

临床数据质量评估指标研究*

田琪 陈雅妮 韩喆僖 吕旭东

(浙江大学生物医学工程与仪器科学学院 杭州 310007)

[摘要] 基于质量评估相关文献和临床数据应用中的数据问题报告,对现有评估知识进行分解、合并和归纳,构建对临床评估问题进行唯一表述的临床数据评估指标,对每项指标所适用的评估手段进行描述。对比以往研究并与国家相关标准进行验证,结果表明本研究提出的指标能更清晰、准确地对应实际需求,有利于促进临床数据质量评估标准化。

[关键词] 数据质量评估;数据问题报告;质量维度;自动化;质量评估指标

[中图分类号] R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2020.10.002

Study on the Clinical Data Quality Assessment Indicators TIAN Qi, CHEN Yani, HAN Zhexi, LV Xudong, College of Biomedical Engineering and Instruments Science, Zhejiang University, Hangzhou 310007, China

[Abstract] Based on the relevant literature of quality assessment and data error reports of clinical data application, the existing assessment knowledge is decomposed, combined and summarized to construct the clinical data assessment indicators that uniquely expressed the clinical assessment problems, and the assessment methods applicable to each indicator are expounded. By comparing with previous studies and verify with relevant national standards, the results show that the indicators proposed in this study can more clearly and accurately correspond to the actual needs, which is conducive to promoting the standardization of clinical data quality assessment.

[Keywords] data quality assessment; data error report; quality dimensions; automatic; quality assessment indicators

1 引言

临床数据是记录临床信息的载体。数据能够准确、完整地反映所记录的临床事实,是保证数据有效利用的重要前提^[1]。临床数据的质量问题会严重影响临床工作正常开展,对患者生命安全和社会经

济发展造成负担^[2]。美国国家卫生信息协调办公室(Office of the National Coordinator for Health Information Technology, ONCHIT)报告中指出可预防医疗错误在美国每年导致超过4万人死亡,患者数据不匹配、诊断数据缺失等数据质量问题是造成医疗错误的重要原因之一。患者数据不匹配也是35%医疗保险拒赔理由,大约每年造成医院120万美元的损失^[3-4]。临床数据质量评估是通过科学手段发现临床数据中质量问题的过程,是了解和改善临床数据质量的重要环节^[5]。研究者基于质量评估结果判定数据是否可用,或针对暴露的数据质量问题采取相应改善措施,提高数据质量^[6]。错误或无效的评估结果可能导致数据应用事故,威胁生命财产安全。数据质量评估流程标准化有利于保证评估结果准确

[修回日期] 2020-10-13

[作者简介] 田琪,博士;通讯作者:吕旭东,教授,博士生导师。

[基金项目] 国家重点研发计划资助项目“冠状动脉粥样硬化病变早期识别和风险预警的影像学评价体系研究”(项目编号:2016YFC1300300)。

性与有效性,然而实现标准化质量评估面临巨大挑战。一方面,虽然多数学者对“临床数据质量是一个包含完整性、合理性等多维度的复杂概念”这一观点已经达成共识^[7],但是每种维度代表哪些临床数据问题仍未达成一致。例如评估人A认为的完整性代表数据项和数据值不能缺失,而评估人B理解的完整性则可能是数据总量不能缺失。由于评估人对于质量维度的理解不一致,两人对数据质量的衡量也就不一致,导致各自评估结果互相毫无意义。另外不同研究者对质量维度的认知不一致会造成评估人没有全面理解维度包含的数据质量问题,导致评估结果无法全面反映数据质量问题。另一方面,对于每种评估问题应当采取哪种评估手段取决于评估人的经验和实施水平。大多数研究都根据评估人经验选择评估方法,方法适宜程度和实施效果因人而异^[7],难以保证质量评估的标准化。因此为促进临床数据质量评估的标准化,首先需要对临床数据质量有准确、完善的认知,其次减少评估流程对人工的依赖,提升质量评估的自动化程度^[8]。

