

MIMIC 数据库在急诊医学临床研究过程中运用的思考

李开源 冯聪 贾立静 陈力 潘菲 黎檀实

100853 北京,解放军总医院急诊科

通讯作者:黎檀实, Email: lts301@sohu.com

DOI: 10.3760/cma.j.issn.2095-4352.2018.05.021

【摘要】 结合重症监护医学信息数据库(MIMIC)的特点和国内外医疗大数据最新研究进展,对MIMIC数据库进行介绍,并详细论述基于MIMIC数据库的急重症临床医学研究方法,提出在急诊医学研究中引入医疗大数据的可行性和必要性。探讨MIMIC数据库在急诊临床医学研究中的作用,以及在医疗大数据背景下试验设计与实施过程中的原则和注意事项。利用MIMIC数据库进行急诊医学的临床研究,可为疾病的早期诊断、风险预警及预后分析提供新的思路,但也具有其局限性,需要进一步开发符合我国国情的急诊医学多源异构数据库,以适应医疗大数据时代的到来,为急诊医学的发展提供新的方向。

【关键词】 MIMIC数据库; 急诊医学; 临床研究; 医疗大数据

基金项目: 国家自然科学基金(81272060, 81701961); 军事医学青年培育项目(16QNP135); 军事医学创新专项(14CXZ005); 军队后勤科研计划项目(BWS14J042, AWS15J004, BWS14J041, 2015XL015, 15WKS03); 解放军总医院大数据研发项目(2017MBD-30)

Discussion of the implementation of MIMIC database in emergency medical study Li Kaiyuan, Feng Cong, Jia Lijing, Chen Li, Pan Fei, Li Tanshi

Department of Emergency, General Hospital of the PLA, Beijing 100853, China

Corresponding author: Li Tanshi, Email: lts301@sohu.com

【Abstract】 To introduce Medical Information Mart for Intensive Care (MIMIC) database and elaborate the approach of critically emergent research with big data based on the feature of MIMIC and updated studies both domestic and overseas, we put forward the feasibility and necessity of introducing medical big data to research in emergency. Then we discuss the role of MIMIC database in emergency clinical study, as well as the principles and key notes of experimental design and implementation under the medical big data circumstance. The implementation of MIMIC database in emergency medical research provides a brand new field for the early diagnosis, risk warning and prognosis of critical illness, however there are also limitations. To meet the era of big data, emergency medical database which is in accordance with our national condition is needed, which will provide new energy to the development of emergency medicine.

【Key words】 MIMIC database; Emergency medicine; Clinical study; Medical big data

Fund program: National Natural Science Foundation of China (81272060, 81701961); Young Military Medicine Practitioner Cultivation Project (16QNP135); Military Medical Innovation Project (14CXZ005); Military Logistics Science Study Planning Program (BWS14J042, AWS15J004, BWS14J041, 2015XL015, 15WKS03); Big Data R&D Project of the PLA General Hospital (2017MBD-30)

随着机器学习技术的不断发展,大数据分析人工智能技术不断成熟,已经在网络、金融、交通、气象等多个领域发挥了重要作用。而在医疗行业,国外的研究者已经通过建立医疗数据库对医疗大数据进行分析利用,并开始对疾病的流行病学、风险预测和分级诊疗等方面展开研究^[1-4],特别是美国麻省理工学院与贝斯以色列女执事医疗中心联合研发的重症监护医学信息数据库(MIMIC)已运转10余年,并对全球的医学研究者免费开放,为全世界急重症研究提供了海量的数据支持,也为急重症的真实世界研究(RWR)提供了“沃土”^[5-6]。我国对医疗大数据的分析研究还处于起步阶段,尤其在急诊医学领域,国内自主研发的医疗数据库几乎为空白。我国人口基数巨大,急诊就诊患者与日俱增,疾病种类繁多,对医疗大数据的分析及研究提出了现实需求。如能将医疗大数据的应用与急诊医学的临床研究有机结合,势必能进一步推动我国急诊医学的发展。

