

基于单病种数据库的临床科研系统的设计与研发

高宇,王奕

(复旦大学附属肿瘤医院,复旦大学上海医学院肿瘤学系,上海 200032)

摘要:基于单病种数据库的临床科研系统针对特定病种的疾病,收集和整理临床数据,对疾病的发病、治疗、转归等进行记录,使得医院系统中临床数据资源价值得到极大提升,促进肿瘤临床科研水平的提高。文章分析了数据中心的现状,介绍了基于肿瘤通用标准数据集的单病种数据模型的架构,详细阐述了临床科研系统的功能,主要包括病例多维度综合查询、临床科研表单录入、随访数据共享等,讨论了系统在实际运用中存在的不足,并对科研系统的数据应用前景进行展望。

关键词:单病种;肿瘤通用标准数据集;临床科研;数据中心

中图分类号:R197 文献标识码:A 文章编号:1004-0242(2017)09-0677-06

doi:10.11735/j.issn.1004-0242.2017.09.A005

Design and Development of Clinical Research System Based on Single Disease Database

GAO Yu, WANG Yi

(Fudan University Shanghai Cancer Center, Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China)

Abstract: Clinical research system based on single disease database collects and processes clinical data, documents the incidence, treatment and outcome of specific disease. The system can greatly enhance the utility of clinical data resources, and improve the clinical research levels. This paper analyzes the present situation of the clinical data center, introduces the structure of single disease data model based on the standardized cancer data sets, elaborates the functions of clinical research system, including multi dimension query of clinical data, the clinical research form entering, the follow-up data sharing and so on. The paper also discusses the shortcomings in practical application and prospects its future development.

Key words: single disease; tumor common standard data sets; clinical research; clinical data center

临床科研是衡量肿瘤医院学术水平的重要标志之一,是催生新知识、新技术的基本手段,是推进医院发展的重要动力,临床科研数据库是科研平台建设中的重要组成部分^[1]。近年来随着“大数据”概念的引入,数据平台以及单病种数据库逐渐被使用到医院的临床科研、数据收集和整理、随访等方面,在医院信息化管理中发挥重要的作用。单病种数据库针对各种特定病种疾病,涵盖了患者的一般资料、诊断治疗过程、后续治疗、随访记录和预后情况等,对疾病的发病、治疗、转归等进行记录^[2]。2012年初,我

院在数据平台基础上建立了单病种数据库,并逐渐开始探索在临床科研上的应用。经过探索尝试后发现,数据平台汇集了全院的临床和运营数据,但它的数据结构及数据关联不一定适合于特定病种科研数据的高效利用,因此我们在数据平台基础上进一步构建了应用数据层,包括面向特定病种的专用数据模型。在构建面向科研应用的单病种数据库数据模型时,我们引入了肿瘤通用标准数据集的概念,同时研发了一系列的科研应用,提供强大的多维度筛选工具、患者临床信息统一视图、完整的课题和实验组管理、提供便捷的科研病例表单(clinic research form, CRF)填写工具、完善的单病种科研用户权限

收稿日期:2016-12-28;修回日期:2017-02-27

通讯作者:王奕, E-mail: tonywang@shca.org.cn

管理以及科研数据导出功能等,支持临床科研的回顾性研究、前瞻性研究以及横断面研究。

1 单病种数据模型的设计

医院数据平台的建立就是将原有系统的数据进行分析和梳理,在物理汇聚的基础上遵循行业或卫生部技术标准所定义的数据模型重新进行建模,按照以患者为中心的方式对临床业务数据进行集中管理^[3]。我院数据平台是通过建立一组统一的数据接口,将各个系统中分散的数据整合汇总,通过对数据的解析、梳理和重构将各个业务系统中的数据成功地汇集到一个数据仓库中。数据平台的核心是以患者为中心,因此将各个复杂的临床业务系统中的病例数据关联在一起,是医院内部资源共享及数据平台建立的必要条件。我院早在2013年实现了EMPI^[4]技术,通过身份证号和姓名相匹配的规则将患者在门诊和住院中的记录整合在一起,再通过企业服务总线(enterprise service bus,ESB)提供规范的接口将EMPI号发布到各个业务系统,针对身份证号缺失的历史数据,则利用住院号和姓名匹配、门诊号和姓名匹配、或者申请单号和姓名匹配的规则进行梳理,尽可能保证数据平台中数据的完整性。为了更高效地利用这些临床数据,我们在数据仓库基础上进一步构建应用数据层(针对专门数据应用优化过的数据模型)。如面向医疗质量管理及医院运营管理的商业智能分析BI的多维Cube模型,以及面向科研应用的单病种数据库数据模型(附录A)。