2 相关研究

Weiskopf 等人在 2011 年通过文献综述,依据各维度在文献中的出现频次,得到 5 种高度关注的的质量维度:完整性、合理性、准确性、一致性、时间性^[7]。Kahn 等人在 2016 年通过文献综述与专家讨论方式,提出包含完整性、合理性、一致性 3 种维度的质量评估框架。该研究中 Kahn 对每种维度都定义了若干评估指标。评估指标是对质量维度相关质量问题的详细描述^[9]。Johnson 等人也基于文献构建一套临床数据质量的本体。该本体中包含完整性、一致性、时间性、合理性 4 个质量维度,每个维度同样定义了若干指标^[8]。Johnson 描述质量评估中质量维度、实施手段等概念之间的关系及相关属性,初步描述评估指标自动化执行的思路^[8]。目前研究中存在一些不足:第一,当前研究者在构建指标时考虑并不全面。例如 Kahn 并未将与需求相关的质量特性纳入指标体系^[9],Johnson 则没有考虑数据重复问题^[8];第二,目前指标并未实现对一

类数据质量问题的唯一定义,研究者仍会对相同指标有不同理解。第三,现有研究主要基于文献知识来构建指标。虽然文献是相关知识的可靠来源,但是指标构建缺少临床实际数据作为基础,降低指标可信度。第四,Johnson 在本体中虽然描述了指标和评估方法之间的关系,但是缺少每项指标与具体实施方法之间的关联。针对以上问题,本研究通过结合文献知识与实际临床数据问题,构建相对全面的对质量评估问题进行唯一描述的数据质量评估指标,促进领域对临床数据质量理解的一致性。此外本研究依据的指标来源对每项指标所适用的评估方法进行描述,构建评估指标与具体评估方法之间的关系,为评估指标的自动化评估研究提供支持。

3 研究方法

3.1 资料收集

3.1.1 文献检索与整理 通过对国内外文献数据库进行检索来收集相关文献。对于英文文献,选择 PubMed 作为检索数据库以提高临床相关文献的检索相关性。以关键字" data quality ", " quality of data " 对标题和摘要进行检索,时间限定为 1998 - 2019 年,共得到 802 篇文献,通过两轮筛选剔除不相关文献。首先,基于文献标题和摘要,剔除研究对象不是电子病历或医院信息系统相关数据或临床研究所收集数据的文献。对无法确定的文献进行全文阅读。本轮筛选排除 267 篇非相关文献。剩余 535 篇文献,基于标题和摘要保留同时满足以下两个条件的文献:研究内容含有数据质量维度或质量评估框架;作者对质量维度、特性提出个人观点。第 2 轮筛选最终排除 509 篇文献。对剩余的 20 篇文献^[6-25]基于引文进行关联查询,对得到的文献采用同样流程进行筛选,最终再纳入 3 篇相关文献^[26-28],累计得到 23 篇相关英文文献。对于中文文献,选择万方、维普和中国知网 3 个文献库。检索条件为主题中关键字为“数据质量”且摘要中包含“临床数据”,时间限定为 1998 - 2019 年。去除重复后共得到 103 篇相关文献。阅读文献标题、摘要,筛选其中与临床数据质量维度或特性相关的文

章。对无法确定的文章进行全文阅读。经过筛选,得到4篇相关中文文献^[29-32]。对其进行关联查询和筛选,发现两篇相关中文文献^[33,34]以及1篇相关英文文献^[35]。综上最终得到24篇英文文献、6篇中文文献。

3.1.2 基于文献的语料集整理 为收集文献中碎片化数据质量知识,对纳入的30篇文献进行全文阅读,摘录其中质量维度的相关描述。摘录遵循以下3条准则以确保内容相关:文献中关于数据质量维度的定义;对维度的解释说明;所列举的实际数据质量问题示例。最终得到包含247个段落或句子的语料集。

3.1.3 问题数据报告收集 在实际临床数据应用中,数据使用者会记录数据中的问题,形成数据问题报告,以便提高数据质量。数据问题报告中包含丰富的数据质量知识,为构建指标提供可用资源。本研究选择区域卫生信息平台构建和临床科研表单数据采集两类临床数据应用场景,收集该场景数据问题报告。构建区域化卫生信息平台是我国基础卫生建设的重要任务,国内各省市均开展相关建设^[36];注册研究是针对性地收集特定诊断或病情数据并开展相关研究工作的一种常见临床研究手段^[37]。针对以上两类常见的应用场景,分别选择两个典型案例进行数据问题收集。案例1:哈尔滨市2020年4-9月间,13家医疗机构产生25份数据问题报告,总计1810条数据问题。每份报告涉及1家医疗机构,报告中的每条数据问题代表该医疗机构1张数据库表中所发现的一类数据质量问题。案例2:山西省某三甲医院2020年5-8月间,面向区域卫生健康信息平台构建的19份数据问题报告,总计1372条数据问题。每份报告涉及该医院一类临床业务数据库表的数据质量问题,其中每条问题代表1张数据库表中的一类数据质量问题。案例3:

国内某冠心病患者注册数据集。该研究数据由16家医疗机构共同参与收集,从2016年至今累积收集超过33000例冠心病患者临床、随访数据。本研究收集该研究2019年6月-2020年1月期间6份数据问题报告,总计207条数据质量问题。案例4:国内某老年糖尿病患者注册研究数据集。该研究数据由全国141家医疗机构共同参与收集,2015-2017年间累积收集2806例老年糖尿病患者(年龄>65岁)诊疗和随访数据。收集该研究2016年7月-2017年1月期间的8份数据问题报告,总计323条数据质量问题。案例1代表在区域卫生信息平台构建中多家不同医疗机构所发现的数据质量问题;案例2则代表同一家机构在长期多次数据递交中所发现的数据质量问题。两个案例结合能有效代表该场景中遇到的数据质量问题,具有典型性。案例3和案例4则是多中心、跨地区、大规模的患者注册研究,具有典型性。

3.2 指标属性设计

为使指标清晰、唯一描述数据质量问题,对指标进行结构化描述十分必要。Johnson曾在研究中提出一套临床数据质量本体,对质量评估相关概念进行定义^[6]。我国卫健委发布的电子病历数据评估相关标准中也对指标进行一定描述^[38-39]。Johnson本体和我国相关标准中出现的11个与质量评估相关的概念,见表1。本研究选择其中9个概念作为指标属性。其中问题描述和评估对象两个属性是区分指标的两个重要属性,分别代表指标所描述的临床问题类型以及指标所适用的数据对象类别。没有选择数据集、评估结果两个概念作为指标属性,因为本研究针对临床数据提出评估指标,不面向其他领域数据集;评估结果也不是描述评估过程的属性。

表1 指标属性以及对应文献依据

属性名称	说明	文献来源依据	是否选择
代码	指标的代码	[7, 39, 40]	√
名称	指标的名称	[7, 39, 40]	√
问题描述	指标的解释和说明	[7, 39, 40]	√

续表 1

数据业务类别	指标所适用数据的临床业务类别,按照评级标准中的分类,包括:病房医师、病房护士、门诊医师、检查科室、检验处理、治疗信息处理、医疗保障、病历管理、电子病历基础和利用	[7]	√
评估对象	指标所适用的数据类别,按照文献中的分类包括:元数据、数据元、数据值、数据记录、多数据值、多数据记录	[7]	√
评估方式	指标适用的实现方式	[7]	√
需求来源	评估的需求来源	[7]	√
任务场景	指标所适用的临床任务场景。“通用”代表适用于一般任务;“科研”代表一般在科研分析任务中才会涉及的指标	[7]	√
计算方法	指标的计算方法	[7]	√
数据集	待评估数据集的说明	[7]	-
评估结果	质量评估的结果	[7]	-

3.3 指标归纳

3.3.1 确定热点维度 临床数据质量维度是用来描述临床数据质量特性的专用概念。统计文献中不同维度出现的次数得到当前关注高的质量维度^[8]。统计所得文献中各维度的出现频次。由于目前缺乏统一术语标准,相同维度可能会使用同义词来表示,例如 consistency 和 conformance, accuracy 和 correctness 等;不同含义的术语在不同研究中也代表相同维度。因此根据维度定义统计相同维度出现的次数并记录相关同义词,最后得到当前关注最高的 5 种维度,构建相关指标。

3.3.2 基于语料分析归纳指标 指标归纳流程,见图 1。(1) 语料分解。从文献中收集的语料可能存在一条语料描述多种质量问题的情况。为实现指标对质量问题的唯一表达,对语料进行分析与拆解。例如从文献^[22]中收集到的语料“数据一致是指相同数据在不同应用、系统中的描述应该一致,例如患者的医疗档案编号。此外相关联的数据也应当不冲突”。该语料描述了两种情况,因此将其拆分为“数据一致是指相同数据在不同应用、系统中