1 对MIMIC数据库的认识

1.1 MIMIC数据库简介: MIMIC数据库是一个多参数、智能化的重症监护数据库,它涵盖了来自不同人群的重症加强治疗病房(ICU)患者住院全程的临床数据资料。目前最新版本的MIMIC-III数据库涵盖了2001至2012年53423例次成年患者(年龄 ≥ 16 岁)的ICU住院信息,还包括了2001至2008年入住ICU的7870例新生儿的临床资料。数据库中包涵38597例成年患者及49785份住院记录。住院患者平均年龄为65.8岁^[5]。该数据库包括临床数据库(clinical data)和生理波形数据库(physiological data)两大部分。每个数据都有对应的时间信息,包括患者出入院时间、实验室检查结果、体征记录、医疗干预记录、护理记录、影像学检查报告以及出院记录等;生理波形数据库记录了床旁监护仪的高分辨率波形数据,如心电、血压、脉搏波以及其他生理参数,如呼吸、血氧饱和度、中心静脉压等。MIMIC数据库将

为流行病学的分析性研究、临床决策的发展以及医学电子设备的研发提供更多样的方法和思路。

1.2 MIMIC 数据库的数据特点: MIMIC 数据库中的数据分为两种类型。① 临床数据: 临床数据包括由护士确认并实时记录的生理指标(如每小时的心率、动脉血压、肺动脉压等)、护理记录及病程记录、连续静脉给药记录、体液平衡情况、患者的人口统计学资料(如年龄、性别、种族等)、影像学检查结果、医嘱、出院摘要、全面的实验室检查结果(如血生化、血常规、动脉血气、尿常规、微生物检查等)。值得注意的是,上述资料不仅局限于患者在 ICU 期间,而是整个住院期间的临床数据^[6]。② 生理学数据: 生理学指标的获取来自每位患者床旁监护系统仪器的记录。每台监护仪获取多参数的生理数据并将其数字化,然后将心率、血压、血氧饱和度等信号分别处理。每分钟记录 1 次生理波形(如心电图、血压、脉搏体积描记图、呼吸)的趋势图,通过上传形成生理波形数据库。数据分别来自心内科 ICU、心脏外科术后 ICU、内科 ICU、外科 ICU 及创伤外科 ICU。

2 MIMIC 数据库在急重症临床研究中的应用

2.1 临床大数据背景下的临床研究与随机对照临床试验(RCT)的区别: 利用 MIMIC 数据库的临床研究更接近 RWR。RWR 是指研究数据来自真实的医疗环境,反映治疗过程和真实环境下患者的健康状态。RWR 的数据来源可以是患者在整个诊治和康复过程中产生的庞大数据。RCT 作为最高等级的临床证据标准,需要严格控制试验条件,在目标样本中给予标准化临床干预。RCT 首先需要设定严格的入选和排除标准,强调的是标准化样本,即纳入同质患病人群。这种高度选择限制了许多 RCT 结果的应用。这些研究中纳入的患者往往代表性差,导致 RCT 所获得的结果很难体现疾病的本质,其外部真实性有一定局限性^[7]。而 RWR 可以纳入合并多种疾病的重症患者,治疗方式的分配为非随机式,更具有临床现实意义。利用大数据进行临床研究可以有效补充在 RCT 中被排除的病例,如患有多种合并症的患者,使得临床研究更为全面和准确^[8-9]。

2.2 医疗大数据背景下的急诊医学临床研究: 急诊医学是医学上唯一以时间描述来命名的学科,“急”字充分体现了这门学科的紧迫性特点。急诊医学疾病种类繁多,病情复杂危重,更为其临床实践增加了难度。面对急重症患者,医生需要准确判断并迅速给予治疗,这就需要医护人员积累大量的临床经验,然而许多治疗手段却缺乏有效的证据支持。利用人工智能和机器学习技术对医疗大数据进行分析,可建立一系列临床辅助决策支持系统,该系统对数据的集成、搜索、处置能力强大,可以对急重症患者进行早期风险预警、病情动态评估及治疗方案优化,从而为急诊医生的临床决策提供参考和支持。

2.3 利用临床大数据选择课题的角度: 以往的临床研究经常关注强烈的因果关系,而基于大数据的风险预测分析研究还有待开发。不同于因果分析,预测分析更适合应用医疗大数据。急诊医学或许可从这种更简便地应用大数据的方式

