

## 基于真实世界数据的疾病管理研究

谭婧<sup>1,2</sup>, 刘兴会<sup>3</sup>, 孙鑫<sup>1,2</sup>

<sup>1</sup> 四川大学华西医院中国循证医学中心, 成都 610041

<sup>2</sup> 四川大学华西医院临床研究与循证评价研究室 (CREAT 团队), 成都 610041

<sup>3</sup> 四川大学华西第二医院产科, 成都 610041

通信作者: 孙鑫 电话: 028-85423040, E-mail: sunx79@hotmail.com

**【摘要】** 近年来, 随着信息技术快速发展和对循证临床实践和决策的需求急剧增加, 真实世界数据在医疗健康领域的作用日益受到重视。基于真实世界数据的疾病管理研究已成为重要发展方向, 其核心是针对疾病负担、诊断、临床预防和治疗、预后管理等多个方面的疾病管理问题, 在严谨的、预先设定的设计基础上, 基于海量、多源和高质量数据开展科学的数据分析和处理, 最终将基于这些研究证据形成疾病管理的指南、路径和技术规范, 从而提高医疗质量和安全。本文系统概述了基于真实世界数据开展疾病管理研究的一些重要问题。

**【关键词】** 真实世界数据; 疾病管理, 研究

**【中图分类号】** R-1 **【文献标志码】** A **【文章编号】** 1674-9081(2019)03-0284-05

**DOI:** 10.3969/j.issn.1674-9081.2019.03.017



开放科学(资源服务)标识码(OSID):

## Research on Disease Management Based on Real-world Data

TAN Jing<sup>1,2</sup>, LIU Xing-hui<sup>3</sup>, SUN Xin<sup>1,2</sup>

<sup>1</sup> Chinese Evidence-based Medicine Center, West China Hospital, Sichuan University, Chengdu 610041, China

<sup>2</sup> Clinical Research and Evaluation Unit (CREAT Group), West China Hospital, Sichuan University, Chengdu 610041, China

<sup>3</sup> Department of Obstetrics and Gynecology, West China Second University Hospital, Sichuan University, Chengdu 610041, China

Corresponding author: SUN Xin Tel: 86-28-85423040, E-mail: sunx79@hotmail.com

**【Abstract】** In recent years, with the rapid development of information technology and the increasing demand for evidence-based healthcare practice and policy decisions, the value of real-world data has received extensive attention. One of the most important applied areas of real-world data is disease management. Its core concepts lie in the application of rigorous and pre-specified designs to data of large volume, multiple sources, and high quality. By doing so, the high-quality real-world evidence for disease burden, diagnosis, prevention, treatment, and prognosis is generated to inform the development of guidelines, clinical pathways, and technical standards for disease management, which ultimately improves healthcare quality and safety. This paper systematically discussed important issues regarding the use of real-world data for research on disease management.

**【Key words】** real world data; disease management, research

*Med J PUMCH*, 2019,10(3):284-288

近年来,现代信息技术飞速发展,医疗部门对医疗质量、医疗资源使用和标准化的要求逐渐加强。同时,整个社会群体(如疾病高危人群、患者等)对健康也日益重视。这些重要而积极的变化,使快速实时收集、储存、归集和整理海量、多维度、多种来源的常规医疗和健康数据成为可能,这些医疗和健康数据常常被称为真实世界数据。国际上,尽管不同机构和组织对真实世界数据的确切定义存在差异,但实质均强调是在真实环境下采集的数据,数据的产生和收集过程与临床实际诊疗保持较好统一<sup>[1-4]</sup>。

目前,真实世界数据一部分是以管理为目的,在具体医疗和管理实践过程中自然形成的数据系统(如医院电子病历数据、医保数据、疾病防控数据、药品不良反应自发报告数据等);另一部分仍然是以研究为目的,在常规医疗环境下收集的数据(如患者登记、人群队列等)。