的描述应该一致,例如患者的医疗档案编号”,“相关联的数据也应当不冲突”两条描述。(2) 语料分类。依据内容对语料进行分类。第 1 类语料是文献作者对相同类型临床数据质量问题的归纳总结,例如“数据包包含任务所需的所有数据项目”;第 2 类语料通常是文献作者用来举例说明的示例,需经过归纳总结才能表示为相应的评估指标,例如“数据应包含患者的血压、心率值”。(3) 指标合并。对于第 1 类语料,通过分析语料语义对相似观点进行合并,从而形成对应的评估指标。语料语义分析主要从其描述的数据对象(数据值、数据记录等)、数据问题(数据缺数、重复等)或数据约束(不为空、不重复等)描述、对应的质量维度(完整性、合理性)3 个方面进行考虑。(4) 指标映射。将第 2 类语料与步骤 3 中已得到的指标进行映射,剔除能映射到指标的语料,避免指标重复。(5) 归纳总结。对于步骤 4 中尚未对应指标的语料,基于其含义归纳总结相应的临床数据质量评估指标。与步骤 3 中类似,将语料描述抽象为 3 个部分:数据对象、数据问题或数据约束以及对应的维度。最终将所有语料转化为相应的评估指标。

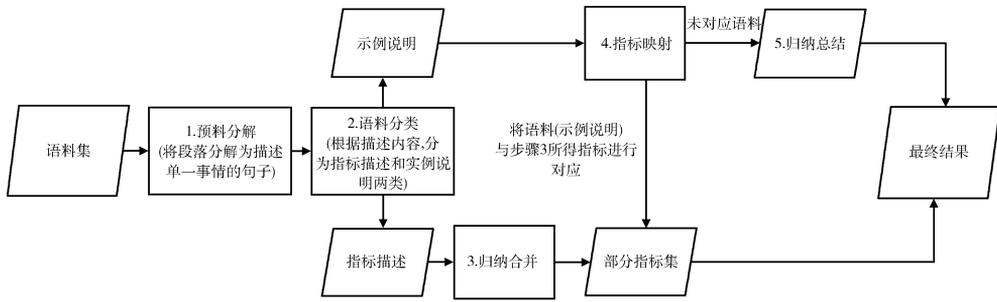


图 1 指标归纳流程

3.3.3 基于数据问题报告归纳指标 本研究所收集的问题数据报告主要包含数据库表名、问题描述、不符合要求数 3 部分内容。其中问题描述部分是对质量问题发生的位置、对象以及对质量问题的描述。问题数据分析总结流程，见图 2。首先，对于数据问题报告中的每条记录，分析问题描述部分，获取两部分关键信息：问题发生的数据对象和

该质量问题的描述。其中根据其数据对象的描述，将其抽象为数据元、数据值或数据记录等。其次，根据所抽象的评估对象以及问题描述，将其与已成立的指标进行映射匹配。若评估的数据对象和质量问题描述都能匹配至对应指标则说明该条问题记录已存在对应指标；若没有对应的指标与问题关联，则针对该问题新建指标。

