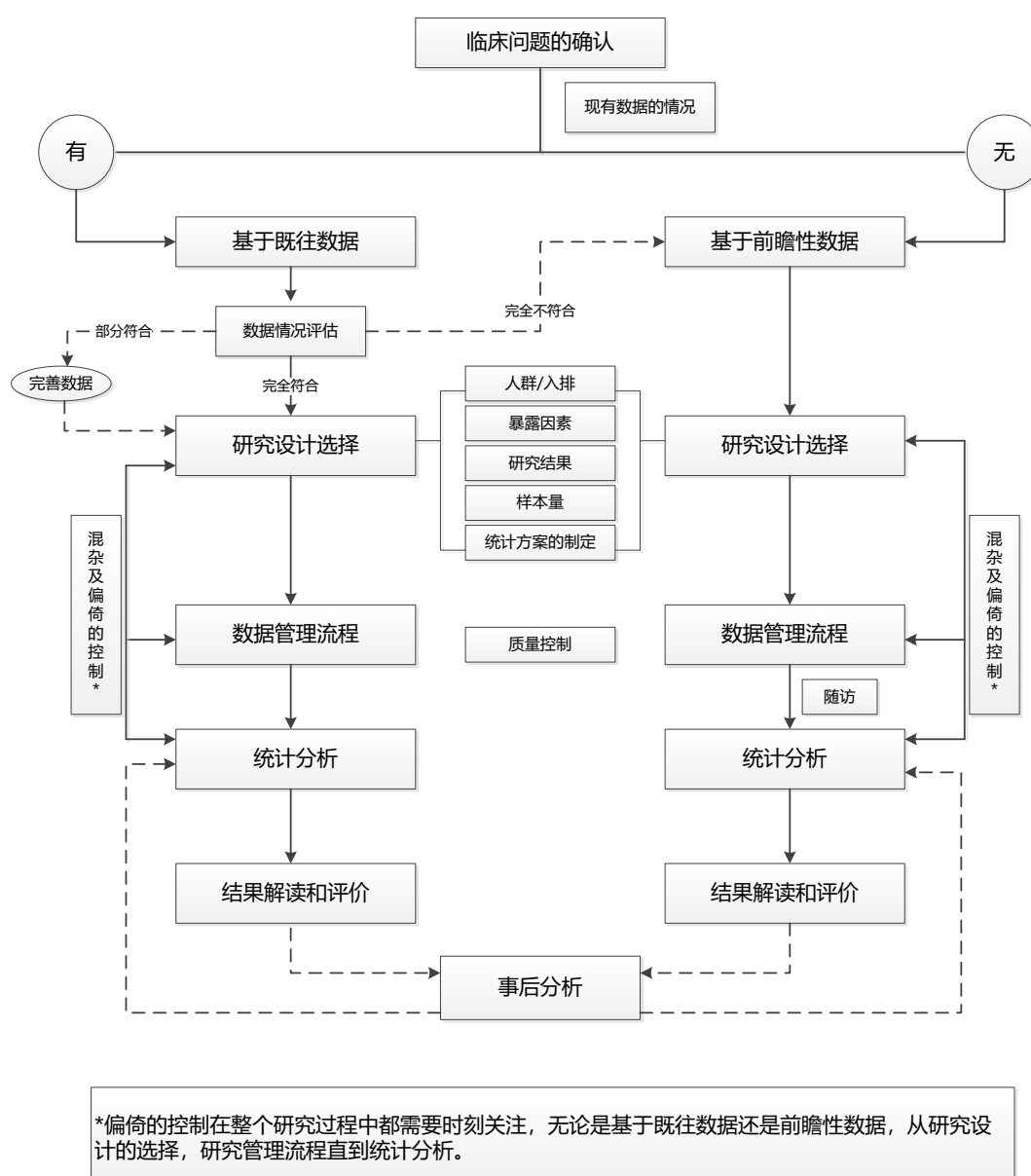


图 1：真实世界研究的思路与流程

四. 临床问题的确定和研究设计常见类型

RWS 通常会围绕着病因，诊断，治疗，预后及临床预测^[4]等相关的研究问题展开。病因研究主要是研究危险因素与疾病之间的关系，并研究引起人体发病的机制，如研究幽门螺杆菌感染与十二指肠溃疡的关系。诊断研究主要是研究某类新方法对特定疾病诊断的准确度研究，以判断新诊断方法的临床价值。治疗性研究主要是研究某类治疗方案对特定疾病的疗效及副作用研究。主要包括两方面：① 治疗方案对特定疾病的疗效研究；② 治疗方案的不良反应研究。预后研究是对疾病发展的不同结局的可能性的预测以及与影响其预后的因素研究，主要包含三大类：① 对疾病的预后状况进行客观描述；② 对影响预后的因素进行研究；③ 对健康相关生活质量的研究。临床预测研究则是寻找出最佳的疾病诊断或疾病转归的预测指标或症状等，主要包括诊断预测研究和预后预测研究。除上述研究外，RWS 也会涉及药物经济学研究等其它研究类型。

RWS 包括观察性研究和试验性研究。其中观察性研究进一步分为描述性研究（病例个案报告、单纯病例、横断面研究）和分析性研究【（巢式）病例对照研究、队列研究】，试验性研究即实效性临床研究（Pragmatic randomized clinical trial, PRCT）。除此以外，一些新型的研究设计如病例交叉设计和序贯设计等也被用在基于现有数据的研究中。下表是常见研究类型对应的临床应用场景及优劣势比较，供参考。

表 2：常见研究类型的优劣势比较

研究设计	拟解决的研究问题	常见临床场景	优势	劣势
病例对照研究	疾病相关的影响因素研究	- 影响疾病发病和预后因素的分析 - 预测研究	- 省时，省成本 - 适合研究罕见病 - 适合研究多个因素与一种疾病的关联	- 因果时序是由果及因，检验病因假说的能力较队列研究弱 - 混杂和偏倚 - 选择偏倚：例如不恰当的选择对照；病例与对照来自不同人群 - 信息偏倚：例如暴露信息不准确 - 回忆偏倚：例如研究者引入的偏倚
横断面研究	- 疾病或状态的分布状态（流行率） - 影响因素	- 疾病的发病患病或死亡率相关调查 - 疾病（或并发症）和影响因素的关联分析	省时，省成本	- 只能了解疾病的流行影响因素 - 判断因果关系的证据等级不高
单纯病例研究	- 罕见病的研究 - 基因环境交互作用的研究	某疾病的分子标记物分析中的组织样本的收集	- 适于医院开展研究 - 特别适合肿瘤及罕见病的研究	- 无对照组 - 所研究疾病的患病率不宜超过 5%