随着对循证临床实践和决策的需求不断加强,这些真实世界数据在医疗领域的作用日益受到关注和重视,尤其在疾病管理研究领域已成为重要的发展方向。基于真实世界数据的疾病管理研究的核心是在严谨的、预先设计的设计基础上,基于大量、高质量数据开展科学的数据分析和处理,以解决在疾病负担、诊断、临床预防和治疗、预后管理等多个方面的疾病管理科学问题,最终基于这些研究证据形成疾病管理的指南、路径和技术规范,从而提高医疗质量和安全。为了让更多临床医生了解该研究体系及其对临床实践和决策的作用,本文系统概述该研究领域的一些重要问题。

## 1 基于真实世界疾病管理研究拟解决的科学问题

疾病管理研究拟解决的问题通常来自于在疾病实际管理过程中遇到的问题,这与生物医学研究中集中解决机制和通路等科学问题存在差异。疾病管理研究的对象可以是患者个体,也可以是医疗机构,甚至是整个医疗服务系统。

对于个体疾病管理研究而言,所涉及内容通常可包括:(1) 确立疾病在特定人群中发生、发展和转归的过程;(2) 了解疾病负担,如了解中国人群2型糖尿病的患病率<sup>[5]</sup>、美国人群支气管扩张的疾病经济负担<sup>[6]</sup>;(3) 探索疾病发生的原因,如明确被动吸烟与慢性阻塞性肺疾病之间的关系<sup>[7]</sup>;(4) 探索新的筛查方法的价值,如探索最佳生物标志物组

合在肺癌筛查中的价值<sup>[8]</sup>;(5) 寻找最佳的预防和治疗措施,如评估高血压治疗策略的比较效果<sup>[9]</sup>;(6) 建立疾病预后管理策略,如分析携带乙型肝炎病毒孕妇的妊娠预后<sup>[10-11]</sup>,预测硬皮病-肺间质病患者3年死亡率<sup>[12]</sup>等。其中,了解疾病负担、优化预防和治疗策略、明确患者预后是患者管理研究中最常见的科学问题。

## 2 不同来源真实世界数据在疾病管理研究中的特征和优缺点

真实世界数据可分为基于常规收集的数据和主动收集的数据<sup>[4]</sup>。前者包括已有的以常规管理为目的产生的医院电子病历数据、区域医疗健康数据和医保数据库等<sup>[13]</sup>;后者以患者登记数据为代表,还包括基于实效性随机对照试验的研究数据<sup>[14-15]</sup>,至少部分数据需要主动收集。不同来源的真实世界数据在疾病管理研究中各有优缺点(表1)。

总的来讲,基于常规收集的真实世界数据是基于管理目的而产生,开展疾病管理研究时样本量较大、耗时短、经费需求少是主要优势,不过多种与医疗管理、数据管理相关的原因都可能显著影响数据质量,以致出现难以完全满足研究需求的情况。此外,验证不同来源数据的数据质量也是基于常规收集数据开展疾病管理研究的重要步骤。主动收集的真实世界数据,由于具有严格的设计和预先安排,在数据质量、变量范围、因果论证的强度方面可能具有优势,但由于其耗时、耗力、经费需求大,需要研究者具备更高的研究设计和管理能力。

尽管这些不同来源的数据对疾病管理研究均具有意义,但患者登记有可能是其中使用最广泛、数据的全面性和质量最能满足开展系统性疾病管理研究要求的数据来源。患者登记是通过主动收集数据的过程(至少包含部分主动收集的数据),从而建立患者数据库的过程<sup>[15]</sup>。根据研究纳入对象的不同,可以分为基于特定疾病的患者登记(如心肌梗死患者登记研究<sup>[16]</sup>)、基于特定暴露的患者登记(如孕妇人群患者登记<sup>[17]</sup>、服用某种新型抗凝药物登记<sup>[18]</sup>)和基于特定诊疗过程的患者登记[如实施经皮冠状动脉介入治疗(percutaneous transluminal coronary intervention, PCI)患者登记<sup>[19]</sup>]。