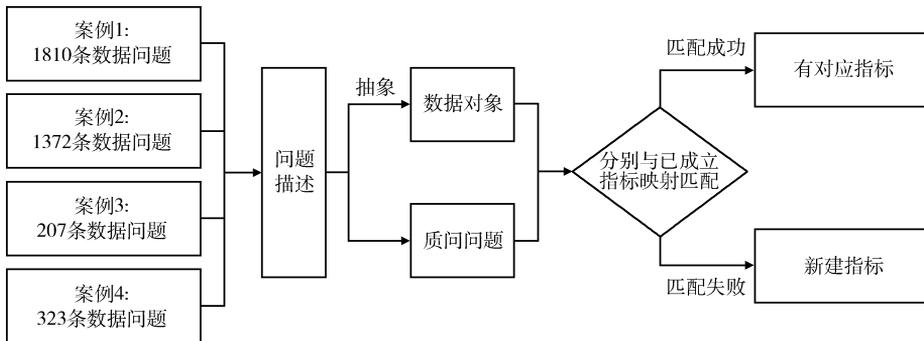


图 2 数据问题归纳为指标的流程

3.3.4 指标信息补充 对于最终得到的指标，按照其指标属性定义，依据指标所来源的文献描述或案例场景的实际情况，对其补充评估方法、适用场景等相关信息。

3.4 结果验证

Kahn 在 2016 年提出一套质量评估框架，在其研究中将个人提出的框架与前人研究框架进行比较，证明个人提出的评估框架具有一定全面性^[9]。我国卫健委发布的《电子病历系统应用水平分级评价标准》^[38]（下文简称国标）是针对电子病历应用评级的质量要求标准，按照维度对不同类别临床数据提出相应的质量评估需求。病房病历是国标中的

一个类别，其中包括患者住院病历所涉及的相关数据，这些数据因为含有丰富临床信息，有很高的二次利用价值，其数据质量问题被广泛关注^[40]。本研究设计以下对比试验来验证研究所提出指标的全面性和描述唯一性。选择国标中病房病历类别相关的质量评估需求作为金标准，将本研究提出的评估指标和 Kahn 框架中所提出的评估指标与国标的质量需求进行对应，比较两种指标的完善程度和对应需求的难易程度。

4 实验结果与讨论

4.1 当前统计的热点维度

基于对 30 篇文献的分析，出现次数最高的前 5

种质量维度和定义,见表 2,分别为完整性(28/30)、准确性(24/30)、时间性(21/30)、一致性(18/30)和合理性(18/30)。

表 2 质量维度、相关术语和定义

维度名称	相关术语	定义
完整性	Completeness; Integrity; Comprehensiveness	数据的存在或缺失
准确性	Accuracy; Correctness; Validity; Reliability	数据是否准确反映所记录的临床事实
时间性	Timeliness; Currency; Temporal	数据实时、及时的程度
一致性	Consistency; Conformance; Concordance; Agreement;	相互关联的数据是否表达一致
合理性	Plausibility; Trustworthiness; Validity; Reliability	数据的表达是否符合专业知识或使用预期

4.2 质量评估指标

4.2.1 指标及其定义 基于 30 篇相关文献以及 3712 条临床实际问题,本研究从 5 种质量维度归纳出 43 项评估指标。各指标代码、名称和指标来源,见表 3-7。43 项指标在文献中均有提及,29 项指标源自数据问题报告。

表 3 数据完整性相关指标

指标编码	指标名称	指标来源
C01	数据元完整	文献、数据问题报告
C02	数据值完整	文献、数据问题报告
C03	记录完整	文献、数据问题报告
C04	数据总量足够	文献、数据问题报告
C05	记录事实完整	文献
C06	信息完整	文献、数据问题报告
C07	上下文完整	文献、数据问题报告
C08	时序数据完整	文献、数据问题报告
C09	数据状态完整	文献

表 4 数据合理性相关指标

指标编码	指标名称	指标来源
P01	数值合理	文献、数据问题报告
P02	标识不重复	文献、数据问题报告
P03	重复值合理	文献、数据问题报告
P04	结论合理	文献
P05	上下文合理	文献、数据问题报告
P06	数据客观	文献、数据问题报告
P07	主索引合理	文献、数据问题报告
P08	冗余数据	文献、数据问题报告
P09	记录出现次数	文献
P10	数据量的变化	文献

表 5 数据一致性相关指标

指标编码	指标名称	指标来源
S01	数值一致性	文献、数据问题报告
S02	内容与编码一致	文献、数据问题报告
S03	计算参数与结果一致	文献
S04	字段默认值一致	文献、数据问题报告
S05	数据关系与结构一致	文献
S06	内容与定义一致	文献
S07	计算方法一致	文献
S08	数据的度量单位一致	文献

表 6 数据时间性相关指标

指标编码	指标名称	指标来源
T01	数据的时效	文献、数据问题报告
T02	数据的时间逻辑	文献、数据问题报告
T03	数据的记录频率	文献
T04	数据的实时程度	文献、数据问题报告
T05	数据的及时程度	文献
T06	持续时间	文献

