

电子病历数据质量评估指标体系构建研究^{*}

杨 柳¹ 李小龙² 李善平³ 吴义熔⁴

(¹ 北京师范大学政府管理学院 北京 100875 ² 三峡大学经济与管理学院 宜昌 443002

³ 三峡大学附属仁和医院 宜昌 443001 ⁴ 北京师范大学人文和社会科学高等研究院 珠海 519087)

〔摘要〕 目的/意义 采用科学系统的方法构建电子病历数据质量指标体系,助力提升电子病历数据质量,为相关政策改进提供借鉴。方法/过程 借助批判性评价技能方案的系统检查清单进行文献综述分析,结合专家意见构建数据质量评估指标体系,使用问卷调查和统计分析验证,采用熵权法和模糊模型相结合的方法综合评价。结果/结论 构建一套包含 16 项指标的电子病历数据质量评估指标体系,对问卷调查结果采用模糊综合评价得到该指标体系综合得分 90.445 分,隶属度等级为“非常重要”(隶属度为 62.5%),说明本研究所采用的数据质量指标符合电子病历使用者的关注重点。

〔关键词〕 电子病历;数据质量评估;指标体系;模糊综合评价

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2023.10.005

Study on the Construction of an Electronic Medical Record Data Quality Assessment Index System

YANG Liu¹, LI Xiaolong², LI Shanping³, WU Yirong⁴

¹School of Government, Beijing Normal University, Beijing 100875, China; ²College of Economics & Management, China Three Gorges University, Yichang 443002, China; ³Affiliated Renhe Hospital of China Three Gorges University, Yichang 443001, China; ⁴Institute of Advanced Studies in Humanities and Social Sciences, Beijing Normal University, Zhuhai 519087, China

〔Abstract〕 **Purpose/Significance** A scientific and systematic method is adopted to construct an electronic medical record (EMR) data quality index system to help improve the quality of EMR data and provide references for the improvement of relevant policies. **Method/Process** Literature review is carried out by using the critical appraisal skills programme system review checklist. The data quality assessment index system is constructed by leveraging literature review in combination with experts' opinions. The questionnaire survey and statistical analysis are used for verification, and the entropy weight method and fuzzy model are combined for comprehensive evaluation. **Result/Conclusion** The data quality assessment index system including 16 indicators is constructed in this study. The comprehensive score of the system based on the questionnaire result is 90.445 points, and the membership level is “very important” (the membership score is 62.5%), indicating that the system is in line with the focus of EMR users.

〔Keywords〕 electronic medical record (EMR); data quality assessment; index system; fuzzy comprehensive evaluation

1 引言

〔修回日期〕 2023-03-06

〔作者简介〕 杨柳,博士研究生,发表论文 3 篇;通信作者:吴义熔,教授,博士生导师。

〔基金项目〕 国家社会科学基金项目(项目编号:20BTQ066)。

随着大数据、物联网和人工智能等技术在医疗健康领域应用日渐深入,数字健康已成为当前社会关注的热点。电子病历数据作为医疗健康大数据最

主要来源^[1]，受到广泛重视。2018 年《关于促进“互联网+医疗健康”发展的意见》明确提出要健全统一规范的全国医疗健康数据资源目录与标准体系^[2]。2020 年《公立医院内部控制管理办法》明确提出应加强医院信息数据质量管理制度的建立^[3]。2021 年《关于推动公立医院高质量发展的意见》强调推进智慧医院建设和医院信息标准化建设^[4]。电子病历数据广泛应用，数量显著增加，但其质量还不尽理想，其质量问题成为广泛关注并亟待解决的热点问题。

2 相关研究

一套系统而科学的数据质量评估指标是提高电子病历数据质量的前提。近年来，学者们以不同方法对数据质量评估指标展开研究。袁莎等^[5]从原始质量、过程质量、结果质量 3 个维度出发，构建包含 17 个二级指标和 53 个三级指标的医疗数据质量评估指标体系；Weiskopf N G 等^[6]通过综述电子病历数据质量评估方法的研究文献，将 27 项指标纳入数据质量的 5 个维度；Ehsani - Moghaddam B 等^[7]从准确性、精度、完整性和一致性等维度，讨论加拿大初级保健数据的质量问题；Fu S 等^[8]从模型开发、模型部署和模型实施 3 个阶段出发，考虑数据质量的完整性、可变性和及时性 3 个维度。国家层面，2018 年国家卫生健康委员会颁布《电子病历系统应用水平分级评价标准（试行）》，要求各级医疗机构对其电子病历系统应用水平进行评级，并将一致性、完整性、整合性和及时性作为其数据质

