

真实世界观察性研究的质量评价工具 ArRoWS 解读



曹雪¹, 孟祥然¹, 王馨¹, 孙铭璘¹, 邓辉², 艾艳珂¹, 刘佳¹

1. 中国中医科学院中医临床基础医学研究所(北京 100700)
2. 陕西中医药大学针灸推拿学院(咸阳 712046)

【摘要】 真实世界观察性研究评估工具 (Assessment of Real World Observational Studies, ArRoWS) 是由英国莱斯特大学糖尿病研究中心莱斯特真实世界证据工作组研发的用于评估真实世界证据研究质量的工具, 具有较好的实用性。ArRoWS 能快速和具体地评估使用电子健康记录信息的真实世界证据研究的质量, ArRoWS 共包含 16 个条目, 其中 9 个条目为通用条目, 另外 7 个条目为与特定研究设计相关的条目。本文就 ArRoWS 的制订背景、制订过程、评价条目、评价标准和使用方法等进行介绍, 以为国内的真实世界研究者提供参考。

【关键词】 观察性研究; 真实世界证据; 评价工具; ArRoWS

Assessment tool ArRoWS for the quality of real world observational studies: an interpretation

CAO Xue¹, MENG Xiangran¹, WANG Xin¹, SUN Minglin¹, DENG Hui², AI Yanke¹, LIU Jia¹

1. Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, P. R. China
2. College of Acupuncture and Moxibustion, Shanxi University of Chinese Medicine, Xianyang 712046, P. R. China

Corresponding author: LIU Jia, Email: marie_liujia@sina.cn

【Abstract】 Assessment of Real World Observational Studies (ArRoWS) is a tool developed by the Leicester Real World Evidence (LRWE) Unit of the Diabetes Research Centre of the University of Leicester in the United Kingdom to assess the quality of real world evidence research, and has been reported to have good practicability. ArRoWS can be used to quickly and specifically assess the quality of real world evidence research that uses electronic health record information. The tool contains 16 items, nine of which are common items, and seven of which are related to specific research designs. The current study introduces the development background, development process, assessment items, assessment criteria, and application methods of ArRoWS and other related aspects, to provide references for real world researchers in China.

【Key words】 Observational studies; Real world evidence; Appraisal tool; ArRoWS

真实世界数据 (real world data, RWD) 是指来自真实医疗环境的研究数据, 反映实际诊疗及医疗服务过程和真实条件下的患者健康状况^[1-2], 其数据来源主要为基于临床或管理目的常规收集的健康医疗数据 (routinely collected health data, RCD) 和基于一定研究目的主动收集的健康医疗数据^[3], 前者包括医院电子病历数据 (electronic medical record, EMR)、健康/安全监测数据、医保数据及其

他健康数据等; 后者包括传统流行病学研究的数据^[3]。基于 RWD 开展的真实世界研究, 是 RWD 转化为真实世界证据 (real world evidence, RWE) 的重要桥梁^[4]。目前 RWD 和 RWE 日益得到医药卫生行业专家和学者的重视, 其主要用途包括以下几方面^[5-8]: ① 医药产品的评价和卫生政策制定 (如中成药有效性评价、药械监管和医保决策); ② 疾病临床研究 (如临床疗效、患者预后等); ③ 辅助管理和临床决策 (如药品上市后监管和临床指南制订); ④ 促进医疗质量。

目前真实世界研究中最广泛使用的研究类型是观察性研究^[9], 观察性研究包括利用真实世界数据和研究型数据开展的队列研究 (包括前瞻性、回

DOI: 10.7507/1672-2531.202206043

基金项目: 国家重点研发计划项目 (编号: 2019YFC1712200); 国际针灸临床研究服务标准研制项目 (编号: 2019YFC1712205); 中国中医科学院科技创新工程项目 (编号: CI2021A05408); 中国中医科学院重点项目 (编号: Z0733)
通信作者: 刘佳, Email: marie_liujia@sina.cn