表7 数据准确性相关指标

指标编码	指标名称	指标来源
A01	计算准确	文献
A02	数据类型准确	文献、数据问题报告
A03	表达合法	文献、数据问题报告
A04	表达精确	文献、数据问题报告
A05	数据关系的粒度	文献、数据问题报告
A06	表达准确	文献、数据问题报告
A07	数据格式准确	文献、数据问题报告
A08	编码/术语准确	文献、数据问题报告
A09	测量准确	文献
A10	元数据准确	文献、数据问题报告

4.2.2 结果验证 本研究选择国标中住院病历相关质量需求作为金标准,比较 Kahn 指标、本研究所提出的评估指标与国标需求的对应唯一性。结果表明国标中关于住院病历的质量需求能够与本研究所提出的指标实现一对一对应;而在与 Kahn 的对应过程中,需求(1-4),(6-7)没有找到对应指标;需求(5)、(12-13)能够明确对应到 Kahn 框架中的某个指标;需求(8-11)能模糊对应到 Kahn 指标。本研究所提出的质量评估指标能够更全面、更准确地对应到实际评估需求。

4.2.3 结果讨论 标准化的临床数据质量评估对于提高临床数据质量意义重大。构建全面、描述唯一的质量评估指标是实现标准化评估的重要前提。本研究结合文献知识与实际数据问题,提出一套临床数据质量评估指标,详细描述指标所适用的评估方法,为实现标准化、自动化的临床数据质量评估奠定基础。本研究选择两种不同类型临床场景数据问题报告,收集的问题条目数量上有较大差异,原因有以下两点:一是在本研究所选择的注册研究采集场景中研究人员将不同机构数据合并后再进行评估,因此同一份报告中不同机构的相同数据问题被合并在一起;二是注册研究场景中各机构会根据报告迅速地采取相关措施,因此后续报告中数据问题会逐渐减少。然而在本研究的分析过程中,同一类型数据问题只要出现一次就会关联相关指标,同类型问题出现的次数不会影响最终指标结果。因此在本研究中两种场景数据问题在数目上的差异没有影响最终建立指标的结果。高效、自动的数据质量评

估技术是研究者一直致力的目标。本研究描述评估指标所适用的评估方法,构建评估需求与评估方法的联系,为研究自动化评估技术提供一定支持。统计指标适用的评估方法,有29项指标(67%)可以通过逻辑规则方法进行评估,其余14项(33%)则适用于统计分析、自然语言处理算法等手段进行评估。这29项指标涉及5个质量维度,其中包括数据值完整、数值合理等被大多数研究或标准所要求的评估指标,因此也说明基于逻辑规则的评估方法是一种相对普适的评估手段。然而逻辑规则的定义与具体的知识相关,目前逻辑规则的创建主要依赖人工完成。基于本研究,研究者能够了解适于采用逻辑规则方式进行评估的指标,这些指标代表了相关临床数据质量问题类型。因此如果能够找到与这些指标相关的评估知识源,则可以在这些知识源的基础上探索自动化评估的可能性。目前逻辑规则评估所涉及的知识包括临床知识、通用常识以及数据知识等。临床指南、临床信息模型中都对这些知识有一定的描述。通过自然语言处理等技术提取其中的相关知识自动化的生成评估规则是具有可行性的自动化评估手段之一。后续本研究基于指标探索以临床信息模型为基础的自动化逻辑规则生成技术。

5 结语

本研究通过对质量评估相关文献和实际数据问题中相关知识进行收集、分解、合并与归纳,基于当前关注度较高的质量维度,即完整性、合理性、准确性、一致性、时间性,构建包含43项对数据问题进行唯一表述的评估指标,描述各项指标所适用的评估方式。通过与传统指标以及国家标准的比较,本文所提出的质量评估指标更全面地涵盖国标中的评估需求,更清晰、准确地映射到现有国家质量评估标准。对于临床数据质量评估的标准化有促进意义。本研究也存在一些不足:首先,即使结合文献与实际数据问题报告总结归纳指标,仍然难以保证指标全面性。收集文献时尽量更多地纳入相关文献,在案例选择上充分考虑典型性,从而最大可能地代表当前实际应用中质量评估的需求。其次,本