		特殊疾病队列	- 在检测基因与环境交互作用时，可信区间更窄 - 因无对照组，从而避免了对照选择所引起的偏倚 - 省时，省成本	除了可出现病例对照研究的病例选择所引起的常见偏倚外，还存在不同亚人群暴露率和基因频率不一致所引起的偏倚
回顾性队列研究	- 病因研究 - 预后研究 - 治疗性研究	- 临床治疗和疗效/结局的生存分析 - 预测研究	- 可以直接获得发病率，直接估计相对危险度 - 因果时序合理，检验病因假说的能力较强 - 了解疾病的自然史 - 获得一种暴露与多种疾病结局的关系 - 相比前瞻性队列研究，省时省成本	- 不适于发病率很低的疾病的病因研究 - 数据和信息的缺失 - 偏倚和混杂 - 矛盾数据需要特别注意和处理
前瞻性队列研究	- 病因研究 - 预后研究 - 治疗性研究	- 检验病因假设 - 临床治疗和疗效/结局的生存分析	- 因果时序合理，检验病因假说的能力较强 - 了解疾病的自然史 - 获得一种暴露与多种疾病结局的关系 - 所收集的资料相对完整可靠，一般不存在回忆偏倚	- 不适于发病率很低的疾病的病因研究 - 要求随访观察，时间周期长 - 失访难以避免 - 随着时间推移，未知的变量引入人群可能导致结局受影响 - 研究的设计要求高，实施难度大，费用高
实效性临床研究	- 预后研究 - 治疗性研究	- 临床疗效/结局（如心梗、生存质量、死亡、成本等）以及安全性的评价 - 成本-效果分析等卫生经济学评价	- 可在不同等级的医疗机构开展研究 - 真实世界患者（异质性相对较大、限制相对少） - 相对灵活可变（可调整方案），更符合日常医疗实际 - 外部可推性较好 - 可以通过随机分组平衡组间已知和未知的预后因素，最大程度提高组间的可比性，从而增强论证强度	- 样本量通常较大 - 其它和前瞻性队列研究相似

五. 基于不同数据源的研究要素

1. 基于现有数据

1.1 现有数据利用

现有数据主要包含电子病历（electronic medical record, EMR），电子健康档案（electronic health record, EHR），医保数据（claims data），出生死亡登记，公共健康监测数据以及区域化医疗数据等，这些数据数量非常庞大，但由于数据的采集并非为某特定研究目的而设计，数据分散，异质性高，数据的完整性及准确性也存在一些问题。另外医保数据一般由各级政府机构掌握，可及性较弱。

1.2 数据可行性评估

数据可行性评估首先基于待研究的临床问题，确定主要研究变量比如待研究的治疗措施，关键基线信息，主要研究结局包括患者主要人口统计学特征、患病史、并发症、合并症和实验室指标等关键数据是否存在；其次对缺失数据的数量和类型的影响进行全面评估，包括主要研究变量及其

他相关研究变量，可以通过抽样或全数据集来检查关键变量的数据缺失程度和模式来实现。另外还需进行数据质量评估：包含数据的准确性，可靠性，完整性及可溯源性等评估。

1.3 研究设计的考量

- **研究人群和入排标准**

回顾性队列和病例对照研究中，首先需要确定研究对象尤其是对照的选择和入排标准的设定。病例对照研究中，对照应尽量选择内部对照，选择没有发生研究结局的人群，且与病例来自同一人群；对照的选择不受暴露因素的影响，即除了暴露因素外，对照应与病例在其他特征上相似；病例对照比例可以从 1:1 到 1:4 不等。单纯病例研究设计则无需对照组。回顾性队列设计中，须根据研究问题清晰定义暴露，比如可以是有/无某治疗方案，暴露的程度如剂量，或者暴露的模式如顺序等等。非暴露组的人群应在除了暴露因素外与暴露组尽可能相似。在确定研究人群的标准后，要在数据库中使用一定的算法和代码来识别研究人群，如 ICD 编码和药品编码。值得注意的是，任何一种编码由于不同机构医疗水平或者电子病历系统平台不同，对于疾病诊断的准确性和完整性都会有所不同。因此，综合各项编码和实验室诊断等联合的识别方式，是基于数据库开展研究的常用方法。

- **暴露因素和研究终点**

暴露因素和研究终点是研究中首要考量的关键因素。在数据可行性评估及纳入和排除研究对象的阶段之后，进一步确认暴露状态及研究终点的准确性和代表性，是否有不同数据佐证关键数据的准确性。如果研究基于已有病例数据，则需特别注意收集数据的完整性、真实性和可溯源性；如果研究基于患者的自我报告或者研究者的回忆，则须强调数据的准确性和真实性，以防出现回忆偏倚。

- **样本量**

RWS 同样需要将样本量的估计作为不可缺少的设计要素之一。对于存在假设检验的分析性研究（如病例对照设计、队列设计等），如果样本量不足，会导致没有足够的把握度（power）去检验提出的问题和假说。基于不同研究类型需要依据不同的统计分析方法，确定重要参数，定义 I 类

错误和把握度，在保证研究具有一定可靠性条件下，估计并确定最小样本量，确保研究同时具备科学性和经济性。另一方面，RWS 往往采用较宽泛的纳排标准，有时需要随访较长时间来研究长期临床结局，充分反映实际的临床实践，因此应在确定最小样本量的基础上，尽可能地扩大样本量以保证其能够覆盖更广泛的患者群体并考虑到较长随访时间导致的失访的可能性。在具有异质性的患者群体中可进行亚组分析，从而拓展研究的意义。具体样本量的估算需要在临床医生、统计师和流行病学家等的合作下共同完成。

• 统计方法的考量

应在研究问题确定后，尽早制定研究方案和统计分析计划，将解决主要研究目的统计方法纳入。

RWS 的统计方法和疗效比较研究 (Comparative effectiveness research, CER) [5] 的分析方法有类似之处。因 RWS 接近临床实际，研究对象的纳入限制较少、人群的异质性较大、自主选择治疗措施等造成潜在偏倚和混杂，因此统计方法更多是关注如何减小和控制偏倚和混杂。常见的有匹配、分层分析和多变量分析。在较多风险因素或者研究因素的情况下，使用多变量分析将多个因素同时纳入模型，会由于共线性等问题，使得模型无法正常运行。倾向性评分匹配 (propensity score matching) 或者分层 (stratification) 则是解决该类问题的常用统计方法。成本效益模型、贝叶斯模型等也常应用于 RWS 的研究设计中。