1.1 单病种数据库的创建

根据临床科研的规律,一般临床科研项目都是由多学科团队发起,而多学科团队一般都是以单病种划分。所以临床科研项目的样本集一般都属于一个单病种,而单病种之间的研究方向相差很大^[5]。因此,我们按照临床医生的经验为每个单病种定义了不同的入库规则,例如,骨软组织肿瘤病种将病理诊断及门急诊、入出院诊断中包含“肉瘤”、“骨瘤”、“神经瘤”等多达73种关键词的患者筛选入库;而乳腺肿瘤则将乳腺外科诊断中包含“乳”关键词的患者及化疗科诊断中包含“乳”关键词且开过抗激素抗肿瘤类药物的患者筛选入库。按照入库规则筛选符合的病例进入单病种数据库后,再抽取该病种关注的诊

疗数据,由于业务系统中有很多非结构化数据,不能直接符合数据集字段的要求,系统在抓取数据的同时,自动对部分半结构化的数据做颗粒化、后结构化处理,以满足回顾性查询所需的数据细化程度。例如,我院病理系统可以设置病种报告模板,单病种数据库则可以按照固定的模式将模板化的病理报告解析成结构化报告。后结构化处理之后,有些不规范的文本格式的字段仍然无法统计,比如大肠癌病理报告里的浸润深度,该数据是用来分类病理分期的,对患者的后续治疗起着至关重要的作用,因此根据临床医生的经验我们将不同写法根据关键词做了归纳,很好地解决了后续无法统计的问题,为临床科研提供更好的支持。

1.2 肿瘤通用数据集标准

在构建单病种数据模型的时候,我们首先定义了肿瘤通用数据集标准,单病种数据库的设计在这个数据集标准的基础上进行的。单病种数据集标准定义了病例人口学信息、临床病史、就诊记录、检验结果、肿瘤标志物、检查结果、病理结果、免疫组化、分子诊断、疾病分期、治疗方案、手术治疗、化疗、放射治疗、靶向治疗、出院小结、费用信息、随访信息等。例如,遵循国家卫生行业标准WS445.10-2014电子病历基本数据集标准^[6]自定义了患者人口学信息数据项,根据实体肿瘤疗效评估标准(RECIST)定义了化疗疗效评估标准。肿瘤数据集标准包含各数据元的中英文名称、编码、数据元说明、值域定义及范围等,为单病种数据模型的构建提供标准化数据基础。

受限于前端业务系统参差不齐的数据完整度和质量,单病种数据库自动收集的数据还存在部分缺失和非结构化。为此,一方面我们可以通过CRF的形式补充录入高质量研究数据,例如,业务系统没有TNM分期数据,科研系统提供病理报告给医生参考,让医生在CRF中填写患者的TNM分期,该数据元事先已经定义好标准值域,确保不同医生录入的数据一致性。另一方面以标准数据集为依据,为业务系统的结构化改造和数据采集提供建议,逐步提高数据源质量,例如可以为我院正在开展的病历文书结构化及医技报告结构化项目提供依据。在数据集标准的帮助下,数据生产、管理和利用形成闭环并在运作中持续自我优化。数据集标准结合专业的临床

科研数据平台服务,使数据的质量、可访问性、安全和隐私都得到保证,其可扩展、可持续建设的特点为院际甚至区域级医疗数据共享创造可能。

2 临床科研系统的研发

在单病种数据库基础上,我们设计并研发了一套临床科研系统。临床科研系统可以展示单病种数据库中的所有病例数据,包含患者的诊疗过程中的完整维度数据;提供一套灵活强大的多维度数据查询筛选工具,用户可以对所有维度数据进行灵活地组合需要查询筛选的数据项及条件;对查询及筛选出的病例进行数据展示或者加入课题组;针对课题组关注的各项数据,再使用电子 CRF 工具填写课题需要的统计资料;CRF 表单录入完成之后,用户可自行选择需要的数据导出到专业的统计工具中做更深入地分析。