研究提出的指标需组织专家进行评估和验证,在后续工作中将组织经验丰富的专家对指标进行讨论和完善^[12]。

参考文献

- Whitelaw F G, Nevin S L, Milne R M. Completeness and Accuracy of Morbidity and Repeat Prescribing Records Held on General Practice Computers in Scotland [J]. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners*, 1996, 46 (404): 181 - 186.
- 彭传薇, 刘琛玺, 李小华. 浅谈医疗数据质量重要性及其影响 [J]. *解放军医院管理杂志*, 2005, 12 (5): 467 - 468.
- Amber Lee Dennis. Case Study: using data quality and data management to improve patient care - DATAVERSITY [EB/OL]. [2020-06-25]. <https://www.dataversity.net/case-study-using-data-quality-data-management-improve-patient-care/>.
- Makary M A, Daniel M. Medical Error—the third leading cause of death in the US [J]. *BMJ*, 2016 (353): i2139.
- Maydanchik A. Data Quality Assessment [M]. New Jersey: Technics Publications, 2007: 25 - 28.
- Lee K, Weiskopf N, Pathak J. A Framework for Data Quality Assessment in Clinical Research Datasets [C]. Washington DC: Annual Symposium Proceedings. AMIA Symposium, 2017: 1080 - 1089.
- Weiskopf N G, Weng C. Methods and Dimensions of Electronic Health Record Data Quality Assessment: enabling reuse for clinical research [J]. *Journal of the American Medical Informatics Association*, 2013, 20 (1): 144 - 151.
- Johnson S G, Speedie S, Simon G. A Data Quality Ontology for the Secondary Use of EHR Data [C]. San Francisco: Annual Symposium Proceedings. AMIA Symposium, 2015: 1937 - 1946.
- Kahn M G, Callahan T J, Barnard J. 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): 18.
- Chen H, Hailey D, Wang N. 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.
- Reimer A P, Milinovich A, Madigan E A. Data Quality Assessment Framework to Assess Electronic Medical Record Data for Use in Research [J]. *International Journal of Medical Informatics*, 2016 (90): 40 - 47.
- Mccormack J L, Ash J S. Clinician Perspectives on the Quality of Patient Data Used for Clinical Decision Support: a qualitative study [C]. Chicago: Annual Symposium Proceedings. AMIA Symposium, 2012: 1302 - 1309.
- Taggart J, Liaw S - T, Yu H. Structured Data Quality Reports to Improve EHR Data Quality [J]. *International Journal of Medical Informatics*, 2015, 84 (12): 1094 - 1098.
- Zhang R, Wang Y, Liu B. Clinical Data Quality Problems and Countermeasure for Real World Study [J]. *Frontiers of Medicine*, 2014, 8 (3): 352 - 357.
- Liaw S - T, Taggart J, Dennis S. Data Quality and Fitness for Purpose of Routinely Collected Data - a General Practice Case Study from an Electronic Practice - based Research Network (EPBRN) [C]. Washington DC: Annual Symposium Proceedings. AMIA Symposium, 2011: 785 - 794.
- Logan J R, Gorman P N, Middleton B. Measuring the Quality of Medical Records: a method for comparing completeness and correctness of clinical encounter data [C]. Washington DC: Proceedings of AMIA Symposium, 2001: 408 - 412.
- Ali S M, Anjum N, Kamel Boulos M N. Measuring Management's Perspective of Data Quality in Pakistan's Tuberculosis Control Programme: a test - based approach to identify data quality dimensions [J]. *BMC Research Notes*, 2018, 11 (1): 40.
- Weiskopf N G, Bakken S, Hripcsak G. A Data Quality Assessment Guideline for Electronic Health Record Data Reuse [J]. *eGEMs (Generating Evidence & Methods to Improve Patient Outcomes)*, 2017, 5 (1): 14.
- Khare R, Utidjian L, Ruth B J. A Longitudinal Analysis of Data Quality in a Large Pediatric Data Research Network [J]. *Journal of the American Medical Informatics Association*, 2017, 24 (6): 1072 - 1079.
- Diazvasquez J F. DataGauge: a model - driven framework for systematically assessing the quality of clinical data for secondary use (2016) [EB/OL]. [2020-10-14]. https://digitalcommons.library.tmc.edu/cgi/viewcontent.cgi?article=1033&context=uthshis_dissertations