量的评价指标^[9]。

电子病历数据质量评估指标体系的研究已取得一定进展，但仍存在不尽如人意之处。主要体现在两个方面：一是指标甄选及构建过程缺乏系统性和科学性，多数根据评估者的经验^[10]，缺乏量化指标^[11]；二是运用场景选择具有局限性，很多研究以电子病历数据的二次使用为目的，忽视了辅助临床护理和医疗管理等使用电子病历系统的最初目标^[12]，导致电子病历数据质量评估指标缺乏全面性。

本研究立足于临床使用、医疗数据管理和二次使用等目的，重点集中在评估指标的遴选和评价主体整体对其重要性的验证上，通过系统开展文献综述分析，结合专家建议，构建电子病历数据质量指标体系。使用问卷调查验证专家经验结果，采用定量与定性相结合的方法分析各指标特性以及综合评价指标体系，弥补既往仅基于专家经验构建指标体系的不足^[13]。

3 资料来源与方法

3.1 资料来源

在 Web of Science 数据库中搜寻基础性研究资料，补充专家推荐文献，形成与研究目的高度相关文献的资料库，作为开展电子病历数据质量评估指标体系构建的基础。

3.2 研究方法

3.2.1 文献分析（图 1）

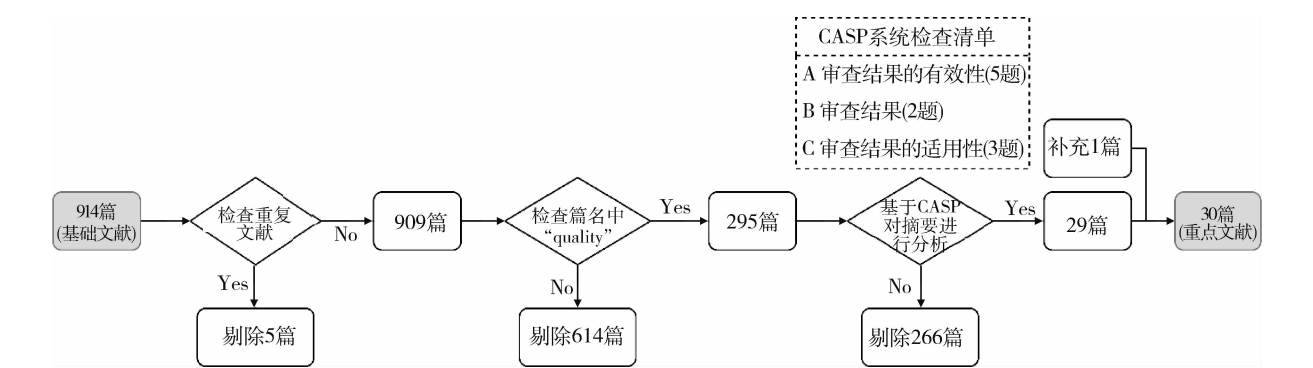


图 1 基于批判性评价技能方案的文献综述流程

以“data quality”和“electronic health record *”或“electronic medical record *”为搜索主题，将在 Web of Science 数据库中检索出的 914 篇文献作为基础资料，去除 5 篇重复文献，对剩余文献以篇名包含“quality”进行筛选，共得到 295 篇。随后，借鉴批判性评价技能方案（critical appraisal skill program, CASP）的系统评价检查清单^[14]展开综述分析，确定重点研究文献 29 篇，补充 1 篇国家标准性文献，形成包含 30 篇重点文献的资料库^[6-8,12,15-40]。

3.2.2 指标提炼及问卷调查 从 30 篇重点文献中，列出文献所含指标，对其术语汉化及标准化，统计出现频次，最后结合专家组意见，提炼出一套电子病历数据质量评估指标，以此为基础设计“电子病历数据质量调查表”。采用方便抽样、滚雪球抽样和随机抽样方法，选取湖北省两家三级甲等医院和一所研究型高校，进行问卷调查。