2.1 多维度数据综合查询及病例数据展示

科研系统基于已经建成的单病种数据库,用户可查看数据库中病例的全方位诊疗数据。在庞大的数据库基础上,提供强大的多维度组合筛选工具,满足用户各种查询需求,例如,大肠癌病种数据库提供结构化病理结果、免疫组化、肠镜活检、手术分类、新辅助放化疗等共 14 个维度(附录 B);脑脊液病种数据库则关注手术耗材、化疗方案、影像图片、手术视频等。

筛选出符合条件的病例后,可以调阅患者的临床信息统一视图,该视图基于临床数据中心,使临床医生可以通过一个清晰、友好的统一视图对患者的就诊信息进行查阅,同时还提供直观的导航模型,从而优化医生的操作流程,使临床医生在短时间内对患者就诊情况有整体了解,为患者提供高效、快捷、准确的医疗服务^[7]。该视图展示病例历次的门急诊及住院的临床资料,包括就诊记录、用药记录、诊断信息、医嘱信息、手术情况、检查、检验、病理、超声、内镜、放射、心电图、病历文书、诊疗过程等一系列信息。另外,系统还提供以时间轴展示患者多个临床指标的工具,用户在同一个时间轴上定义多个观察指标(如化验指标)及治疗处置(如放化疗或者药物治疗等),以观察分析多个指标与治疗处置之间的变化关系,是临床信息统一视图更精细化的应用,针对单

病种的科研病例,深入展现诊疗信息中有科研价值的部分^[8]。

2.2 CRF 设计及填写

用户根据条件筛选出满足课题组需要的病例后,可以统一添加到课题组,再使用课题匹配的 CRF 填写内容。系统提供了第三方系统入组接口,将入组功能嵌入临床业务流程,用户在看诊期间可随时收录符合条件的病例,及时填写 CRF。用户入组之后,系统提供了灵活的病例表单填写工具,实现系统的半自动化填写,在填写表单数据项的同时,系统会自动在右侧屏幕定位并展示和该数据项相关的患者临床信息,精准提供参考信息(附录 C)。表单的设计基于肿瘤通用标准数据集,这样不同研究课题的数据可以十分便捷地整合共享,因为它们对一个字段的定义和值域设置完全一致。标准数据集的引入,使得科研数据更加标准化、规范化,减少无效数据,提高临床科研数据的效率及准确率。表单填写完成之后,用户可以根据设置好的权限导出需要的数据,再导入到专业统计软件做更深入地分析。

表单设计可分阶段管理,同步于临床诊疗的疾病全程数据管理。例如术前检查、术中情况、术后用药、化疗、放疗、随访等阶段,更方便、清晰地管理患者的整个就诊流程;表单填写支持项目动态扩展,例如患者一般会有多次化疗记录,但每个患者的化疗次数可能都不相同,因此,系统支持用户在填写化疗记录的时候动态添加记录数。

为了方便用户自由设计表单,系统提供了强大的 CRF 可视化配置界面及数据字典维护工具。在设计表单的时候,系统支持模块化管理,用户设置好模块内容后,该模块可以发布出来以供重复利用。为保证数据值域的一致性,系统提供了大量基于标准数据集导入的数据元供用户使用,用户可以方便地拖拽数据元至相应的表单单元格中。在使用数据集标准的同时,要对其进行维护更新,因此系统还提供了简单易操作的数据字典维护工具。

2.3 随访数据共享

对肿瘤患者来说,随访数据的质量和随访率直接影响临床科研的质量,绝大部分研究都与患者的转归息息相关。目前大多数科研随访资料是由医生个人收集整理,形式多样,安全性较差,且医生本身医疗任务很重,没有时间逐一对患者进行随访和辨