另外，利用已有数据库开展的预测研究也是常见的 RWS 类型之一，是对疾病各种结局发生概率及其影响因素的研究。传统的统计方法包括 Logistic 回归和 Cox 回归以及列线图 (nomogram)，可以用来预测疾病转归或者并发症的发生概率；另外，近年来发展出的基于真实世界大数据的机器学习 (machine learning) 的方法也是用于预测研究的常用工具。

• 处理缺失数据

在 RWS 中，数据的缺失是一个不可避免的问题。预防策略和统计调整可以减少缺失数据对研究结果的影响，提高结果的可靠性。关注不同研究类型可能出现的数据缺失也可帮助减少缺失数据对研究结果影响。

使用电子 EHR、EMR 或者医疗保险等数据源会出现不同的问题。由于研究者往往无法采集额外的数据，因此研究前数据可行性评估是非常重要的环节。例如，对于临近死亡的患者，症状可能记录不完整，而这种缺失信息可能使这些患者的相关数据看上去比实际状况更健康而造成研究偏倚。有时数据的缺失并非随机的，如果忽略缺失数据的内在规律，那么会丢掉重要的信息。对于缺失数据，在能够溯源的情况下，则尽可能补全相应缺失数据；对于无法溯源的情况，则展开探索性分析，明确缺失值在各个研究因素中的分布情况，判断其分布是否随机，如有偏倚，则需考虑后期分层分析。

1.4 其他常见问题

• 不同数据库整合的问题

对于不同数据库的整合，一般建议首先对各自数据库中数据进行质量评估和分层分级，确定好需要整合的内容。在整合过程中，矛盾的数据是重中之重，需要重点关注。建立统一的数据标准，将不同数据库的数据结构进行标准化处理。在整合过程中对于无法整合的数据进行处理，同时需要注意因整合造成的系统误差，以及关注整合后的主要研究因素，暴露因素及主要混杂因素的影响。

• 分子标记物相关研究的注意事项

开展分子标记物相关的 RWS，由于采集生物样本的难度以及其测量的特殊性，研究设计需要注意以下的问题。

基于已有数据库的回顾性研究需要考虑患者的基线特征和代表性的问题，由于缺失的生物样本而导致的数据缺失，多重检验导致的“假阳性”结果，数据分析和结果解释中的偏倚，以及研究结果是否具有可重复性等。

前瞻性的研究则需考虑，可能有生物样本但是没有详细的患者基本信息的情况；样本量小，且患者人群缺乏代表性；实验检测方法是否经过验证；有分子标记物测量结果，没有或仅有有限临床结局数据（如总生存率，无进展生存期等）的情况；临床数据或者研究终点数据的质量等。

- **用数据库开展研究，知情同意和伦理审批的相关事宜**

真实世界研究的数据收集，患者的信息有可能会成为研究资料，因此患者本身在某种程度上可能也已成为“受试者”，所以真实世界研究同样需要符合伦理的要求。因此某些情况下，真实世界研究也需获得伦理审查委员会的备案或批准。

知情同意是保证研究符合伦理要求的一个重要环节，它是一个持续的完整过程，保护受试者权益。因真实世界研究多不是预设方案，尤其是回顾性研究，收集的大量信息多属于历年数据，且涉及到对大规模病例或是生物样本数据进行研究，所以获得每一位受试者同意其数据用于临床的研究有很大的挑战。如果经伦理审查，认为课题研究不大于最小风险，且研究者使用受试者数据不会对患者造成不利的影响和受试者重要隐私信息的泄露，一般可考虑免除知情同意^[6,7]。知情及伦理的相关具体要求需根据各伦理委员会的实际要求而定。

2. 基于前瞻性数据

2.1 前瞻性数据

前瞻性数据主要包括如临床试验的补充数据，PRCT，注册登记研究（Registry），健康调查，公共健康监测等，该类数据在收集之前已确定具体的研究目的，需要收集的数据也很明确，故数据比较规范比较标准，完整性准确性比较好。

2.2 研究设计的考量

基于前瞻性数据的 RWS 一般采用队列研究的设计，队列研究的展开首先起始于要研究的问题，基于健康结果的病因或者风险因素的假说，通过对研究对象的追踪，观察不同群组健康结局的发生情况，进而建立暴露因素和健康结局的联系。

临床研究中，往往有计划地招募患者参加，进而随访患者疾病的复发、好转、痊愈、死亡等，研究不同的治疗或者某些疾病特点是否和不同疾病结局的相关关系。前瞻性队列研究的样本量和随访时间同样重要，研究的样本量越大，随访的时间越长，观察到的健康结局数量越多。开展前瞻性队列研究的主要考虑因素有以下几点：

- **研究人群的选取**

疾病注册或患者注册登记就是系统性地收集某些特点的患者，比如某些诊断、治疗或疾病症状等，这样的登记研究可能使用患者的病历记录和相关信息，而不是所有的可以获得的庞大的数据。不同的疾病注册登记的样本量或者收集信息的深度和广度可能是不一样的，可能是一个单中心或多中心的研究。考虑到不同的研究目的和执行的可能，这样的研究也可以设定一些入排标准。一般来说，严格的入排标准，是为了加强研究本身的内部有效性（internal validity），宽泛的入排标准，会提高研究结果的广泛代表性或外推性（external generalizability）。可能很难做到研究结果的内部有效性和外推性两全其美，在设计阶段，如何入选患者需要在临床医生和流行病学家等合作下共同完成评估资源和操作的可能性，平衡研究的内部有效性和外部可推性。

- **基线调查的研究内容的尽量丰富完整**

前瞻性队列研究的基线研究，本身也是一个横断面的研究。在计划和展开研究早期阶段，可以考虑是否要利用横断面信息，一方面回答具体的科学问题，提示未来继续研究的基础；另一方面，也是对数据的可行性和部分数据质量的检验。在这一分析过程中，应该特别关注对主要暴露因素，甚至暴露水平的测量和评估。例如主要暴露因素效度的研究（validation study），为未来的主要研究目的的分析奠定质量基础。另外，在考虑收集主要研究暴露因素和相关基线数据之外，应该考虑未来研究拓展的潜在可能性，尽可能收集丰富完整的数据。这样的基线数据会定义患者在研究基线时间点更多的暴露因素和暴露水平，使验证研究假说或研究其他问题成为可能。最后，基线信息的完整采集，对于在后续分析阶段控制偏倚和混杂有重要的作用。