无论哪种类型,患者登记在设计和实施中需要满足下列基本条件:在临床“自然”状态下收集数据;有预先设计,使得数据收集具有明确的目的性(不一

表 1 不同来源真实世界数据在疾病管理研究中的优势和不足

典型数据来源	优势	不足	适宜的研究领域	备注
基于常规收集的真实世界数据				
医院电子病历数据	诊疗过程数据相对最完善；医嘱信息完整，能获得院内用药数据；实验室检查数据准确性高；能获得影像学数据	无院外诊疗数据；通常无法获得患者生活行为习惯数据；无院外随访数据；数据缺失可能较多；非结构化数据需要转化和验证	疾病负担评估；防治策略评估；诊断价值评估；预后管理策略	疾病诊断准确性与医疗机构及医生诊断水平密切相关
区域性医疗数据库	能获得患者在多家医疗机构的就诊数据；能获得患者既往患病和就诊信息	医疗机构间数据质量有差异；患者链接可能不完善；药品编码可能不完善；数据缺失可能较多	疾病负担（区域内人群）；病因探索；预后管理策略；部分疾病自然史	数据完整性、结构化程度与区域化数据库的成熟度密切相关
医保数据库	人群代表性好；费用数据准确性高	疾病诊断可能不准确；缺乏诊疗过程数据；无患者自报数据	疾病负担评估；部分病因探索；部分疾病自然史	往往只能获得患者主要疾病诊断，且缺乏多种混杂变量数据
主动收集的真实世界数据				
患者登记	数据准确性较高；人群代表性较好；能获得患者自报数据；能够获得长期随访数据	资源需求高、研究时间长、协调管理难度相对更大	疾病自然史；疾病负担评估；病因探索；防治策略评估；预后管理策略	数据质量与研究设计、执行情况密切相关
基于实效性试验的数据	随机过程能控制多种无法观测的混杂因素	资源需求高、研究时间长、协调管理难度相对更大	防治策略评估	往往一个研究仅针对一个确切的疾病防治问题

定是明确的研究目的，可能是基于某种管理目的开展的登记)；采用统一的指标定义和数据收集方式建立标准化、统一的患者数据库；至少有一部分数据是主动收集。因此，患者登记为开展疾病管理领域的相关研究提供了良好的数据来源。

在疾病自然史研究中，患者登记通过前瞻性随访，能获得研究对象疾病发展及过程的详细记录。由于对患者采用相同的测量和数据收集方式，在疾病诊断的准确性方面显著高于其他数据来源。在疾病负担研究中，比如开展基于罕见病的患者登记，能在一定程度上估计罕见病在人群中的患病率，为医疗管理决策提供依据。在病因研究中，可以整合多种其他来源数据，比如环境暴露、生活行为习惯等，探索多种可能的治病危险因素。在防治策略评估中，除直接利用观察性数据评估特定治疗措施的疗效和安全性外，还可基于患者登记人群开展实效性随机对照试验，如基于国家综合冠状动脉造影和血管成形术注册数据库开展实效性随机对照试验，探索 ST 段抬高型心梗患者行 PCI 前常规进行血栓抽吸是否能减少短期死亡率（血栓抽吸+PCI 与仅 PCI 比较），以获得更高质量的疾病防治证据。此外，患者登记是预后研究的优势领域，通过对患者开展前瞻性随访，能获得患者自报结局、长期终点结局等数据，是开展预后预测研究必要的数据库资源。

然而，由于患者登记需要主动收集数据，涉及前瞻性随访，因此，需要的人力、物力和财力要高于常

规收集的数据库研究。在研究开始前，研究者需充分权衡研究目的、拥有资源、人员配备和目标人群。大多数研究者希望通过建立特定疾病的患者登记以回答或解决在该疾病管理中的多个问题。但过多增加研究内容和扩大目标人群的范围可能带来相反结果——由于研究资源不足和协调的负担增加，数据准确率降低、失访率和数据缺失增加可在很大程度上削弱患者登记数据库的价值。