别,效果很不理想,因此,建立一套合理的随访机制,既减轻医疗工作者的负担,又保证了随访的效率^[9]。科研系统支持各病种的专科随访,同时,也可以与我院的常规随访系统共享数据。我院的随访中心负责全院的常规随访,为了防止对患者进行二次随访,引起患者的反感,科研系统支持多个来源随访数据的接入与共享。用户在做单病种专科随访的时候,可以查看到常规随访的数据,用作参考;常规随访系统也可以查看到专科随访的数据,实现数据的共享,提高数据的利用率。同时,该系统支持随访方式的多样性,不管随访数据的来源如何,都可以在 CRF 填写时用作参考。例如,我院常规随访采用病史随访加电话随访的方式;前列腺病种采取门诊随访的方式,在患者术后复诊的时候医生会录入一份随访文书;乳腺病种采取患者微信主动随访的方式。单病种数据库会接入所有的随访数据,在科研系统里做专科随访的时候用作参考。

2.4 存在的不足

临床科研系统在为医院临床科研提供有效服务的同时,还存在一些不足。首先,内部临床业务系统结构化程度不够,诊疗记录及医技报告大都是以文本的格式存储,单病种数据库虽然做了一部分后结构化处理,将诊疗记录按照 XML 格式解析,将病理报告按照病种模板格式解析,但是解析出的字段仍然是不规范的文本格式,不同医生的写法不同,字符格式不同,都会造成字段无法统计。因此,根本的解决办法是将前端业务系统进行结构化改造,实现电子病历及医技报告结构化。以此为基础,单病种科研系统才能提供更多更有效的数据项供临床科研使用。其次,科研系统提供数据导出功能,用户可根据自己的需要将病例筛选出来,再将这些病例的诊疗数据导出,但由于系统后台数据库是将不同类别的诊疗数据存放在不同表里,目前系统无法将病例的所有诊疗数据关联导出,需要用户人工关联或者信息管理人员后台导出,后续系统还需要针对这一功能做升级,使得系统可根据前台用户输入的条件自动生成关联脚本将病例诊疗数据一起导出。最后,在统计功能上,系统目前只提供单病种数据库的入库例数、课题数、各课题的病例数等简单的供医院管理人员查看的统计,而针对研究人员使用的更专业的数据统计功能则需要用户将科研数据导出到专业统计分析软件再进一步处理,将来系统可根据用户的

需要开发专业的数据统计分析功能,将大大提高临床医生的科研效率。

3 结论与展望

综上所述,我院的数据平台充分利用了医疗卫生信息资源,建立以患者为中心的临床信息整合及分析解决方案,实现医疗卫生机构的科学化、现代化管理,为数据应用提供了强有力的支持。目前,我院共构建了乳腺、大肠、胃、宫颈、卵巢、脑脊柱、前列腺、肾、膀胱、肝脏、骨软组织及甲状腺肿瘤共 12 个单病种数据库,入库的病例数总共已达到 99 万多,正在开展的课题数共 32 个,已覆盖我院大部分科室。以如此庞大的数据量为基础,医生可随时方便快捷地创建课题、自定义 CRF 表单、筛选符合条件的病例、半自动化录入科研数据项、数据导出,最后进行专业的数据分析。该临床科研系统为我院医生进行学术研究提供了极大的技术支持和数据保证。为了保证数据的一致性,我们遵照国家的各项标准整理了肿瘤通用标准数据集作为支撑,但为了更适用于单病种科研,还需要根据更专业的标准,结合临床医生的工作经验整理更详尽深入的单病种数据子集作为补充,更好地为全院的大数据利用做支持。在健康卫生领域,大数据也能发挥其巨大作用,对于患者的诊疗策略更加需要基于“数据分析”而得出,而非传统的经验和直觉^[10],虽然相比其他领域具有更加棘手的研究问题,但对临床研究的推动更具效率和期望,能够实质性辅助临床研究和管理决策,并将成为信息时代进行大规模临床研究不可缺少的基础条件平台^[11]。因此除了回顾性分析,我们未来还可以利用大数据做数据挖掘,为临床决策提供支持。数据挖掘可以有效地预测临床数据和患者病历信息相关的指标,通过对比大量患者的诊疗数据分析个性化治疗的药物和治疗方案,帮助临床医生迅速、准确地为患者设计出治疗方案,降低医疗风险,减少医疗浪费,充分体现“以病人为中心、以医疗质量为核心”的理念,从而提高医院的整体医学水平和医院的核心竞争力^[12]。

参考文献:

- [1] Yu XY,Zhang HP. Experiences of clinical research

database for tumor specific hospital[J]. Chinese Journal of Medical Science Research Management,2011,24(6):422-423. [于新颖,张焕萍. 肿瘤专科医院临床科研数据库建设探讨[J]. 中华医学科研管理杂志,2011,24(6):422-423.]