顾性、双向性队列研究)、横断面研究、注册研究、病例-对照研究、病例系列和病例报告^[10]。国家药品监督管理局药品审评中心发布了《用于产生真实世界证据的真实世界数据指导原则(试行)》^[2],旨在促进基于医院信息系统,如电子健康记录(electronic health records, EHR)等数据库开展的真实世界观察性研究,为干预措施的结局评价提供更广泛的证据。有文献研究发现,观察性研究证据质量有待提高,主要原因包括:一些真实世界数据库由于结构或设计问题,可能不适合正在研究的人群;研究开展过程中研究设计不规范,如使用有偏差的抽样方法;缺乏对混杂因素的调整,或不适当的分析方法导致不正确或不可靠的结论;对于多源异构数据的治理和分析缺乏恰当的方法;对RWE的评价方法不统一等,这些问题影响了高质量RWD的产生和医疗人员对RWE合理地解释与使用,最终降低了医疗卫生决策效率。目前观察性研究的证据评价工具仍不够全面,而近年来真实世界的证据正蓬勃发展,因此亟需一个恰当的工具对真实世界研究的证据质量实施评估。本文通过对ArRoWS工具进行解读,为真实世界研究的证据质量评价提供参考。

1 制订背景

现有的评估真实世界观察性研究工具,多是对经典研究设计类型直接形成的真实世界证据展开评价。据报道,用于观察性研究的方法学质量评价工具目前可查到22种^[11-12],主要包括3种形式:一是量表式(scales),例如纽卡斯尔-渥太华量表(Newcastle-Ottawa Scale, NOS)^[13],是由多个评估研究质量和偏倚风险的项目组成,对每个项目进行评分,并综合得出一个总评分;二是清单式(checklists),例如英国牛津循证医学中心文献严格评价项目(Critical Appraisal Skill Program, CASP)^[14],是由多个评估研究质量和偏倚风险的项目组成,不对每个项目进行评分;三是条目式(items),例如Crombie条目^[11],由单个成分组成,与临床研究方法相关,结果可能会有偏差,如分配隐藏和盲法。但是由于不被广泛接受或适用性不强,其中大多数已不常使用或不被推荐使用。而且这些工具没有考虑到真实世界证据评价的特殊性,缺乏对电子代码的报告、数据库的质量评估及临床适用性评价。2015年开发的RECORD(reporting of studies conducted using routinely collected data)清单^[15],是用于评估使用日常常规医疗数据研究报告的完整性,但缺乏报告研

究的质量和临床相关性评价。所以,对于采用真实世界实践或医疗数据进行的观察性研究,目前还没有较为成熟且公认的评估工具,并且许多现有的工具不允许快速评估,要求评估人员有专门的知识 and 培训来使用该工具。2021年3月,来自英国莱斯特大学糖尿病研究中心和莱斯特真实世界证据工作组的临床流行病学专家们在美国《流行病学年鉴》上发表了题目为“Development, content validation, and reliability of the Assessment of Real-World Observational Studies (ArRoWS) critical appraisal tool”的专论,标志着ArRoWS的正式形成^[16],该团队在真实世界证据领域具有丰富的研究经验。该工作组还与英国国家卫生研究院合作开发了ERICAS网站(<https://www.ericas.org>)^[17],目的是提高心脏代谢医学中高质量真实世界证据的可用性和可见性,并使用ArRoWS工具来评估一项研究是否能产生高质量的真实世界证据,并将选定的研究发布到ERICAS网站上,有助于临床医生和研究人员将真实世界证据应用于临床决策。本文介绍的ArRoWS工具,是一种实用性较好且可快速评估真实世界研究质量的工具。

2 制订过程

最初的ArRoWS工具是在2007年开发,旨在评价观察性研究系统回顾的信息^[18],并未形成具体的条目。本次工具研发在此基础上更新了文献检索,纳入2006年至2018年3月新发表的文献,还纳入了灰色文献,共借鉴118个观察性研究的评估工具,通过文献综述确定了两个适用类型的领域:①所有研究设计的内容领域;②特定研究设计的内容领域(例如队列研究、病例-对照研究)。综合现有观察性研究评估工具的条目,形成初步的ArRoWS条目池,工作组再通过专家调查问卷的形式检验工具内容效度的可理解性、相关性和全面性,ArRoWS评价工具的单个条目水平的内容效度指数为0.64~1.00,总量表的内容效度指数为0.91,内容效度良好。工作组进一步以与心脏代谢医学相关的大样本观察性研究为例,由成对的评估人员独立使用ArRoWS进行评级,来检验工具的信度和实用性,使用校正Kappa值测定ArRoWS的重测信度和评价者间信度,评价者之间的一致性有较大差异(校正Kappa值为0.18~0.71)。工具研发历时15年,在前期研究积累基础上,形成了明确的评价内容,最终发布的16个条目中包含9个通用条目和7个与特定研究设计相关的条目。本文对其进