### 3 基于真实世界数据的生殖围产管理研究实践：经验与体会

近年来，笔者所在的中国循证医学中心临床研究、评价与转化（clinical research, evaluation and translation, CREAT）研究团队以围产生殖领域为研究对象，利用患者登记、医院电子病历数据、前瞻性调查等多源数据，针对围产和生殖领域中的疾病管理科学问题，从孕前因素的考虑到分娩结局的全局视图，整合应用多种大数据和人工智能技术，开展系列研究，建立了疾病全周期研究管理模式（图 1）。现展现 3 个研究实例。

#### 案例 1：前瞻性调查研究

与临床专家团队合作，于 2016 年在全国 6 大区 24 家医院开展了针对 12 000 余名孕妇铁缺乏和缺铁性贫血的多中心横断面调查，发现中国孕妇经实验室检查诊断的铁缺乏（48.2%）和缺铁性贫血（13.9%）

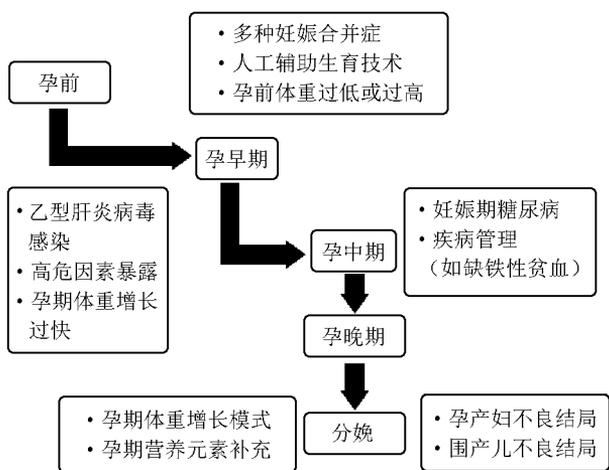


图 1 基于真实世界数据的围产领域全周期研究管理模式

患病率较高，远高于临床医生的常规认识。同时，笔者团队分析了孕前体质量指数（body mass index, BMI）和孕期体重增长与孕妇缺铁性贫血的相关性<sup>[20]</sup>，采用多水平模型考察不同中心是否存在聚集效应，并考察孕前 BMI 和不同孕期是否存在交互作用。研究发现，孕前低体重和孕期体重增长过快的孕妇有更高的铁缺乏和缺铁性贫血风险。这项研究的意义在于：缺铁性贫血是一种影响孕产妇和新生儿健康的重要妊娠期并发症，然而对我国孕妇铁缺乏和缺铁性贫血的状况尚不清楚（最近一次全国调查开展于 2004 年）。随着我国孕产妇孕期生活水平普遍较高，甚至有专家认为缺铁可能不再是孕产妇管理的关键问题。该项研究结果明确指出了目前孕期铁缺乏的现状，也为制定本土化的临床指南和孕期管理提供了有意义的决策支持。

### 案例 2：登记数据库研究

基于 2009 至 2010 年四川省 8 家医院 33 000 多名多中心孕妇登记数据库，笔者团队开展了系列研究：（1）分析了我国西部地区孕妇严重不良结局的患病率，并建立了针对多种合并症下的孕产妇严重不良结局预测模型<sup>[21]</sup>；（2）针对我国乙型肝炎高流行的现状，探索孕妇携带乙型肝炎病毒是否有更高的孕产妇不良结局和围产儿不良结局风险<sup>[10-11]</sup>。研究过程中进一步发现，既往研究对孕妇携带乙型肝炎病毒是否增加妊娠期糖尿病的风险存在争议。针对该问题，研究团队进一步基于全球已发表的真实世界数据，使用预后 Meta 分析技术，合成真实世界证据，最终证实了孕妇携带乙型肝炎病毒显著增加妊娠期糖尿病风险的结论<sup>[22]</sup>。（3）人工辅助生育技术已经成为解决不孕不育的主要手段，在孕产妇中的使用比例持续增加。然而，不论