- **样本量和研究深度的平衡**

有些情况下，前瞻性队列研究限于研究条件，研究某些科学问题受限于样本量小，没有足够的把握度来验证假说，在这种情况下，鼓励多个群组研究的共同研究(cohort consortium)，来扩大样本量，促进科研合作。同时，研究的大样本也相应地减少了抽样误差。在小样本量研究的情况下，尽可能在数据采集的深度上努力，在创新性方面做出探索。具体样本量计算可根据研究实际情况在临床医生、流行病学家和统计师的合作下共同完成。

- **提高患者依从性，长期随访患者**

前瞻性队列研究可能会随访数十年，未来研究的健康结局不仅可能包括电子病历、死亡记录、保险登记等电子化平台提供的信息，也应该包括未来可能随访等获得的信息。有些研究时间可能会超过研究者最初计划的跟踪时间，所以提高患者依从性，避免失访至关重要：

- a. 充分知情：患者脱落或失访的两个常见原因分别是不良反应或频繁检查，让患者充分了解研究流程，提前告知可能不良反应并进行相应预防措施可有效提高患者依从性。
- b. 人文关怀：研究过程中的医患互动，可以发放健康信息，更新研究结果等，争取患者的配合；可帮助患者及时解决研究中碰到的相关问题，增强医患间的相互了解，提高患者依从性。
- c. 定期提醒：定期系统或人工提醒可提高患者用药及访视的依从性。
- d. 多种沟通方式：保留患者及家属多个长期的联系方式，包括住址，手机，微信，电子邮箱，线上等，并及时更新，可有效避免患者换号后导致的失访。对于失访的患者，也应在不同时间段电话或者通过其它联系方式争取，减少失访。

• 失访及缺失数据考量

研究者需要在研究开始时针对缺失数据制定计划，尽可能防止数据缺失，同时为处理重要变量的缺失数据制定计划。

失访是前瞻性队列研究的一个重要问题。在研究随访期间，某些研究对象由于各种原因不能或不愿意，或者死亡等原因不能继续参与研究。在需要长期随访的研究中，失访发生的可能性更高，需要更加重视。一般来说，失访低于 5% 的时候，引起的偏倚比较小，但是失访超过 20% 的时候，就必须加以重视和分析。如果失访的原因与暴露因素和研究结局有关联，即便是 < 20% 的失访也可能会引起偏倚^[8]。如果失访的原因是完全随机的，和暴露因素以及研究结果没有联系，失访低于 60% 一般是可以接受的。如果失访在某些明确因素的条件下是随机的，而且低于 60%，在分析阶段，通过对这些确定因素的分层分析，可以控制失访带来的偏倚。此外，即便失访不一定会带来偏倚，仍会影响研究的准确性，或者说导致研究可信区间变宽。因此，存在失访的情况下，应

该进一步分析暴露组和非暴露组失访比例是否有显著性不同，失访是否会和一些关键指标存在关联，进而判断失访条件下得到的研究结果是否低估或者高估了实际情况。

2.3 其他衍生研究类型

• 队列研究中的巢式病例对照研究

巢式病例对照研究是在现存队列研究基础上的病例对照研究。在随访研究人群后，选择新发患者作为病例，然后通过年龄和性别等因素配比，选择对照组，进而比较病例和对照的暴露因素是否不同，来研究疾病和暴露的关系。一般来说在暴露因素的测量难度大、费用比较高或者病例比较罕见的情况下，适用选择巢式病例对照研究。

• 实效性临床研究

尽管实效性临床研究（PRCT）在 RWS 中划分在试验性研究，然而其研究设计和前瞻性队列研究有相似之处。其主要用于常规临床实践中疗效对比研究，另外，成本效果分析（cost-effectiveness analysis, CEA）也是 PRCT 的关键组成部分。随机化分组是 PRCT 的关键，用来提高组间可比性，减少选择偏倚。通常选择两种待比较的临床干预措施或方案，采用相对宽泛的入选标准，允许不同研究对象间存在临床异质性，以保证试验结论能最大程度地外推。研究中干预方案的低标准化是 PRCT 重要特征之一，即研究在实际临床情况或常规临床实践下进行，并尽可能减少对常规治疗的干预。PRCT 对照组很少选用安慰剂，通常选用常规或目前公认最佳的临床治疗方法^[9]。1954 年美国病毒学家 J. Salk 主持实施的临床研究“索尔克脊髓灰质炎疫苗试验”（Salk field trial of the polio vaccine），以及 2015 年由 National Patient-Centered Clinical Research Network 主导的“阿司匹林心血管获益研究”（Aspirin Study, ADAPTABLE），都属于 PRCT^[10]

六. 数据管理流程及数据质量控制

1. 数据管理流程

RWS 研究流程管理的核心整体来说还是加强数据质量，提高研究效率，控制研究成本。研究者可充分发挥软件，移动端和人工智能结构化等新技术功能。加强电子数据采集系统（EDC）定制，增强系统逻辑核查功能，让系统自动进行数据核查，时时保障数据质量；从研究设计层面应该减少不必要的随访和检测，尤其是来院随访。充分利用在线随访功能，即充分让患者参与，让患者报告结局，自动进行患者随访提醒、问卷量表推送，患者端数据采集及医患沟通，提高随访效率和质量；数据点分级管理，强制性保证关键数据的收集的准确真实性。从研究监查与技术层面进行技术创新：使用移动 APP 进行数据源收集，线上中心化录入配合文本识别（OCR）和智能结构化功能以节约录入成本。同时远程监查以减少差旅成本。

2. 数据质量控制

数据质量控制是确保研究数据真实、准确、可靠的关键。研究各个阶段都需要对可能影响数据质量的各个因素和环节进行控制，涉及从数据收集、处理、到统计分析报告的全过程。参照数据质量评价 ALCOA+原则^[11-13]，数据的可溯源性、完整性、一致性、及准确性等指标在 RWS 尤为重要，需要重点关注。数据质控则需要建立完善的 RWS 数据质量管理体系、完善的标准操作流程（SOP）以及人员定期的培训。主要包括：

- 1) 保证数据源质量，确保数据源信息的完整性和准确性，减少数据源本身的缺失和偏差。临床病历作为关键数据源，其不仅要符合病历书写规范、医院三级质控要求等，还应提高病历质控标准以满足科研需要。
- 2) 在采集数据前，制定详细的研究设计方案和分析计划。评估确立采集字段，确认关键字段已被收集，制定相应的 CRF 和数据库架构；
- 3) 建立数据采集和录入的标准指南，确保录入数据与数据源的一致性。对于录入过程中的任何修改，需要提供修改原因并留下完整的稽查轨迹。