表 1 ArRoWS 质量评价工具条目

序号	条目内容	评级标准
通用条目		
1	研究问题或目的是否明确?	好/中等/差/不清楚
2	该研究样本是否能代表目标人群?	好/中等/差/不清楚
3	是否提供样本量、检验效能计算或不确定度的测量方法(例如:是否提供了置信区间或标准误差)?	是/否
4	暴露措施是否明确且合适?	好/中等/差/不清楚
5	结果是否明确且合适?	好/中等/差/不清楚
6	混杂因素是否清晰明确且合适?	好/中等/差/不清楚
7	统计分析是否清晰明确且合适?	好/中等/差/不清楚
8	该研究的局限性是否明确且合适?	好/中等/差/不清楚
9	作者是否从他们的研究结果中得出了适当的结论?	好/中等/差/不清楚
队列研究		
10	随访方法是否合适?	好/中等/差/不清楚
11	随访的时间是否足够长以确定结果?	好/中等/差/不清楚
12	如果作者正在测量治疗效果,那么该分析方式(例如匹配,倾向性评分,工具变量)是否合适?	好/中等/差/不清楚
病例-对照和比较效益研究		
13	作者是否解释了他们对病例和对照的选择?	好/中等/差/不清楚
14	如果是一个匹配的病例-对照研究,作者是否描述了他们的匹配标准?	好/中等/差/不清楚
15	如果是一个匹配的病例-对照研究,在分析中是否考虑了匹配?	好/中等/差/不清楚
电子数据库研究		
16	作者是否列出/引用的(从先前的文献中)代码集(ICD代码,读取代码)相关测试、程序、治疗和临床事件?	好/中等/差/不清楚

行详细介绍和解读,条目内容和评价结果见表 1。

3 条目解读

3.1 通用条目

3.1.1 条目 1: 研究问题或目的是否明确? 解读: 由于研究问题应该指导研究和引导 ArRoWS, 因此研究者必须明确地陈述研究问题或目的, 并有明确的理由。真实世界证据研究必须促进临床决策, 因此, 研究问题或目的必须具有重要的临床意义。

3.1.2 条目 2: 该研究样本是否能代表目标人群?

解读: 研究样本应该能够代表目标人群。因此, 必须明确研究目标人群的选择, 抽样方法必须适当, 以使样本能够代表目标人群。不具有代表性的研究样本可能会导致产生偏倚。研究者应充分解释抽样方法, 包括纳入与排除标准、招募方法和应答率。

3.1.3 条目 3: 是否提供样本量、检验效能计算或不确定度的测量方法(例如: 是否提供了置信区间或标准误差)? 解读: 在真实世界证据研究中, 存在着一定程度的不确定性。因此, 量化统计参数的不确定性很重要。此外, 必须计算检验效能来推断研究是否有统计学差异。研究者应该提供样本量、检验效能计算或不确定度的测量方法(例如置信区间或标准误差)。但由于研究类型不同, 这些措施可能不适用。可对不确定度的大小进行

讨论。这是唯一使用是/否来进行评级的条目。

3.1.4 条目 4: 暴露措施是否明确且合适? 解读: 准确的暴露特征是分析真实世界证据研究的一个关键因素。对暴露因素的评估有两种方式。首先, 必须明确定义暴露变量, 包括考虑到暴露过程中可能出现的错误分类和数据的可靠性(例如患者报告的数据通常不如由医疗保健专业人员直接收集的数据可靠)。其次, 如果暴露不明显, 作者应该证明该暴露因素对回答研究问题或目的的适当性。

3.1.5 条目 5: 结果是否明确且合适? 解读: 与条目 4 中的暴露类似, 结果的准确描述是分析真实世界证据研究的一个关键因素。结果需要能明确回答研究问题或目的。作者还应该评估是否遗漏了重要的结果。

3.1.6 条目 6: 混杂因素是否清晰明确且合适?