[2] Zhang FL,Tao H. Application progress of single disease database technology in the ophthalmic informatization management [J]. Journal of Medical Informatics,2013,34(3):44-46. [张芳丽,陶海. 单病种数据库技术在眼科信息化管理中的应用进展 [J]. 医学信息学杂志,2013,34(3):44-46.]

[3] Lu ZH,He Y,Ju HN,et al. Experiences of clinical data repository (CDR) implementation [J]. China Digital Medicine,2016,11(3):116-118. [陆兆辉,何毅,巨华宁等. 医院临床数据中心(CDR)及应用的建设体会[J]. 中国数字医学,2016,11(3):116-118.]

[4] Wikipedia. Enterprise master patient index[EB/OL]. http://en.wikipedia.org/wiki/Master_Patient_Index,2014-08-20.

[5] Huang ZH,Wang Y,Li Y,et al. Application of single disease data center for clinical research [J]. Chinese Hospital Management,2014,34(2):56-58. [黄宗浩,王奕,李渊等. 单病种数据中心在临床科研中的应用 [J]. 中国医院管理,2014,34(12):56-58.]

[6] WS445-2014. Basic dataset of electronic medical record: medical record summary[S].[WS445-2014. 电子病历基本数据集 第1部分:病历概要[S].]

[7] Wu XF,Wang L,Wang PJ. Construction on clinical data center based on electronic medical records [J]. Journal of

Medical Informatics,2016,37(10):44-47. [吴晓芬,王磊,王培军. 基于电子病历的临床数据中心构建[J]. 医学信息学杂志,2016,37(10):44-47.]

[8] Yu G. The application and design of the clinical scientific research platform based on the medical big data[J]. China Digital Medicine,2016,11(9):15-17. [俞高. 基于医疗大数据的临床科研平台应用设计 [J]. 中国数字医学,2016,11(9):15-17.]

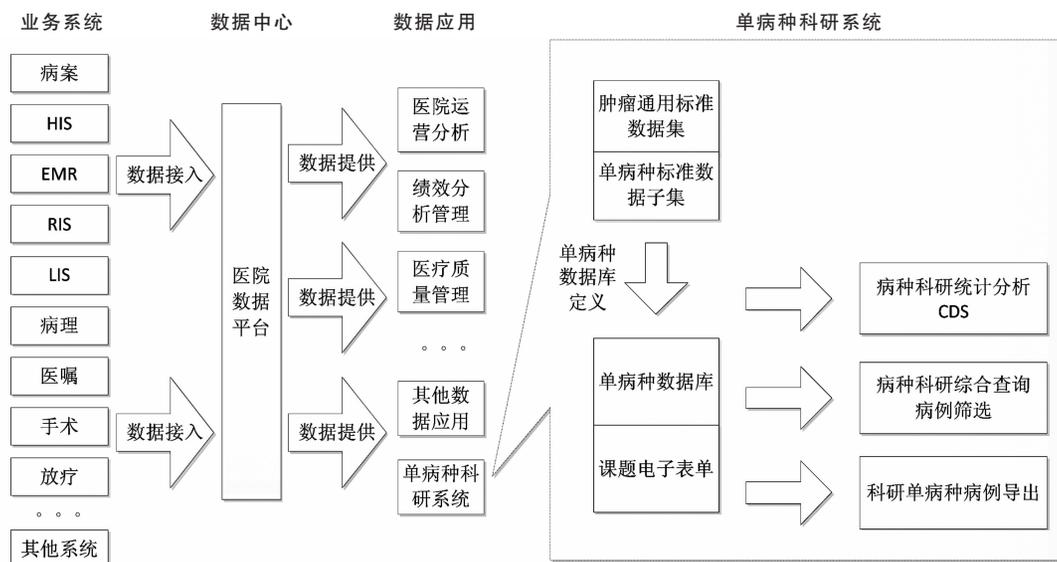
[9] Tang XD,Yan H,Luo J,et al. Application of discharge follow-up system in clinical research [J]. China Digital Medicine,2015,10(5):115-117. [唐晓东,闫辉,罗娟,等. 出院随访系统在临床科研中的应用 [J]. 中国数字医学,2015,10(5):115-117.]