4) 制定完善的数据质量管理计划，确立关键字段；制定系统质控和人工质控计划，确保数据的真实性、准确性和完整性。数据源核查确认是保证研究数据真实完整的必要措施之一，RWS 涉及到大规模的数据，可充分利用系统实时自动逻辑核查来加强质控，降低人工质控成本；对于关键字段，可进行 100%原始数据核查（SDV），其他字段可根据实际情况降低核查率。

5) 数据标准化，建议使用标准化字典。RWS 信息来源复杂，数据的标准化是保障数据质量的基础和关键环节。

为保障 RWS 的发展，保证数据的可溯源性和一致性，可运用新技术，充分利用电子化系统，增强系统逻辑核查功能等，加强 RWS 的数据质量。

七. 偏倚和混杂的控制

临床研究中误差的来源可以分为两类：一类是随机误差（random error）；一类是系统误差（systematic error），即我们经常提到的偏倚。在 RWS 的设计和实施中，偏倚是一个需要特别审视的问题。

1. 选择偏倚（selection bias）的控制

选择偏倚是在 RWS 中比较常见的偏倚。大多数情况下可以通过科研设计来减少或者消除，如采用匹配或者随机的方法。比如临床科研常见的一个实例，某研究数据库里有 3000 例患者的数据，经过一系列入排之后仅有 500 人可用于分析，如果其中一部分人因为数据缺失被排除研究，那么就要考虑是否存在选择偏倚的问题，即数据全面完整的患者和因数据缺失而排除的患者之间是否可比，这是一个需要考虑和评估的要点。如果信息缺失的患者因为某种治疗疗效不佳退出，甚至死亡而导致失访，那么排除后的数据而产生的结果会严重影响研究结果的准确性、可靠性以及外推性。

对于选择偏倚需采取的控制方法：

1) 严格掌握研究对象的纳入或排除标准。第一，不患与研究因素有关的其它疾病。例如研究吸烟和肺癌的关系，那么就要将慢性阻塞性肺病（COPD）排除出研究人群。第二，在某些方面与病例

组可比。如对照组和病例组在基线信息上尽量可比，如年龄，性别，病情严重程度，经济状况等。

2) 尽量提高应答率，减少失访率，并对失访的患者进行评价。

3) 尽可能采用多种对照。理想的是以人群中全体病例和非病例(或其有代表性的样本)作为研究对象。如以医院病例为研究对象，宜在多个医院选择对象，且最好有 2 个对照组，其中一个对照组来自社区一般人群；在队列研究中，最好也应设多种对照，以减少选择偏倚对结果的影响。

PRCT 的设计也可以视为控制选择偏倚的一种手段。通过随机分组平衡组间已知和未知的预后因素，最大程度提高组间的可比性，将偏倚最小化，从而增强论证强度。后续的随访过程遵从观察性研究的原则，即研究在常规临床实践下进行，尽可能减少对常规治疗的干预，反映真实临床实践中的情况。

2. 信息偏倚 (information bias) 的控制

信息偏倚主要来自资料收集和解释过程中的错误信息，比如问卷的问题，生物标本的问题，数据管理的问题，设计与分析的问题。而产生这些不正确信息的原因可以是研究对象本身的记忆误差，也可以由研究者的态度或方法不当所致，更重要的是在研究设计过程中对调查表设计，指标设立和检测方法的选择缺乏科学性和合理性，因此要在研究的不同阶段控制和消除影响信息准确性的各种因素。

对于信息偏倚需采取的控制方法：

- **研究设计阶段：**

在研究设计中对暴露因素必须有严格，客观的定义，并力求指标定量化。要有统一、明确的疾病诊断标准，调查表的问题应易于理解和回答，例如：当询问对象是否吸烟时，首先要明确对应于本次研究的吸烟的定义，如：“每日吸烟一支以上连续一年以上。”调查前应开展预调查，充分估计调查实施过程中可能遇到的问题以及各调查项目的可行性。

研究对象应清楚地了解本次研究的目的、意义和要求，以获取其配合和支持。对于涉及生活方式和隐私的问卷，应事先告知对象所有应答均获保密并将得到妥善保管，必要时可采用匿名问卷。

调查员需经过严格培训，能正确理解调查的意义、方法和内容，能严谨客观地从事资料收集工作。研究者应定期检查资料的质量，并设立质量控制程序。

- **资料收集阶段：**

信息偏倚与调查对象的记忆程度有关，在研究中可对同一内容以不同的形式重复询问，以帮助调查对象回忆并检验其应答的可信性。如询问吸烟暴露年数时可问：“你一生中共吸过几年烟”和“你一生中哪几年吸烟，哪几年不吸烟”。为了便于对象理解并准确地定量，可在询问中使用实物如杯子，量匙等来为某些暴露因素如每日饮酒量、盐摄入量等定量。向研究对象提供有关因素的实物照片也可以帮助对象直观地回答。

为了避免主观诱导对象、除了严格培训调查员外，在临床试验和某些现场研究中，应尽可能采用“盲法”以消除主观因素对研究结果的影响。但在采用“盲法”的同时需考虑其伦理学可行性。

研究中的各种测量仪器、试剂和方法都应标准化。应使用同一型号的仪器并定期校验，试剂必须是同一品牌，同一来源并力求同一批号，检测方法要统一，由专人测定。

对于信息偏倚，除了在方法学上杜绝其来源外，对其所致的错误的分类结果，可进一步在资料分析过程中加以测量、校正，并进行相应的灵敏度分析。

3. 混杂 (confounding) 的控制

混杂是一类在 RWS 中值得全程关注却又无法完全避免和控制的因素，只能尽可能的识别和控制以减小混杂对结果的影响。识别和选择混杂因素，通常是结合临床和流行病学专业知识进行选择。

对于较成熟的领域，任何已有证据提示为混杂的变量都应该考虑；可以基于文献回顾以及研究人员专业领域的知识和理解。对于崭新的领域，考虑那些与疾病有关也可能与暴露有关的因素。如果难以确定，可以在资源允许的条件下，考虑对所有与疾病有关的因素都进行测量，尽可能的收集更多的数据点。