解读: 混杂因素可能会导致暴露和结局之间出现虚假关联, 应该通过统计方法进行控制。混杂因素应该是清晰明确、预先确定且合理的, 关于混杂因素的数据应该是可靠的。此外, 对不可观测的混杂因素也需重视。

3.1.7 条目 7: 统计分析是否清晰明确且合适?

解读: 其他研究人员可以使用相同的真实世界数据集来产生相同或类似的结果, 具有可重复性。因此, 作者应该提供一个全面的分析方法说明, 包括

如何处理缺失数据和失访。应确定统计分析方法是否合适且可以明确回答研究问题或目的。在大多数情况下,对结果的稳定性判断可通过敏感性分析方法来解决。

3.1.8 条目 8: 该研究的局限性是否明确且合适?

解读: 随机对照试验通常被认为是临床研究中评价因果效应的金标准^[19],与随机对照试验相比,真实世界证据研究本身存在一定程度的偏倚。在内容效度分析中,专家小组认为减小偏倚与评估真实世界证据研究的质量高度相关。在该研究的局限性范围内,应该考虑更广泛的偏倚。因此,本条目旨在评价作者是否充分承认并解释了真实世界证据研究的局限性及其对研究结果的影响,并考虑到如何减小偏倚。

3.1.9 条目 9: 作者是否从他们的研究结果中得出了适当的结论? 解读: 经过专家小组对内容效度进行分析后增加了该条目。重要的是要确保不能将关联推断为因果关系。此外,研究结果不应适用于目标人群之外的人群。

3.2 特定研究设计的条目

3.2.1 队列研究 ① 条目 1a: 随访方法是否合适? 解读: 随访方法必须是明确的,包括受试者失访的程度。与研究中保留的受试者相比,失访者不应该有额外的属性,因为这可能会混淆这种关联。② 条目 2a: 随访的时间是否足够长以确定结果? 解读: 对于真实世界证据研究,一个重要的考虑因素是确保随访的时间足够长以确定结果。因此,应事先确定随访时间,可通过先前的研究和/或生物学或临床知识来确定。③ 条目 3a: 如果作者正在测量治疗效果,那么该分析方式(例如匹配、倾向性评分、工具变量)是否合适? 解读: 经过专家小组对内容效度分析后,增加了该条目。由于大多数真实世界证据研究包括非随机选择的样本,基线特征的差异可能因治疗组而不同。因此,作者在测量治疗效果时,应该使用合适的分析方法来解释这一点。

3.2.2 病例-对照研究和比较效益研究 ① 条目 4a: 作者是否解释了他们对病例和对照的选择? 解读: 所选的对照组必须能代表目标人群,选择对照时应使选择性偏倚最小化,同时为研究提供足够的统计效能^[20]。病例和对照应来源于同一人群,病例的纳入与排除标准也必须是明确和合理的。② 条目 5a: 如果是一个匹配病例-对照研究,作者是否描述了他们的匹配标准? 解读: 匹配病例和对照可以减少混杂因素,并提高研究效率^[21]。还可以

控制难以量化的因素,如基于邻域级属性的匹配。作者应该充分解释匹配过程,并证明用于匹配的标准。③ 条目 6a: 如果是一个匹配病例-对照研究,在分析中是否考虑了匹配? 解读: 对于匹配病例-对照研究是否必须进行匹配分析存在一些争论^[22-24]。但是,在分析过程中,必须以某种方式控制匹配因素。

3.2.3 电子数据库研究 条目 7a: 作者是否列出/引用的(从先前的文献中)代码集(ICD 代码、读取代码)相关测试、程序、治疗和临床事件? 解读: 经过专家小组对内容效度分析后,增加了该条目。这个条目有两个目的,首先,应该充分提供所使用代码的细节,以便其他研究人员能够重复这项研究。其次,代码应该从相关引用的文献中预先确定。如果可能的话,作者应该提出所使用代码的敏感性和特异性。