3.2.3 统计分析 共回收 195 份有效问卷，采用

SPSS 26.0 和 SPSSAU 进行描述性统计以及推论统计分析，包括人口统计学分析、一致性分析、结构效度分析等。

3.2.4 综合评价 使用熵权法确定权重，结合模糊模型综合评价。在开展模糊模型综合评价过程中，评价集是评价者对评价对象可能做出的各种结果组成的集合，表示为 $\{v_1, v_2, \cdots, v_5\}$ ，其中， v_1 为非常不重要， v_2 为不重要， v_3 为一般， v_4 为重要， v_5 为非常重要，将其分别赋值为 $\{55, 65, 75, 85, 95\}$ 。这种定量与定性相结合的方法能提升“模糊”的客观性，使评价结果更加科学^[13]。

4 结果

4.1 电子病历数据质量评估指标

构建一套包含 16 项指标的电子病历数据质量评估指标体系，见表 1。

表 1 电子病历数据质量评估指标体系

指标	注释
准确性 ^[6-7,12,15-20,22,24-25,27-29,31-40]	记录的患者信息与患者真实情况相一致
完整性 ^[6-8,12,15-34,37-40]	记录的患者信息详细完备
时效性 ^[6-8,12,15-20,24-25,27-29,32,35-38,40]	系统中记录的患者状况及时有效
一致性 ^[6-7,12,15-19,22-24,27-29,31-33,36-37,39-40]	数据的内外部达到一定程度的一致
精细性 ^[7,32,37,40]	数据定性或定量达到一定的精度
规范性 ^[15,17,19,21,23,25-26,30,35-37]	数据以标准化格式存储、处理和流通
唯一性 ^[7,22,28,33,39-40]	数据记录无重复性；有些数据必须保持唯一
可信性 ^[15,17-18,25,28-29,32,35,37,39]	数据来源于专业性的机构；数据常被审查
合理性 ^[6,12,15,19-21,26-27,30,32,34]	数据取值是否合理等
可追溯性 ^[16,22,28,40]	保证电子病历数据访问及变更踪迹的可审计性
可移植性 ^[40]	数据能被存储、替换或转移，并保持已有质量
可用性 ^[7,19,32,35-37,39-40]	数据可使用的程度应符合数据管理者声称的水平
可访问性 ^[7,15-16,18,32,34,36-40]	数据是否可以很容易地被访问和提取，界面友好
关联性 ^[7,18-19,25,32,36-37]	数据之间存在某种应有的关联
适用性 ^[7,19,22,28]	所记录的内容适用于患者的健康管理和疾病诊治
可理解性 ^[16,18-19,34-40]	数据的格式及内容易于理解

4.2 统计分析及结构效度

4.2.1 人口统计学 被调查者中，女性占比（57.9%）略高于男性（42.1%）；医生（47.2%）

和护士（34.9%）明显多于临床研究人员（11.3%）和管理人员（5.6%）；年龄 41~50 岁（28.7%）的最多，其次是 28~34 岁（26.7%），28 岁以下和 35~40 岁占比相当（20%、19.5%）；

临床工作 15 ~ 25 年占比 (26.2%) 最高, 其次是 5 年以下 (25.1%); 使用电子病历 5 ~ 10 年的占比最多 (47.2%), 其次是 5 年以下 (32.8%), 其中有超过 74% 的被试者对电子病历使用熟练。被调查者平均年龄为 36.74 岁, 平均临床工作 13.73 年, 平均电子病历使用 7.92 年。可见被调查者思想成熟、临床经验丰富且熟悉电子病历系统, 调查结果能够反映电子病历使用者的真实体验, 具备一定的专业性和代表性。

4.2.2 一致性 Cronbach's α 系数为 0.980, Kendall's W 系数为 0.730 ($\chi^2 = 2\,266.213$, $P < 0.05$)。说明本问卷具有非常高的内部一致性, 且参与调查的 195 名被试者具有较强的评分一致性, 问卷答案一致性高, 数据可信度高。

4.2.3 结构效度 采用主成分分析法探索性因子分析: 取样适当性 (Kaiser - Meyer - Olkin, KMO) 值为 0.959, Bartlett's 球性检验的显著性 $P = 0.000$, 说明非常适合因子分析; 提取 1 个公因子数, 总方差解释率为 77.112%, 载荷 0.819 ~ 0.930, 说明本研究提炼的 16 项指标, 能够从整体上作为数据质量评估指标, 形成的数据质量评估体系具有较高的结构效度, 且结构稳定、内容准确有效。