对于混杂的常用的控制方法如下：

- **限制**：研究设计开始阶段，针对某些可能的混杂因素，对研究对象的入排标准加以限制是常用的控制混杂的方法，然而这样会使得研究对象减少。优点在于筛检大量个体，只入选其中一部分，增加了研究对象的同质性（homogeneity），提高了研究的内部有效性，但是缺点是降低了研究效率，可能会影响结果的外推。而限制范围如果太宽或不当，又可能有残余混杂。
- **匹配**：队列研究中，匹配是暴露者与非暴露者在某因素上的匹配，一旦匹配，原则上即可完全控制匹配因素引起的混杂，不必在统计分析阶段进一步控制。而病例对照研究中，匹配后还需要按照匹配因素进行分层分析，这是控制混杂的必要条件，并且，匹配并非直接控制混杂，而是提高了控制混杂的效率。值得注意的是要防止匹配过度。
- **分层分析**：目的在于估计和控制混杂的影响，评估和描述效应修正因子（modifier）不同水平分层中的研究结果。方法是将研究资料按照混杂因素来进行分层。分层分析缺点在于：一次只能分析一种暴露-疾病关联；连续性变量转变为离散性变量，丢失了一定程度的信息，可能造成残余混杂；需要控制的混杂较多时，分层则不是最明智的选择。
- **多因素分析**：将多个可能的混杂因素变量引入分析模型，可以是多元线性模型，logistic 回归，比例风险模型，因子分析等。要点在于在设计阶段和数据采集阶段将可能的混杂因素收集，以便后续纳入分析，有效地控制混杂。
- **倾向性评分**：在比较不同的治疗方法对疾病的安全性和有效性评价的研究中，不同组别的患者往往没有可比性，比如采取手术治疗的患者和保守治疗的患者在疾病研究程度、年龄和经济负担能力等方面会有较大的不同，这些因素都是评价手术或保守治疗效果的重要的混杂因素，如果不能得到控制和调整，得到的结果往往是不准确的或者是错误的。为了同时控制多种不同混杂因素在两组人群分布的不平衡，可以采用使用倾向性评分的方法。倾向性评分是指在一定协变量条件下，一个观察对象可能接受某种处理（或暴露）因素的可能性。当研究涉及到的混杂因素过多时，如 20 - 30 个，进行匹配则变得不现实，严重影响了样本量，而全部进入统计模

型会因共线性等问题，使得统计模型无法正常估计效应。在观察性研究中，通过对倾向性评分的分层分析、匹配、回归分析，权重等方法，以期达到控制混杂的目的^[14]。

八. 证据等级评价

1. 正确认识 RWS 证据等级

循证医学金字塔证据分级（hierarchy of evidence）方法^[15]最早诞生于治疗研究领域，证据分级的概念提出来之初，将所有 RCT 的证据级别置于其他研究之上，由此导致普遍的观念认为 RWS 的证据等级或可信度低于 RCT 研究产生的证据。然而，不同的研究问题需要选择不同类型的研究证据，因为有些临床问题难以用 RCT 设计去回答。例如，针对糖尿病并发慢性肾炎的患者 10 年的心血管死亡率评估这一问题，最优的证据来源不是 RCT，而是高质量的队列研究。美国 FDA 于 2016 年 12 月在《新英格兰医学杂志》上发文指出^[3]，RWS 与其他证据的本质区别不在于研究方法和试验设计，而在于获取数据的环境。2017 年 8 月美国 FDA 发布《使用真实世界证据支持医疗器械注册审批指南》提出随着真实世界临床研究的增多，可以利用这些证据用来支持医疗器械的监管决策^[2]。可见 RWE 甚至可以作为药械审批的支持性证据内容，是循证医学体系中的有效补充。因此，把 RWS 的证据级别简单划分在金字塔证据分级法中的某个或某几个级别并不适合，也不意味着通过 RWS 所产生的证据等级一定低于 RCT 证据，两者往往是为了回答不同临床问题而产生的不同研究设计，在证据级别上不具备简单的可比性。

2. RWE 评价方法

随着医学研究领域的发展，证据评价体系也需要相应拓展以适应新的需求。对 RWS 的证据等级评价应该依据选取的研究设计与研究问题之间的相关性、研究质量控制程度，以及选取的研究数据的可靠性进行评价。本文借鉴循证医学金字塔证据分级方法^[16]，将常见的 RWS 类型按照证据等级的高低进行排序（图 2），同时建议研究者参照一系列影响证据等级的关键因素来客观评估真实世界证据的证据级别，例如，以下表格（表 4）中的关键点均是提高真实世界证据等级的常见关键因素^[17]，反之，则降低研究的证据级别。

图 2：常见真实世界研究的证据等级

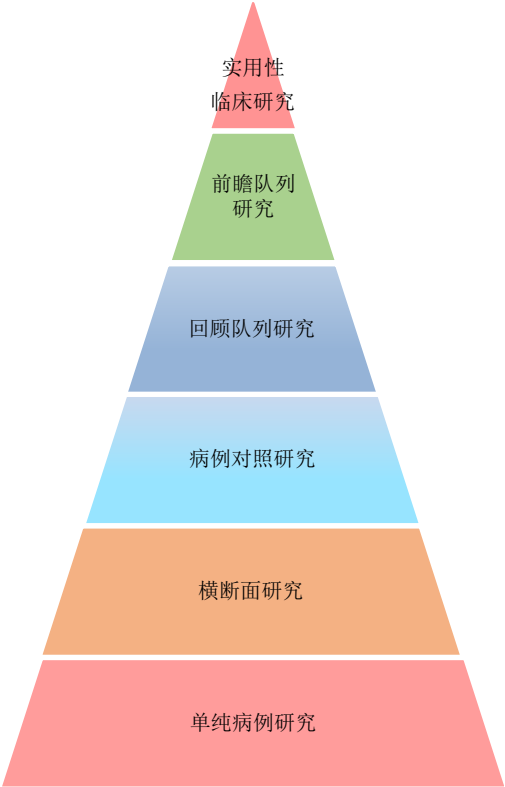


表 3：提高真实世界研究证据等级的关键因素

- 1 有效数据的样本量足够
 - 2 前瞻性设计，治疗/暴露因素和健康结局时间跨度合理
 - 3 研究人群的选择具有代表性
 - 4 有明确合理的纳入及排除标准，保证研究的内部有效性
 - 5 对治疗/暴露因素、健康结局和主要混杂因素评估准确
 - 6 控制和分析偏倚、混杂因素和数据缺失
 - 7 数据收集的完整性高
 - 8 数据来源可靠、准确
 - 9 有清晰的质量控制
 - 10 随访成功率高，对不完整数据或失访病例进行评估*
 - 11 使用统计分析方法适当
 - 12 对结果的分析客观可靠
 - 13 研究结论与研究问题相关性高
 - 14 横向比较既往同类研究
 - 15 研究结果得到既往作用机制和动物实验等证据支持
 - 16 罕见疾病研究**
- 注： *：针对前瞻性研究
**：针对单纯病例研究