4 评价标准

每个条目的评价选项有“好”、“中等”、“差”和“不清楚”。ArRoWS 研究工作组遴选出 7 个影响高质量真实世界证据产生的关键条目,分别为条目 1、2、4、5、6、7、8、9。如果作者未能完全说明某条目,则该条目应该被评为“中等”;如果作者根本没有解决这个条目,它应该被评为“差”或“不清楚”。根据这些回答,评价者应该为该研究提供一个总体评级和评价其产生真实世界证据的能力。一些研究可能会在一些条目上被评为“好”,但有一个关键条目不符合(例如大量数据的缺失、测量的不一致性或不确定的测量方法),这会严重影响研究产生高质量的真实世界证据的能力。因此,这项研究的总体分级可能只是“中等”。另一方面,研究可能有一些小的缺陷,但作者已经认识到并适当地控制了这些缺陷,那么这些研究可被评为“好”。若研究总体上被评为“质量较差”,并不一定意味着是质量较差的研究。这个分级方法仅仅意味着这项研究降低了产生真实世界证据的效用。被评为“良好”的研究将由临床医生进行进一步评估。建议不要将单个项目的评级合并得出一个综合等级。相反,评估者应该考虑对每个项目的不良评级对研究提供真实世界证据能力的潜在影响(表 2)。

5 报告模式

ArRoWS 评价结果推荐用表格的形式进行呈现。表 3 举例展示了 ArRoWS 的评价结果^[25-26]。如

表 2 真实世界研究总体等级

质量等级	评级标准
好	无关键条目不符合和非关键条目缺陷得到合适的控制：该研究提供了临床相关的、高质量的真实世界证据
中等	无关键条目不符合和非关键条目缺陷没有得到适当的控制：该研究提供了真实世界证据，但证据的效用有限
差	一个或多个关键条目不符合：该研究没有提供可靠的真实世界证据

表 3 ArRoWS 评估的表格模板展示

纳入研究	Hsu 2022 ^[25]	Schechter 2021 ^[26]
研究类型	电子数据库研究	队列研究
条目1	好	好
条目2	好	好
条目3	是	是
条目4	好	好
条目5	好	好
条目6	好	好
条目7	好	好
条目8	好	好
条目9	好	好
条目1a		好
条目2a		好
条目3a		好
条目4a		
条目5a		
条目6a		
条目7a	好	
质量总评级	好	好

果一个研究的样本量很小，则认为该研究样本无法代表其目标人群。除了条目3的评级为是或否外，所有其他条目都被评为良好、中等或差。每个条目的评级对研究提供真实世界证据的潜在能力需被单独考虑。当10个条目（9个核心条目和1个特定条目）中有8个条目被评为良好时，总体等级可被评为良好。

6 讨论

目前使用评估工具评价观察性研究质量的研究并不广泛，仅有少数研究探讨文献质量评价方法，如基于NOS量表对中国吸烟人群相关队列研究、对老年谵妄危险因素和预后队列研究、对中国人群的伤害预防的病例-对照研究和队列研究及中医药疗效评价队列研究评价文献质量的研究^[27-31]。并且以上都是针对传统观察性研究，目前探讨真实世界观察性研究文献质量的研究更少，如对真实世界研究中急性心力衰竭患者30天再住院率的系统评价^[32]，其中只纳入了队列研究，而对其他类型研究因缺乏统一偏倚风险评价工具，故未进行评价。目前常用的观察性研究的质量评价工具^[33]，包括NOS量表适用于评价病例-对照研究和队列研究，

CASP清单适用于队列研究与病例-对照研究，美国卫生保健质量和研究机构标准（Agency for Healthcare Research and Quality, AHRQ）推荐用于横断面研究评价，不适用于电子数据库研究。RECORD清单和GRACE（good research for comparative effectiveness）清单^[15]用于真实世界研究规范报告，但不能用于评价研究的质量。《中医药真实世界研究技术规范-证据质量评价与报告》^[34]该标准推荐评估真实世界证据的质量时可借鉴ArRoWS工具，ArRoWS的评价对象是真实世界研究，可适用于评价队列研究、病例-对照研究、比较效益研究和电子数据库研究，是一种通用的评价工具，运用的研究类型比较广泛，该工具的优势还在于当一个系统评价中纳入多种研究类型的时候，只需要用ArRoWS工具就可以评价不同纳入研究的质量，减少使用不同工具带来的偏差，省去了使用不同工具的繁琐过程，能为研究者评价文章的质量提供更多依据。