从管理层面，还是研究层面都希望了解人工辅助生育技术是否会增加孕产妇严重不良结局的风险。现有数篇研究报告了两者之间的相关性，但笔者团队发现上述研究对多胎妊娠在两者中的作用认识不足，因此专门调查多胎妊娠是否是人工辅助生育技术与孕产妇不良结局的中介效应（待发表），为采用人工辅助生育技术的人群选择合适的胚胎移植个数提供决策依据。

### 案例 3：基于医院电子病历数据库的研究

笔者从华西第二医院电子病历数据库中回顾性地提取了 2013 年 1 月 1 日至 2014 年 12 月 31 日期间分娩的单胎妊娠数据。在 10 000 多名孕产妇（110 000 余次体重重复测量）中筛选出 4567 名（47 699 次体重重复测量）未发生任何预设不良事件的孕产妇，根据世界卫生组织亚洲人群 BMI 分类标准，采用多水平样条模型拟合了不同孕前 BMI 分类孕期体重增长的模式。该研究初步建立了针对我国孕产妇不同孕前 BMI 分类的孕期体重增长范围和适宜增长速度，同时发现对我国孕妇人群的增长情况与美国医学研究所（Institute of Medicine, IOM）的推荐存在重要差异<sup>[23]</sup>。该研究的意义在于：我国目前尚无针对中国孕妇人群孕期体重增长的推荐意见；临床多采用美国 IOM 的孕期体重推荐意见。由于美国的指南基于高加索人群，并不适用于其他种族人群，尤其我国孕妇身高、体重等指标与高加索人群差异明显。从孕产妇管理角度来讲，使用 IOM 的标准可能并不能实现最优。此研究证实了该管理科学假设，并强调了建立适用于中国孕妇人群体重增长的重要性。

## 4 小结

真实世界数据在解决疾病管理科学问题方面具有巨大价值。根据不同研究目的，选择最合适的高质量数据资源，进行严格的设计和分析可获得高质量研究证据，解决具体的管理科学问题。但同时，基于真实世界数据开展疾病管理研究面临重要挑战，尤其是如何获得高质量数据方面。不同来源的数据存在不同的优势和不足；即使同类的真实世界数据，数据质量仍可能存在显著差异。无论对于基于常规还是主动收集的的真实世界数据，如何保证数据完整性和质量是核心基础。此外，还需要通过预先设定的、合理的设计以及严格的数据分析，才能确保高质量数据变成证据。

目前，随着大数据技术的发展，AI 技术日益用于真实世界数据的应用，从而辅助开展基于真实世界数据的疾病管理研究。这些技术包括了实现文本数据

向结构化数据的转换、在预测模型中使用机器学习技术等,但这些技术的适用性仍需进行进一步验证。充分认识临床团队和方法学团队各自的优势,开展核心技术和数据资源的整合是成功的关键,搭建从提出科学问题,到整合数据资源,再到实现和应用关键技术的完善技术链,可为促进真实世界数据向疾病管理决策转化起到有益推动作用。