九. 案例分析

“以人群为基础的甲状腺癌患者死亡概率的预测和评价研究”，是 2013 年发表在 Journal of Clinical Oncology 的一篇基于已有数据库的 RWS 文章^[18]。我们参考 STROBE 质量评估标准和指南流程中涉及的相关要素^[17]，对该研究进行了如下分析。

背景/理由	解释研究的科学背景，及报告特定研究的理由 介绍了影响甲状腺癌死亡的主要信息，包括发病和死亡数据，同时早期诊断甲状腺癌和报告无临床症状甲状腺癌等。突出长期生存的甲状腺癌患者可能死于其它原因。这些信息有助于了解和强调开展该研究的必要性。
目的	陈述具体目标，如果有特定研究假说，明确陈述 研究目标非常明确具体，即估计甲状腺癌患者死于甲状腺癌、非甲状腺的癌症、非癌症的机率，估计甲状腺癌的预后，为医生提供适用于计算和预测患者个体预后情况的工具。
研究环境和研究对象	描述研究环境、地点和时间点，包括入排研究对象、暴露因素和数据收集 该研究是以人群为基础的基于已有数据库的研究。上述关键因素都得到呈现，并通过图示的方式体现入排标准，明确展示获得最终研究人群的方法。
变量	清楚地定义所有的结局、暴露、预测因子、潜在的混杂因素及影响效应。如适用，给出诊断标准 该研究涉及的变量较为全面（主要包括诊断年龄、性别、种族、病理亚型、肿瘤侵犯、肿瘤大小、淋巴结累及放疗，主要变量缺失的患者将被排除在外。） 诊断标准采用的是病理学诊断金标准：病理诊断为甲状腺癌。 研究认为死因记录错误造成的潜在错分可能是死因分析的局限性之一，然而根据以往文献，总体灵敏度较高（87%）。
数据来源	对每一个兴趣变量，给出数据来源及评估（测量）的方法细节。如果研究变量不止一组，描述组别间评估方法的可比性 数据均来自美国 SEER 数据库，该数据库样本量大、有代表性、覆盖人群广、数据质量较好。介绍这些变量的同时，着重说明了暴露因素和结局变量，包括甲状腺癌、死亡、死亡原因等变量的收集和定义。
偏倚	描述应对潜在偏倚来源的措施 研究分析了人群选取的代表性、甲状腺癌的诊断的客观性等，在研究设计和变量测量阶段均考虑了潜在的偏倚的影响；同时，通过对竞争风险的分析，控制竞争风险带来的可能的偏倚。
样本量	解释研究样本量如何达成及研究样本量的合理性 该研究选取了 SEER 数据库中时间跨度为 15 年的 2.9 万例样本，该研究使用现有数据且样本量足够大且不增加额外负担，故未涉及样本量计算。
统计方法	(a) 描述所用的所有统计方法，包括控制混杂因素的方法

	主要使用 Fine & Gray 比例风险回归模型构建竞争风险模型。竞争风险模型使用基于贝叶斯信息准则的模型选择技术构建。为了减少偏倚，使用了外部队列校正的模型预测死亡概率和观察累积死亡发生率，所有的 p 值采用双边检验。值得注意的是，研究将人群分作两组，用于建立风险模型和验证风险模型，以验证结果的有效性。
	(b) 解释如何解决数据缺失 对于入排标准中的关键变量缺失、未知等病例，做排除处理，不纳入统计分析型。
主要结果	(a) 报告未校正的估计值，如果适用，给出混杂因素校正后的估计值及其精确度 (95%置信区间)。阐明按照哪些混杂因素进行了校正以及选择这些因素进行校正的原因。 研究建立比例风险回归模型，开展多因素分析，对混杂因素进行控制的同时也改善了模型的预测效率。
	(b) 如对连续变量进行分组，要报告每组观察值的范围 初次诊断年龄、自诊断以来的年数以 10 年为跨度分为 4 组，肿瘤大小自 <1.0cm 到 >4cm 被分为 4 组。本研究设定了随机分组后的验证队列，并以此评估了未校正和校正后的结果，用于评估模型的可信度，并对三种死因的分辨力在 95%置信区间进行评估，验证模型精确度。
局限性	讨论研究的局限性，包括潜在的偏倚或不准确的来源。讨论任何潜在偏倚的方向和大小。 指出选择比例风险回归模型的原因，与特定原因风险模型法的比较。由于非甲状腺癌的和其它癌种的重叠性，据此截尾可能会引起偏倚。 作者考虑死因的错分会影响结果的有效性，但是也强调依据之前的死亡患者尸检统计，肿瘤患者的死因记录相对较为准确（灵敏度 87%和阳性预测值为 85%）。
可推广性	讨论研究结果的普适性（外部验证性） 得益于足够大的样本量和代表性，并使用了验证队列，得出在 95%置信区间的准确度。有较好的外推性。

十. 新技术方案及未来展望

21 世纪是数据引导决策的时代。计算机技术支持的临床实践所产生的大量临床数据为进行科学研究奠定了基础。在海量数据背景下, 可基于现有数据库中的数据进行分析, 来支持不同种类的业务。随着机器学习的发展, 图像识别, 文本识别, 声音识别等新技术可以快速帮助临床数据结构化, 大数据分析, 云储存及云计算等技术的逐渐成熟以及区域链技术 (blockchain) 的提出和应用, 使多维度数据整合及一体化管理成为可能。这些都给 RWS 带来了前所未有的便利, 为 RWS 的开展和长期应用奠定了基础。

真实世界证据作为一个较新的概念, 还存在着一些有争议的内容, 其相关技术方法也还有待完善, 目前仅作为药品和医疗器械审批决策的补充证据。然而, 真实世界证据对于未来健康和医疗的推动不可低估。本指南仅针对 RWS 提出建议性思考, 在具体临床科研实践中, 研究人员可根据具体情况灵活参考和使用该指南。希望该指南能够帮助研究人员快速入门并解决临床实际问题。鉴于该指南为初版, 还有非常多不足和需要优化的空间, 非常期待大家的宝贵意见和反馈, 在此基础上我们将逐步完善, 在未来的 3 到 5 年时间里, 建立起完全符合中国实际需求的 RWS 指南体系。