4.3 模糊综合评价

利用熵权法得到的权重集对该指标体系模糊综合评价, 使用加权平均型 $M(\wedge, +)$ 算子和最大隶属度法则分析, 得到综合得分 90.445 分, 隶属度等级为“非常重要” (隶属度为 62.5%), 判定本研究构建的数据质量指标体系对数据质量评估非常重要, 表明该体系指标符合电子病历数据使用者的关注重点。

5 结语

本研究基于 CASP 系统评价检查清单展开文献分析, 结合专家意见, 构建一套包含 16 项指标的电子病历数据质量指标体系, 问卷调查结果表明, 该电子病历数据质量评估指标体系得到电子病历使

用者的认可。

本研究的创新点主要包括 3 方面。研究方法方面, 开展系统性文献综述、问卷调查及量化分析等, 将定量与定性研究相结合的方法是对既往研究缺乏定量分析的有益拓展, 选取的指标与电子病历使用者对电子病历数据质量评估指标的关注重点相吻合。应用方面, 重点评估指标的确立将有效地对电子病历数据的产生、存储、使用、二次使用、报废等全生命周期管理提供指引, 达到提升电子病历数据质量的重要目标。政策促进方面, 为与电子病历数据质量相关国家政策的改进提供新视角, 可为扩充《电子病历系统应用水平分级评价标准 (试行)》数据质量评价指标提供参考。

研究局限主要在于抽样限制。一是调查地区缺乏医疗信息化水平欠发达地区, 未来可采取分层抽样的方式, 对中国不同经济及信息化水平地区展开调查, 为指标体系的优化获取更全面的样本。二是调查群体以电子病历系统使用者为主, 未来将纳入电子病历数据质量评估实践专家等第三方的批判性意见, 为形成更具权威性的指标体系提供多视角支撑。未来将开展不同类型主体对电子病历数据质量的价值取向分析, 进一步验证本研究得到指标体系的可靠性。

参考文献

- 1 戴明锋, 孟群. 医疗健康大数据挖掘和分析面临的机遇与挑战 [J]. 中国卫生信息管理, 2017, 14 (2): 126 - 130.
- 2 国务院办公厅. 关于促进“互联网+医疗健康”发展的意见 [EB/OL]. [2022 - 10 - 13]. http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm.
- 3 国家卫生健康委员会, 国家中医药管理局. 公立医院内部控制管理办法 [EB/OL]. [2022 - 10 - 26]. <http://www.nhc.gov.cn/caiwusi/s7785t/202101/1b6d1b06b398486082263107bb074801.shtml>.
- 4 国务院办公厅. 关于推动公立医院高质量发展的意见 [EB/OL]. [2022 - 10 - 26]. http://www.gov.cn/zhengce/content/2021-06/04/content_5615473.htm.
- 5 袁莎, 沈丽宁. 医疗数据质量评价指标体系研究 [J]. 卫生软科学, 2020, 34 (12): 12 - 16.
- 6 WEISKOPF N G, WENG C. Methods and dimensions of electronic health record data quality assessment: enabling re-