[10] Zhang ZH. Big data and clinical research [J]. Journal of Clinical and Pathological Research,2014,34(5):492-497. [章仲恒. 大数据与临床科研 [J]. 临床与病理杂志,2014,34(5):492-497.]

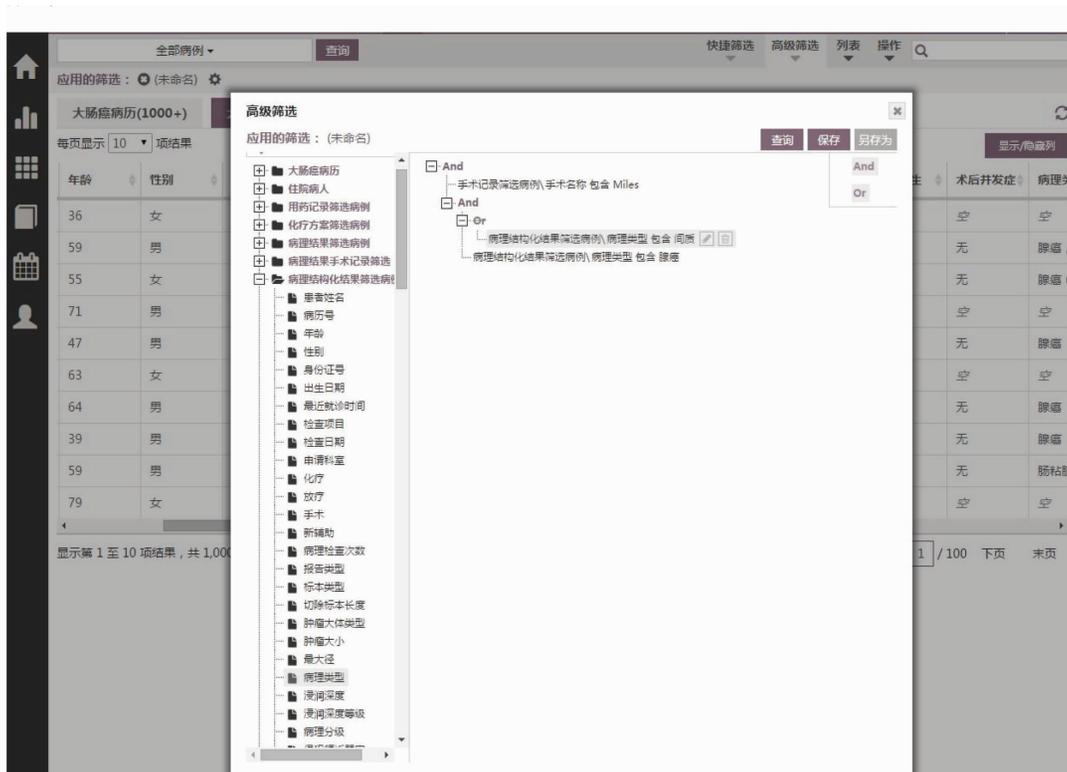
[11] Liu BY,Zhou XZ,Li P,et al. A unified clinical and research information platform toward individualized medicine [J]. China Digital Medicine,2007,2(6):31-36. [刘保延,周雪忠,李平,等. 个体诊疗临床科研信息一体化平台[J]. 中国数字医学,2007,2(6):31-36.]

[12] Zhao Y,Wang Y,Yan GT,et al. Application of clinical decision-making support system based on data warehouse in the hospital[J]. Digitalized Hospital,2016,31(7):95-97. [赵妍,王颖,闫国涛,等. 基于数据仓库的临床决策支持系统在我院的应用[J]. 医院数字化,2016,31(7):95-97.]

附录 A 单病种数据库数据模型



附录 B 多维度数据综合查询



附录 C 半自动填写 CRF 表单