该工具提供了足够的灵活性，且易于使用，可以应用于许多类型的真实世界证据研究，从描述性研究到比较效益研究，以及基于电子病历数据开展的真实世界研究，可以使真实世界证据研究的评估标准化。经研发小组验证，ArRoWS工具有较好的内容效度，是一种值得推荐的方法学质量评价工具。但是，使用该工具报告时会遇到一些常见问题，包括代表性评估（医疗保险电子病历和自愿参与的数据库）、偏倚来源（持续时间和选择偏倚）和混杂因素的控制。关于ArRoWS的更广泛的适用性还有待进一步的验证，目前只有国外研究者使用该工具对在PubMed检索的真实世界证据研究进行了质量评价^[35]，并使用STaRT-RWE（Structured Template and Reporting Tool for RWE）模板作为报告真实世界证据研究实施情况的结构化模板^[35]。ArRoWS作为新发布的方法学质量评价工具，还需要时间和实践去不断完善，需要评价者在使用过程中提供宝贵的建议。评价人员之间的一致性有待进一步的提高，还需要进一步的指导和/或培训。

参考文献

- 1 孙鑫, 谭婧, 唐立, 等. 重新认识真实世界研究. 中国循证医学杂志

- 志, 2017, 17(2): 126-130.
- 2 国家药品监督管理局. 用于产生真实世界证据的真实世界数据指导原则(试行). Available at: <https://www.cde.org.cn/main/news/viewInfoCommon/2a1c437ed54e7b838a7e86f4ac21c539>.
 - 3 王雯, 谭婧, 任燕, 等. 重新认识真实世界数据研究: 更新与展望. 中国循证医学杂志, 2020, 20(11): 1241-1246.
 - 4 孙鑫, 谭婧, 王雯, 等. 建立真实世界数据与研究技术规范, 促进中国真实世界证据的生产与使用. 中国循证医学杂志, 2019, 19(7): 755-762.
 - 5 王雪, 乌兰其其格, 杨悦. 美国应用真实世界证据辅助医药产品监管决策的研究. 中国药事, 2020, 34(9): 1057-1062.
 - 6 王丽茹, 刘东辉, 邢杰, 等. 基于真实世界数据的结直肠癌预后特征识别. 国际遗传学杂志, 2021, 44(6): 401-408.
 - 7 孟若谷, 卓琳, 乔瑞, 等. 真实世界证据在上市后药品安全性监测与评价中的应用价值. 中国药物警戒, 2021, 18(7): 624-627.
 - 8 任静, 王永乐, 刘亭亭, 等. 双抗治疗轻型缺血性卒中或短暂性脑缺血发作——从临床指南到真实世界研究. 中国全科医学, 2021, 24(21): 2696-2700, 2706.
 - 9 明丹丹, 李军, 许璇, 等. 真实世界研究的方法及其在临床研究中的应用. 中国药房, 2018, 29(15): 2138-2143.
 - 10 中华中医药学会《中医药真实世界研究技术规范》制订组. 中医药真实世界研究技术规范——证据质量评价与报告. 中医杂志, 2022, 63(3): 293-300.
 - 11 Ma LL, Wang YY, Yang ZH, *et al.* Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better. *Mil Med Res*, 2020, 7(1): 7.
 - 12 Zeng X, Zhang Y, Kwong JS, *et al.* The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. *J Evid Based Med*, 2015, 8(1): 2-10.
 - 13 Wells G A, Shea B, O'Connell D. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available at: https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
 - 14 Critical Appraisal Skills Programme. CASP (case control study checklist). Available at: <https://casp-uk.net/casp-tools-checklists/>.
 - 15 廖星, 章轶立, 谢雁鸣. 真实世界研究标准: RECORD清单和GRACE清单的解读. 中国中药杂志, 2015, 40(24): 4734-4738.
 - 16 Coles B, Tyrer F, Hussein H, *et al.* Development, content validation, and reliability of the assessment of real-world observational studies (ArRoWS) critical appraisal tool. *Ann Epidemiol*, 2021, 55: 57-63.
 - 17 Coles B, Tyrer F, Khunti K. Improving quality in cardiometabolic medicine real-world evidence. *Prim Care Diabetes*, 2019, 13(2): 188.