- use for clinical research [J]. Journal of the American medical informatics association, 2013, 20 (1): 144–151.
- 7 EHSANI – MOGHADDAM B, MARTIN K, QUEENAN J A. Data quality in healthcare: a report of practical experience with the Canadian primary care sentinel surveillance network data [J]. Health information management journal, 2021, 50 (1/2): 88–92.
- 8 FU S, WEN A, SCHAEFERLE G M, et al. Assessment of data quality variability across two EHR systems through a case study of post – surgical complications [C]. Washington, DC: AMIA 2022 Annual Symposium, 2022.
- 9 国家卫生健康委员会. 电子病历系统应用水平分级评价标准 (试行) [EB/OL]. [2022 – 10 – 18]. http://www.gov.cn/xinwen/2018-12/09/content_5347261.htm.
- 10 叶琪, 赵亮, 阮彤, 等. 区域卫生数据用于临床疗效分析的可用性研究 [J]. 大数据, 2018, 4 (3): 13–23.
- 11 梁萍, 张智如. PDCA 循环用于电子护理病历质量控制的效果研究 [J]. 中国医药指南, 2019, 17 (11): 289–290.
- 12 DIAZ – GARELLI J F, BERNSTAM E V, LEE M, et al. Datagauge: a practical process for systematically designing and implementing quality assessments of repurposed clinical data [J]. eGEMs (Generating evidence & methods to improve patient outcomes), 2019, 7 (1): 32.
- 13 刘燕, 梁钻容. 基于熵权 – 模糊综合评价法的医学检验专业学生的人文素养评价指标体系研究 [J]. 中华临床实验室管理电子杂志, 2021, 9 (3): 182–186.
- 14 CASP. CASP checklist [EB/OL]. [2022 – 03 – 24]. <https://casp-uk.net/casp-tools-checklists/>.
- 15 ITURRY M D, ALVES – SOUZA S N, ITO M, et al. Data quality in health records: a literature review [C]. Chaves: 16th Iberian Conference on Information Systems and Technologies (CISTI 2021), 2021.
- 16 董冲亚, 姚晨, 高嵩, 等. 加强医院临床研究源数据管理, 提高我国临床研究数据质量 [J]. 中国循证医学杂志, 2019, 19 (11): 1255–1261.
- 17 MAKELENI N, CILLIERS L. Critical success factors to improve data quality of electronic medical records in public healthcare institutions [J]. South African journal of information management, 2021, 23 (1): 1–8.
- 18 BADR N G. Guidelines for health it addressing the quality of data in EHR information systems [C]. Prague: 12th International Conference on Health Informatics, 2019.
- 19 BIAN J, LYU T, LOIACONO A, et al. Assessing the practice of data quality evaluation in a national clinical data research network through a systematic scoping review in the era of real – world data [J]. Journal of the American medical informatics association, 2020, 27 (12): 1999–2010.
- 20 TERRY A L, STEWART M, CEJIC S, et al. A basic model for assessing primary health care electronic medical record data quality [J]. BMC medical informatics and decision making, 2019, 19 (1): 30.
- 21 KAMDJE – WABO G, GRADINGER T, LÖBE M, et al. Towards structured data quality assessment in the German medical informatics initiative: initial approach in the MII demonstrator study [M]. Amsterdam: IOS Press, 2019.
- 22 GARCÍA D – L – C R, SÁEZ C, MUÑOZ – SOLER V, et al. Construction of quality – assured infant feeding process of care data repositories: definition and design (Part 1) [EB/OL]. [2023 – 03 – 05]. <https://doi.org/10.1016/j.compbimed.2015.09.024>.
- 23 JETLEY G, ZHANG H. Electronic health records in is research: quality issues, essential thresholds and remedial actions [J]. Decision support systems, 2019, 126 (1): 113137.
- 24 JOHNSON S G, SPEEDIE S, SIMON G, et al. Application of an ontology for characterizing data quality for a secondary use of EHR data [J]. Applied clinical informatics, 2016, 7 (1): 69–88.
- 25 DUNGEY S, BELOFF N, PURI S, et al. A pragmatic approach for measuring data quality in primary care databases [C]. Valencia: IEEE – EMBS International Conference on Biomedical and Health Informatics (BHI), 2014.
- 26 LEE K, WEISKOPF N, PATHAK J. A framework for data quality assessment in clinical research datasets [C]. Washington, DC: AMIA 2017 Annual Symposium, 2017.
- 27 WEISKOPF N G, BAKKEN S, HRIPCSAK G, et al. A data quality assessment guideline for electronic health record data reuse [J]. eGEMs (Generating evidence & methods to improve patient outcomes), 2017, 5 (1): 14.
- 28 SÁEZ C, MARTÍNEZ – MIRANDA J, ROBLES M, et al. Organizing data quality assessment of shifting biomedical data [M]. Amsterdam: IOS Press, 2012.
- 29 FEDER S L. Data quality in electronic health records research: quality domains and assessment methods [J]. Western journal of nursing research, 2018, 40 (5): 753–766.
- 30 KAHN M G, CALLAHAN T J, BARNARD J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data [J]. eGEMs (Generating evidence & methods to improve patient outcomes), 2016, 4 (1): 1244.

(下转第 38 页)