附录 I 国外指南参考

近年来国际上多个临床研究组织和监管机构先后颁布了关于 RWS 的指南或工作指导意见，例如有 Good Pharmacoepidemiology Practice (GPP)，ISPOR RWD 指南，US FDA Device RWD 指南等，本文以表格的形式简要总结几个重要国外指南的要点。

表 4：国外参考指南概要

2015 年 6 月	规范药物流行病学研究指南 (Good Pharmacoepidemiology Practices , GPP ^[19])	1.国际药物流行病学学会 (The International Society for Pharmacoepidemiology , ISPE) 为帮助研究者解决药物流行病学研究的计划、实施与评估等问题发布的指导意见。 2.从方案制订, 责任、人员、设施、资源投入和合同各方, 研究实施 (包括受试者保护) , 信息发布, 不良事件报告和档案管理作出了阐述 3. 相比 2007 版 GPP , ①新版 GPP 的应用范围更广, 纳入了疗效比较研究 (comparative effectiveness research, CER) 和安全信号评估; ②重视研究方案注册, 对注册程序提出了相应要求; ③完善伦理诚信要求, 补充了使用二手数据来源开展研究的伦理学原则; ④规范不良事件报告原则, 在使用二手数据来源开展药物流行病学研究时, 研究者没有报告数据库中单个不良事件的义务, 当数据库包含医疗记录或叙述性文字记录时, 研究者也应报告其中涉及到的药品不良反应。
2016 年 12 月	美国国会《21 世纪治疗法案》 ^[20]	1. 定义 RWE “是除随机对照试验以外的其它来源获取的关于医疗产品使用、获益及风险相关的数据” 2.建议 RWE 在监管决策中的两项用途：用来支持上市后药物扩大适应证的批准；用来支持或满足上市后药物研究的相关需求。
2017 年 8 月	美国 FDA《使用真实世界证据支持医疗器械注册审批指南》 ^[2]	1.进一步诠释 RWD 和 RWE 的定义 1) 将 RWD 定义为从传统临床试验之外的来源收集的数据, 即从不同来源常规收集的与患者健康状况或者医疗服务相关的数据。 2) 将 RWE 定义为是基于 RWD 分析形成的与医疗产品使用、获益及风险相关的临床证据。 2. 提出对于 RWD 数据标准的要求：相关性(Relevance)和可靠性(Reliability)。 3.明确指出不应将指南过度理解为对现有 FDA 用于监管决策的证据评价标准做出任何的修改或变更；指出随着真实世界临床研究的增多, 可以利用这些证据用来支持监管决策, 为了缩小 RWE 距离可用于支持监管决策的证据之间的差距, 应减少 RWE 固有的偏倚潜质, 无论是回顾性或前瞻性采集的 RWD 均需良好的研究设计, 并且同传统临床试验一样, 应在获取和分析数据前制定研究方案及分析计划。
2017 年 9 月	ISPOR 真实世界证据工作 报告 ^[21]	1.认可真实世界数据的重要性和价值, 明确真实世界数据可以带来诸多获益 2.承认真实世界数据的局限性, 强调回顾性或前瞻性观察研究或数据库研究需要严格评估偏倚和混淆因素的来源, 并且进行相应的调整减少这些因素对研究结果的影响 3.建议依据研究问题 and 环境来评定真实世界研究的证据级别

附录 II 参考文献

1. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, Center for Biologics Evaluation and Research. Use of Real World Evidence to Support Regulatory Decision Making for Medical Devices. Guidance for industry and FDA staff. Document issued on August 31,2017.
2. 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]. 2017.10.08.
3. Sherman RE, Anderson SA, et al. Real-World Evidence - What Is It and What Can It Tell Us? N Engl J Med. 2016 Dec 8;375(23):2293-2297
4. https://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Clinical_Queries_Filters
5. 观察性疗效比较研究的方案制定:使用者指南 平装 -Priscilla Velentgas (作者), Nancy A.Dreyer (作者), Parivash Nourjah (作者), Scott R.Smith (作者), 等, 詹思延 (译者). 2014.8.1.
6. National Ethics Advisory Committee. Ethical Guidelines for Observational Studies: Observational Research, Audits and Related Activities. Revised edition. Wellington, New Zealand: National Ethics Advisory Committee; 2012.
7. U.S. Department of Health & Human Services. Health and Human Service (HHS) Subjects Regulations. 45 CFR 46.116[EB/OL];2009.
8. Sackett DL, Strauss SE, Richardson WS, et al. Evidence-based medicine: how to practice and teach EBM. Edinburgh: Churchill Livingstone, 2000.
9. 唐立等.实效性随机对照试验：真实世界研究的重要设计. 中国循证医学杂志，2017，17（9）：999 - 1004
10. Jones WS, Roe MT, Antman EM, et al. The changing landscape of randomized clinical trials in cardiovascular disease. J Am Coll Cardiol 2016; 68(17):1898-1907. pmid:27765193
11. Center of Drug Evaluation, CFDA. Technical Guidelines for Data Management in Clinical Trials (临床试验数据管理工作技术指南) [S]. 2012.
<http://www.cde.org.cn/news.do?method=largeInfo&id=312673>.
12. FDA. Guidance for Industry Computerized Systems Used in Clinical Trials [S].
<http://www.fda.gov/ohrms/dockets/98fr/04d-0440-gdl0002.pdf>.
13. EMA. Reflection Paper on Expectations for Electronic Source Data and Data Transcribed to Electronic Data Collection Tools in Clinical Trials [S]; 2010.

http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/08/WC500095754.pdf.

14. Rosenbaum, Paul R.; Rubin, Donald B. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*. 1983, 70 (1): 41–55.
15. Campbell DT , Stanley JC. Experimental and quasi-experimental designs for research[M]. Chicago , IL: Rand Mc- Nally College , 1963: P5 - 6.
16. <http://www.gradeworkinggroup.org/>
17. STROBE statement--checklist of items that should be included in reports of observational studies (STROBE initiative). et al. *Int J Public Health*.(2008)
18. Yang L, Shen W, Sakamoto N. Population-based study evaluating and predicting the probability of death resulting from thyroid cancer and other causes among patients with thyroid cancer[J]. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 2013, 31(4):468.
19. Public Policy Committee , International Society of Pharmaco- epidemiology. Guidelines for good pharmacoepidemiology practice (GPP) [J]. *Pharmacoepidemiol Drug Saf*, 2016, 25(1): 2-10
20. H.R.6 - 21st Century Cures Act. <https://www.congress.gov/bill/114th-congress/house-bill/6>
21. Garrison, Louis P. et al. Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report. *Value in Health*, 2017, Volume 10 , Issue 5 , 326 – 335