

区域全民健康信息平台居民就医数据采集和管理策略

汤娟华 朱琴芬 赵光炜 吴亚平

(常州市卫生信息中心 常州 213003)

〔摘要〕 分析数据采集方式对质量的影响, 阐述数据质量分析与评估具体要求, 包括完整性、原始性、唯一性、一致性等方面, 从业务、信息、监管等方面提出数据质量管理主要措施。

〔关键词〕 患者流向分析; 数据采集; 数据质量; 节点

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2019.11.004

Data Acquisition and Management Strategy for Resident Medical Data on Regional National Healthcare Information Platform

TANG Juanhua, ZHU Qinfen, ZHAO Guangwei, WU Yaping, Changzhou Health Information Center, Changzhou 213003, China

〔Abstract〕 The paper analyzes the influence of data acquisition methods on data quality, expounds the specific requirements of data quality analysis and evaluation in terms of completeness, originality, uniqueness, consistency, etc., and puts forward the main measures of data quality management from the aspects of business, information, supervision, etc.

〔Keywords〕 analysis of patient flow; data acquisition; data quality; node

1 引言

患者流向分析是指从居民就医数据入手, 分析患者的就诊习惯和行为以及相关疾病和诊断数据, 提出政策性的引导和建议。分析的数据来源是本地区的居民就医数据。这些数据分散在各二、三级医院和社区, 必须依靠区域全民健康信息平台(以下简称区域平台), 通过卫生专网采集各医院和社区的数据, 区域平台将采集的数据进行汇聚整合, 按主题形成数据仓库或数据集市, 支持

课题研究。数据的充分利用完全依赖区域平台的数据质量。

2 数据采集方式对质量的影响

2.1 两种数据采集方式

目前采用的数据采集方式是医院上传数据到前置机, 区域平台从前置机采集。医院提供的数据来源有两种, 一是业务系统数据库, 二是医院数据中心或者医院信息集成平台(以下简称医院平台)。第1种方式采集的数据直接来自于业务系统, 数据采集架构和节点, 见图1。各业务系统都需要和采集程序做接口, 接口成本高; 各业务系统数据字段不同、标准不一导致数据异构, 增加数据转换的难度; 如果业务系统发生变更, 采集程序不一定能及

〔收稿日期〕 2019-05-08

〔作者简介〕 汤娟华, 硕士, 高级工程师, 发表论文4篇;
通讯作者: 吴亚平, 副教授。

时发现并做出相应调整, 导致程序不稳定; 采集到的数据医院一般不会利用, 因此医院不能发现其中的错误, 而且上传到前置机的数据来源发生变化, 区域平台中心库更是无法感知; 由于医院数据的多源性, 数据采集定位不易掌握与控制, 业务系统变更导致数据来源发生变化, 区域平台无法感知并及时进行调整。这些都会在数据采集的第 1 环节造成不可控的质量问题。第 2 种方式是从医院数据中心获取数据, 数据采集架构和节点, 见图 2。

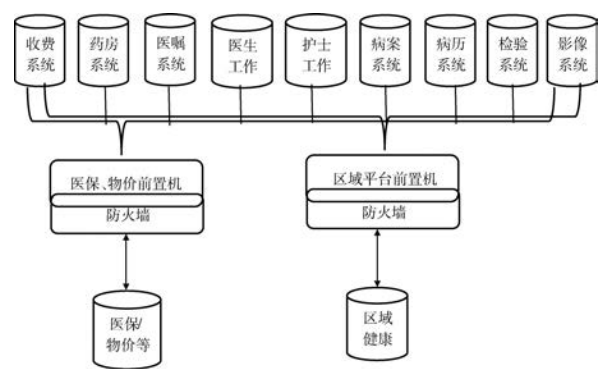


图 1 业务系统数据采集架构和节点

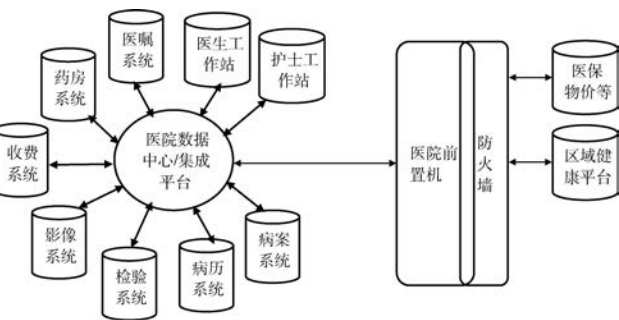


图 2 医院数据中心数据采集架构和节点

2.2 采集方式对比

医院建立数据中心, 通过医院平台在内部首先实现数据级的集成, 完成完整性、一致性、唯一性、原始性、及时性等基本的数据质量控制。在医院内部提供一个信息共享和利用的平台, 同时为医院提供统一的信息对外出口, 避免单个业务系统与多个外部系统做各种接口, 降低数据采集的工作量和难度。通过医院数据中心上传区域平台, 由于区域平台与医院使用同源数据, 容易发现并及时纠正