础。医院只有读懂标准、理解标准，才能更好地发挥标准“指挥棒”作用，促进医院信息化高质量发展，为健康中国建设发挥积极作用。

参考文献

- 1 刘晓强, 华永良, 薛成兵. 我国医院信息化发展历程浅析 [J]. 中国卫生信息管理杂志, 2016, 13 (2): 142 - 152.
- 2 郑涛, 范晨. 互联互通标准化成熟度测评在医院信息化建设中的作用 [J]. 中国卫生信息管理杂志, 2018, 15 (5): 588 - 591, 606.
- 3 原卫生部. 电子病历系统功能应用水平分级评价方法及标准 (试行) [EB/OL]. [2022 - 11 - 04]. <http://www.nhc.gov.cn/zyygj/s3593/201512/a335bd9490164b928f27276b0161daa9.shtml>.
- 4 原国家卫生和计划生育委员会. 国家医疗健康信息区域 (医院) 信息互联互通标准化成熟度测评方案 (2017 年版) [EB/OL]. [2022 - 09 - 05]. <http://www.nhc.gov.cn/mohwsbwstjxxzx/s8553/201709/06ac0b5228674ce487405de8e939d40b.shtml>.
- 5 国家卫生健康委员会. 全国医院信息化建设标准与规范 (试行) [EB/OL]. [2022 - 04 - 13]. <http://www.nhc.gov.cn/ewebeditor/uploadfile/2018/04/20180413162542120.pdf>.
- 6 国家卫生健康委员会. 关于印发电子病历系统应用水平分级评价管理办法 (试行) 及评价标准 (试行) 的通知 [EB/OL]. [2022 - 12 - 07]. <http://www.nhc.gov.cn/zyygj/s7659/201812/3cae6834a65d48e9bfd783f3c7d54745.shtml>.
- 7 国家卫生健康委员会. 关于印发《医院智慧服务分级评估标准体系 (试行)》的通知 [EB/OL]. [2022 - 03 - 18]. <http://www.nhc.gov.cn/zyygj/s3593g/201903/9fd8590dc00f4feeb66d70e3972ede84.shtml>.
- 8 国家卫生健康委员会. 医院智慧管理分级评估标准体系 (试行) [EB/OL]. [2022 - 03 - 15]. <http://www.nhc.gov.cn/zyygj/s3594q/202103/10ec6aca99ec47428d2841a110448de3.shtml>.
- 9 国家卫生健康委员会. 医院信息互联互通标准化成熟度测评方案 (2020 年版) [EB/OL]. [2022 - 08 - 06]. <http://www.nhc.gov.cn/mohwsbwstjxxzx/s8553/202008/e80dafa1334c44c38f644602406a4973.shtml>.
- 31 BOTSIS T, HARTVIGSEN G, CHEN F, et al. Secondary use of EHR: data quality issues and informatics opportunities [EB/OL]. [2023 - 03 - 05]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041534/>.
- 32 JOHNSON S G, SPEEDIE S, SIMON G, et al. A data quality ontology for the secondary use of EHR data [EB/OL]. [2023 - 03 - 05]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4765682/>.
- 33 AERTS H, KALRA D, SÁEZ C, et al. Quality of hospital electronic health record (EHR) data based on the international consortium for health outcomes measurement (ICHOM) in heart failure: pilot data quality assessment study [J]. JMIR medical informatics, 2021, 9 (8): e27842.
- 34 CHO S, WENG C, KAHN M G, et al. Identifying data quality dimensions for person - generated wearable device data: multi - method study [J]. JMIR mhealth and uhealth, 2021, 9 (12): e31618.
- 35 HINDS A, LIX L M, SMITH M, et al. Quality of administrative health databases in Canada: a scoping review [J]. Canadian journal of public health - revue canadienne de sante publique, 2016, 107 (1): e56 - e61.
- 36 LABERGE M, SHACHAK A. Developing a tool to assess the quality of socio - demographic data in community health centres [J]. Applied clinical informatics, 2013, 4 (1): 1 - 11.
- 37 CHEN H, HAILEY D, WANG N, et al. A review of data quality assessment methods for public health information systems [J]. International journal of environmental research and public health, 2014, 11 (5): 5170 - 5207.
- 38 FARZANDIPOUR M, KARAMI M, ARBABI M, et al. Quality of patient information in emergency department [J]. International journal of health care quality assurance, 2019, 32 (1): 108 - 119.
- 39 ÁLVAREZ S R, BERISTAIN I A, EPELDE U G, et al. TAQIH, a tool for tabular data quality assessment and improvement in the context of health data [EB/OL]. [2023 - 03 - 05]. <https://doi.org/10.1016/j.cmpb.2018.12.029>.
- 40 国家药品监督管理局. 人工智能医疗器械质量要求和评价第 2 部分: 数据集通用要求 [EB/OL]. [2021 - 09 - 30]. <https://www.nifdc.org.cn/directory/web/nifdc/infoAttach/38b8b027-8b43-43c1-b744-c0ac891b5ec8.pdf>.

(上接第 32 页)