 - 18 Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol*, 2007, 36(3): 666-676.
 - 19 陈耀龙, 李幼平, 杜亮, 等. 医学研究中证据分级和推荐强度的演进. 中国循证医学杂志, 2008, 8(2): 127-133.
 - 20 Waldram A, McKerr C, Gobin M, *et al.* Control selection methods in recent case-control studies conducted as part of infectious disease outbreaks. *Eur J Epidemiol*, 2015, 30(6): 465-471.
 - 21 Rose S, Laan MJ. Why match? Investigating matched case-control study designs with causal effect estimation. *Int J Biostat*, 2009, 5(1): Article 1.
 - 22 Pocock S, Kleinbaum D, Kupper L, *et al.* Epidemiologic research: principles and quantitative methods. *Biometrics*, 1983, 39(3): 822.
 - 23 Breslow NE, Day NE. Statistical methods in cancer research. Volume I - The analysis of case-control studies. IARC Sci Publ, 1980, (32): 5-338.
 - 24 Pearce N. Analysis of matched case-control studies. *BMJ*, 2016, 352: i969.
 - 25 Hsu CW, Tseng WT, Wang LJ, *et al.* Comparative effectiveness of antidepressants on geriatric depression: real-world evidence from a population-based study. *J Affect Disord*, 2022, 296: 609-615.
 - 26 Schechter M, Melzer-Cohen C, Rozenberg A, *et al.* Cardiorenal outcomes with sodium/glucose cotransporter-2 inhibitors in patients with type 2 diabetes and low kidney risk: real world evidence. *Cardiovasc Diabetol*, 2021, 20(1): 169.
 - 27 艾飞玲, 胡葵茹, 石钰霖, 等. 基于纽卡斯尔-渥太华量表对中国吸烟队列研究文献的质量评价. 中华疾病控制杂志, 2021, 25(6): 722-729.
 - 28 张蒙, 张雪梅, 杨璐, 等. 我国老年谵妄危险因素队列研究文献的质量评价. 护理研究, 2016, 30(33): 4206-4207.
 - 29 杨璐, 杨佳, 张泽梅, 等. 我国老年谵妄预后队列研究的文献质量评价. 护理研究, 2016, (1): 80-82.
 - 30 何琼, 黄渊秀, 康文婧, 等. 2001-2010年我国伤害预防病例对照研究及队列研究文献质量评价. 中华疾病控制杂志, 2014, 18(10): 913-916.
 - 31 王梅, 王建华, 张抗, 等. 中医药疗效评价队列研究的方法学质量评价. 中医杂志, 2016, 57(16): 1379-1383.
 - 32 张学成, 刘岩, 张晶晶, 等. 急性心力衰竭患者30天再住院率: 基于真实世界研究与随机对照试验比较的Meta分析. 中国循证医学杂志, 2021, 21(10): 1203-1210.
 - 33 曾宪涛, 刘慧, 陈曦, 等. Meta分析系列之四: 观察性研究的质量评价工具. 中国循证心血管医学杂志, 2012, 04(4): 297-299.
 - 34 中华中医药学会. 中医药真实世界研究技术规范证据质量评价与报告(试行). Available at: <https://www.cacm.org.cn/2021/06/30/13874/>.
 - 35 Khambholja K, Gehani M. Use of structured template and reporting tool for real-world evidence template for critical appraisal of the quality of reporting of real-world evidence studies: a systematic review. *Value Health*, 2022. DOI: 10.1016/j.jval.2022.09.003.

收稿日期: 2022-06-09 修回日期: 2022-11-25

本文编辑: 张洋