- 21 Canadian Institute for Health Information. CIHI's Information Quality Framework | CIHI [EB/OL]. [2020-06-05]. <https://www.cihi.ca/en/document/cihis-information-quality-framework>.
- 22 Davoudi S. Data Quality Management Model (2015 Update) - Retired [EB/OL]. [2020-10-14] <https://bok.ahima.org/doc?oid=107773#.X4ab9egzaUk>
- 23 Dixon B E, McGowan J J, Grannis S J. Electronic Laboratory Data Quality and the Value of a Health Information Exchange to Support Public Health Reporting Processes [C]. Washington DC: Annual Symposium Proceedings. AMIA Symposium, 2011: 322-330.
- 24 Liaw S T, Rahimi A, Ray P. Towards an Ontology for Data Quality in Integrated Chronic Disease Management: a realist review of the literature [J]. International Journal of Medical Informatics, 2013, 82 (1): 10-24.
- 25 Vanbrabant L, Martin N, Ramaekers K. Quality of Input Data in Emergency Department Simulations: framework and assessment techniques [J]. Simulation Modelling Practice and Theory, 2019 (91): 83-101.
- 26 Weiskopf N G, Hripcsak G, Swaminathan S. Defining and Measuring Completeness of Electronic Health Records for Secondary Use [J]. Journal of Biomedical Informatics, 2013, 46 (5): 830-836.
- 27 Brown J S, Kahn M, Toh S. Data Quality Assessment for Comparative Effectiveness Research in Distributed Data Networks [J]. Medical Care, 2013, 51 (Suppl 3): S22-29.
- 28 Kahn M G, Raebel M A, Glanz J M. A Pragmatic Framework for Single-site and Multisite Data Quality Assessment in Electronic Health Record-based Clinical Research [J]. Medical Care, 2012, 50 (Suppl 2): S21-29.
- 29 张瑶. 影响临床数据质量的因素及数据采集方法的选择 [D]. 上海: 复旦大学, 2012.
- 30 刘智, 商洪才, 翟静波. 基于网络的中医药临床试验在线数据核查 [J]. 天津中医药, 2014, 31 (11): 670-673.
- 31 陈朝华, 黄钦, 邓亚中. 临床试验数据管理质量评价指标体系 [J]. 药学学报, 2015, 50 (11): 1374-1379.
- 32 董冲亚, 姚晨, 高嵩. 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.
- 33 何丽云, 刘保延, 梁志伟. 临床研究中数据管理及其质量评价 [J]. 中国新药与临床杂志, 2005, 24 (11): 916-919.
- 34 芦殿荣, 李冬云, 陈信义. 临床研究课题资料收集过程中的数据管理及其质控探讨 [J]. 中医药管理杂志, 2008, 16 (10): 753-755.
- 35 U. S. Food and Drug Administration. Code of Federal Regulations Title 21 [EB/OL]. [2020-08-15]. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcr/CFRSearch.cfm?CFRPart=58>.
- 36 中国政府网-信息统计中心. 2016年全国卫生统计信息中心主任工作会在沈阳召开 [EB/OL]. [2020-10-08]. <http://www.nhc.gov.cn/mohwsbwstjxxzx/s8555/201606/8114e70a5d624dbb833b6d23ce53c9f2.shtml>.
- 37 Min L, Tian Q, Lu X, et al. An OpenEHR Based Approach to Improve the Semantic Interoperability of Clinical Data Registry [J]. BMC Medical Informatics and Decision Making, 2018, 18 (1): 49-56.
- 38 国家卫生健康委办公厅. 关于印发电子病历系统应用水平分级评价管理办法及评价标准的通知 [EB/OL]. [2020-06-05]. http://www.gov.cn/xinwen/2018-12/09/content_5347261.htm.
- 39 国家卫生计生委统计信息中心. 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]. [2020-06-05]. <http://chiss.org.cn/hism/wcmpub/hism1029/inform/201709/P020170904374922732849.pdf>
- 40 Singer A, Yakubovich S, Kroeker A L. Data Quality of Electronic Medical Records in Manitoba: do problem lists accurately reflect chronic disease billing diagnoses? [J]. Journal of the American Medical Informatics Association, 2016, 23 (6): 1107-1112.