数据质量问题, 可以更好地保障区域平台获取的数据质量。比较上述两种基本方式, 建议使用第 2 种方式。

2.3 数据质量控制节点

以上两种数据采集方式中有 3 个重要的数据质量控制节点: 一是数据采集定位节点, 如果医院建有数据中心, 而且对数据的一致性与唯一性有管制机制, 那么该环节的数据质量控制是有保障的。如果医院没有统一管理的数据中心, 直接到业务系统采集数据, 那么数据采集质量完全依靠采集人的能力, 即对医院众多异构业务系统及数据的熟练掌握程度, 这种方式缺乏机制保障, 数据质量具有极大的不确定性。二是前置机节点, 通过提取转换加载 (Extract, Transform and Load, ETL) 工具对采集数据进行清洗、转换、装载等操作^[2], 建立定时按规范上传区域平台的机制。三是区域平台中心库节点, 对已经通过前置机初步校验欲装载进入区域平台数据进行完整性、一致性和唯一性等控制, 尤其是唯一性控制。最后通过 ETL 工具按主题进行数据分发, 形成支持个案数据、就诊地区、疾病诊断、手术、费用、质量、效率等主题的数据仓库或数据集市。

为提高区域平台对各接入节点状态以及数据质量的监控, 平台需建立完整的监管服务功能。集成网络设备监管系统的监控数据, 实现接口数据质量评估, 同时建立完善的数据监控机制, 从而对接入节点的网络、硬件状态及数据采集情况 (分类采集数量、采集成功率、质量评估结果等) 进行综合展示, 以指导各接入机构进行相应的改造和接口优化^[3]。

3 数据质量分析与评估

3.1 概述

区域平台数据库面向规模数据应用, 与支持业务流程的业务系统数据库管理方式有质的不同。业务系统侧重于根据业务流程设计, 满足业务开展需要, 各类业务信息系统在数据集成和统一服务上考

虑并不多。但从诊疗过程来看,这些数据之间在时间、因果、相关关系上又有联系,而这些联系的背后往往是数据的潜在价值。因此经区域平台管理的数据在结构上要面向主题,数据质量上要达到完整性、原始性、唯一性、一致性、逻辑性、及时性等方面的要求。

3.2 完整性

是对数据的基本要求。完整性至少要表现为事件、表单、表项完整。如每天的门诊服务人次是否记录、采集完整,是事件完整。门诊产生多少张处方、检查检验申请及相关报告等是否记录、产生完整,是表单完整。表单项目是否填写、采集完整,是表项完整。在医院平台的患者 360 统一视图表现为某患者在某次门诊、住院事件中实际产生的数据是否记录、采集完整。

3.3 原始性

要求相关数据无论经过多少次的映射转换、拆分合并操作必须采集保留原始原样记录。如临床诊断不能只采集 ICD 编码及其经映射转换后的标准分类名称,还必须采集、记录医生录入的原始临床诊断;药品处方不能只采集药品编码,还要采集处方原始的药品名。医院归档的病历文书是最原始也是最可靠的文档,在采集过程中要原样保留。

3.4 唯一性

往往容易被忽略但却是决定数据质量的关键。唯一性是指数据中心库对同一个数据或指标不能有两个以上的取值或解释,即不能有二义性。如计算药占比时需要医院总收入数值,当定义为门急诊费用加住院总费用时,其他费用等就不能计算在内,在统计指标数据元中只能有一个总收入。如果管理需要两个以上的总收入计算口径,则必须在名称表达(符号)和数据元标识符上予以区分。数据库表定义的 ID 主键要唯一,标识居民的身份也要有唯一性,身份证号不能有重复记录。保障平台数据唯一性的主要技术手段是元数据管理。

3.5 一致性

是指数据中心的中心库存储及使用的数据均执行或符合一个数据标准,包括语法和语义标准。也就是同平台的主数据保持一致。保障平台数据一致性的主要技术手段是主数据管理。

3.6 逻辑性

指描述同一对象时不能有逻辑矛盾,如医学知识的矛盾,对病案首页数据的稽核中很容易出现逻辑性错误。包括居民就医数据中有级别高(2、3、4 级)的手术记录,但没有麻醉记录;出生日期不能小于当前系统日期;项目金额不等于项目单价乘以项目数量;男性患者出现女性解剖名称或疾病症状体征等,或者女性出现男性疾病症状体征等,这些都是逻辑性错误。