## 参 考 文 献

- [1] 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 Health*, 2017, 20: 858-865.
- [2] Resnic FS, Matheny ME. Medical Devices in the Real World [J]. *N Engl J Med*, 2018, 378: 595-597.
- [3] U. S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, et al. Draft Guidance for Industry and Food and Drug Administration Staff-Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices, 2016 [EB/OL]. <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf>.
- [4] 孙鑫,谭婧,唐立,等.重新认识真实世界研究[J].*中国循证医学杂志*, 2017, 17: 126-130.
- [5] Yang W, Lu J, Weng J, et al. Prevalence of diabetes among men and women in China [J]. *N Engl J Med*, 2010, 362: 1090-1101.
- [6] Seitz AE, Olivier KN, Steiner CA, et al. Trends and burden of bronchiectasis-associated hospitalizations in the United States, 1993-2006 [J]. *Chest*, 2010, 138: 944-949.
- [7] Yin P, Jiang CQ, Cheng KK, et al. Passive smoking exposure and risk of COPD among adults in China: the Guangzhou Biobank Cohort Study [J]. *Lancet*, 2007, 370: 751-757.
- [8] Jung YJ, Katilias E, Ostroff RM, et al. Development of a Protein Biomarker Panel to Detect Non-Small-Cell Lung Cancer in Korea [J]. *Clin Lung Cancer*, 2017, 18: e99-e107.
- [9] Moura CS, Daskalopoulou SS, Levesque LE, et al. Comparison of the Effect of Thiazide Diuretics and Other Antihypertensive Drugs on Central Blood Pressure: Cross-Sectional Analysis Among Nondiabetic Patients [J]. *J Clin Hypertens (Greenwich)*, 2015, 17: 848-854.
- [10] Tan J, Liu X, Mao X, et al. HBsAg positivity during pregnancy and adverse maternal outcomes: a retrospective cohort analysis [J]. *J Viral Hepat*, 2016, 23: 812-819.
- [11] Tan J, Huang S, He G, et al. Maternal hepatitis B surface antigen carrier status and its impact on neonatal outcomes: a cohort study of 21947 singleton newborns in China [J]. *J Matern Fetal Neonatal Med*, 2017, 30: 2219-2224.
- [12] Morisset J, Vittinghoff E, Elicker BM, et al. Mortality Risk Prediction in Scleroderma-Related Interstitial Lung Disease: The SADL Model [J]. *Chest*, 2017, 152: 999-1007.
- [13] 王雯,刘艳梅,谭婧,等.回顾性数据库研究的概念、策划与研究数据库构建[J].*中国循证医学杂志*, 2018, 18: 1-7.
- [14] 唐立,康德英,喻佳洁,等.实效性随机对照试验:真实世界研究的重要设计[J].*中国循证医学杂志*, 2017; 17: 999-1004.
- [15] 谭婧,程亮亮,王雯,等.患者登记研究的策划与患者登记数据库构建:基于观察性设计的真实世界研究[J].*中国循证医学杂志*, 2017, 17: 1365-1372.
- [16] Hofmann R, James SK, Jernberg T, et al. Oxygen Therapy in Suspected Acute Myocardial Infarction [J]. *N Engl J Med*, 2017, 377: 1240-1249.
- [17] Cohen LS, Viguera AC, McInerney KA, et al. Reproductive Safety of Second-Generation Antipsychotics: Current Data From the Massachusetts General Hospital National Pregnancy Registry for Atypical Antipsychotics [J]. *Am J Psychiatry*, 2016, 173: 263-270.
- [18] Yang X, Li Z, Zhao X, et al. Use of Warfarin at Discharge Among Acute Ischemic Stroke Patients With Nonvalvular Atrial Fibrillation in China [J]. *Stroke*, 2016, 47: 464-470.
- [19] Weipert KF, Bauer T, Nef HM, et al. Use and outcome of thrombus aspiration in patients with primary PCI for acute ST-elevation myocardial infarction: results from the multinational Euro Heart Survey PCI Registry [J]. *Heart Vessels*, 2016, 31: 1438-1445.
- [20] Tan J, Qi YN, He GL, et al. Association between Maternal Weight Indicators and Iron Deficiency Anemia during Pregnancy: A Cohort Study [J]. *Chin Med J (Engl)*, 2018, 131: 2566.
- [21] Tan J, Liu XH, Yu C, et al. Effects of medical co-morbidities on severe maternal morbidities in China: a multicenter clinic register study [J]. *Acta Obstet Gynecol Scand*, 2015, 94: 861-868.
- [22] Tan J, Mao X, Zhang G, et al. Hepatitis B surface antigen positivity during pregnancy and risk of gestational diabetes mellitus: A systematic review and meta-analysis [J]. *J Viral Hepat*, 2018, 25: 1372-1383.
- [23] Tan J, Ren Y, Qi Y, et al. The pattern of gestational weight gains among Chinese women: a repeated measure analysis [J]. *Sci Rep*, 2018, 8: 15865.

(收稿日期: 2018-11-05)