3.7 及时性

是指数据采集时间、处理时间和平台展现时间要做到实时、准时或者按照约定。

4 数据质量管理主要措施

4.1 概述

数据质量管理是一个动态过程,数据质量的提高需要持续不断地改进。数据质量的保障除上述数据采集过程必须予以关注并有效控制的几个主要环节外,还应该从业务、信息、监管 3 方面着手解决。

4.2 业务

随着对医院管理及医院信息化要求的不断提高,有的业务系统需要进行标准化改造,或进一步优化和创新,通过改变业务流程和管理模式使业务更加流畅、管控更加有效,从而提升数据质量。业务部门是业务系统建设的主导,要主动参与,信息部门主动配合。

4.3 信息

信息部门加强日常业务培训,明确数据采集的

管理要求并进行细化。如采集数据要求二级及以上医院每天晚上自动上传当天门诊和出院患者数据。区域平台对采集数据进行业务校验,有挂号必须有诊疗数据,有诊疗数据必须有挂号数据,有出院记录必须要有入院记录等。医院要强调操作手册的重要性,规范操作人员业务行为,避免不合规、非标准的数据进入系统。区域平台要将其相关数据与医院统计数据比对,定期公布。

4.4 监管

从监控和管理角度,各个采集节点之间环环相扣。上传方需要先按标准转换、对应数据,完成后请业务方或接收方核验,通过后再进行数据上传。采集时各个软件方利用编译程序调用服务,如果将上传、采集数据过程进行界面流程式的智能节点配置,采集程序将更加稳定。日常工作中对前置机的应用程序、网络状况和后台服务定时检查,最好能做到监测和报错短信提醒,如有问题双方都能及时发现并处理。接收方要定期查看前置机的校验服务、数据库状态、主控制器等是否正常运行。

4.5 PDCA 循环

除以上各环节必须采取的积极应对措施外,建立数据质量持续改进、提高的机制更为重要,即建立数据质量戴明环(PDCA 循环)。首先,建立数据质量评估指标体系,用一组指标来客观地描述数据质量。其次,为每个数据质量指标创建可操作的测量方法。最后,定期形成数据质量报告。该质量报告

必须能够明确指出具体的数据质量问题,供相关部门有针对性地整改,以数据质量报告促进医院业务、信息化、平台及数据中心建设,对整改情况也需要提供整改评估报告,以此形成完整的 PDCA 循环。

5 结语

全民健康医疗已经进入云计算、大数据、物联网、“互联网+”、区块链、人工智能的时代,上述先进技术都将以区域健康信息平台 and 医院信息集成平台为基础展开。数据质量是各级各类平台的生命线,现有数据的分析和利用、大数据应用的广阔前景都需要基于一定数据量的积累和数据质量的保证,对数据的标准化、完整性、可获得性提出更高要求。希望形成一套确保数据质量的可信机制,这些可信数据将为医学研究、临床决策、运营管理、医疗质量管理提供更好支撑,为医疗大数据、人工智能在全民健康领域的应用奠定坚实基础。

参考文献

- 1 蔡晋龙,陈金雄. 基于集成平台的医院信息系统集成方案[J]. 中国医疗器械杂志, 2017, 31 (6): 420.
- 2 张岩龙,龙虎,时琳. 面向健康医疗的资源整合分析方法及平台建设[J]. 中国卫生信息管理, 2017, 2 (1): 25-26.
- 3 倪宁. 区域卫生信息平台省市对接实践[C]. 济南: 2018 (13th) 中国卫生信息技术/健康医疗大数据应用交流大会, 2018: 48-51.

关于《医学信息学杂志》启用 “科技期刊学术不端文献检测系统”的启事

为了提高编辑部对于学术不端文献的辨别能力,端正学风,维护作者权益,《医学信息学杂志》已正式启用“科技期刊学术不端文献检测系统”,对来稿进行逐篇检查。该系统以《中国学术文献网络出版总库》为全文比对数据库,可检测抄袭与剽窃、伪造、篡改、不当署名、一稿多投等学术不端文献。如查出作者所投稿件存在上述学术不端行为,本刊将立即做退稿处理并予以警告。希望广大作者在论文撰写中保持严谨、谨慎、端正的态度,自觉抵制任何有损学术声誉的行为。

《医学信息学杂志》编辑部