中获益^[10]。预测分析作为一种简便的启发式的危险分层方法,在急救护理过程中扮演着重要的角色。这些分析工具都具有共同的特征:标准简洁、易于计算、对于特定人群独立有效。应用医疗大数据可以将海量的信息转化为更为精确的临床预警工具,这些临床预警工具最大的价值在于就诊早期对患者疾病的诊断和预后等给予准确的参考意见。

MIMIC 数据库是全球唯一免费使用的重症数据库,并已运转 10 余年,其中有详实的患者临床资料,世界各地的研究者只需要获得许可即能自由使用。Dunitz 等^[11]利用 MIMIC 数据库研发了一种实时算法,通过将感染性疾病患者进行危险分层,在患者发展为脓毒性休克前进行危险归类。他们通过提取实时记录的动脉血压、心率等连续数据,并评估心排血量和外周阻力,利用二元辨别分析和受试者工作特征曲线(ROC)等方法,研发了多种危险分层工具,可有效预测高乳酸血症和循环衰竭的发生,提高分诊的效率和准确度,并对患者的病情进行了准确评估。Mandelbaum 等^[12]对 MIMIC 数据库进行提取分析,探讨肌酐升高、尿量、住院病死率与肾脏替代治疗之间的联系,为急性肾损伤(AKI)的分类和肾功能衰竭的预警及早期干预提供了新的思路。因此,通过缜密的试验设计和合理的统计学方法,可充分利用 MIMIC 数据库研发出多种操作性强、准确率高、快速有效的实时危险预警系统,这也为试验设计的选题提供了多种思路和参考。

2.4 各种暴露因素的选择与数据提取: MIMIC 数据库含有庞大的数据量,如何选取合适的暴露因素尤为重要。首先,在选择暴露因素时要与数据库结构特点相适应,全面了解数据库中哪些种类的参数指标,再从临床数据库及生理波形数据库中有针对性地提取。在初步选择过程中可尽量多地纳入指标。如 Henry 等^[13]在研究脓毒性休克早期预警评分时纳入 54 项指标,再通过监督式学习方法进行建模,建立了 TREWScore 评分,在特异度为 67% 时,敏感度达到 85%,可有效预警急重症患者脓毒症的发生,为脓毒症的早期干预提供了可能性。Lee 等^[14]在研究体液平衡与预后关系时,按照研究需要及数据特点分类,并根据体液平衡四分位数(可在数据库中计算得到)进行分层,将机械通气、血管活性药物使用、AKI 发生及肾小球滤过率 < 60 mL/min 设为二元变量,将种族及 ICU 类型设为多重变量,将年龄、简化急性生理学评分(SAPS)、实验室指标、静脉输液量、住院时间设为连续变量,再通过不同的统计学方法分析,得出体液平衡与各项预后的流行病学关系。其次,暴露因素的选择也要“量体裁衣”,应选择数据库中可以得到数据,同时又要符合试验设计的要求。

从 MIMIC 数据库中提取数据一般通过两种方式:① 研究者可以获取一种临床数据的平面文件文本,然后应用相关的数据库大图,根据自身的需求来重建数据库。② 研究者可以通过有密码保护的服务器进入数据库。这需要使用者熟悉数据库的结构,并应用结构化的查询语言对数据进行搜索,从而提取数据,然后通过统计学或机器学习等方法,在线下进行分析^[6]。然而从 MIMIC 数据库中提取数据是一

个十分复杂的过程,这就需要临床医生、统计学专家及相关的数据库专家通力合作,相互补充专业盲点,在数据挖掘过程中不断反馈与完善试验设计,才能获取到准确数据,然后选择相适应的统计学或机器学习方法,进行系统分析,从而获得预期结果^[15]。

3 利用 MIMIC 数据库进行急诊医学临床分析的局限性及展望

3.1 MIMIC 数据库的局限性:首先, MIMIC 数据库中包含的数据全部为患者住院期间数据,缺乏住院前门诊治疗数据资料。尤其是 ICU 重症患者,其中很大一部分来自急诊科,而急诊科的治疗及护理数据对于急重症患者的风险预后评估是十分重要的,缺乏相关临床资料会增加较多的混杂因素,使得研究结果的可靠性下降。因此, MIMIC 数据库横断面式的结构就决定了其在急重症临床研究中存在不可避免的局限性。其次, MIMIC 数据库中的医疗数据全部来源于美国波士顿的贝斯以色列女执事医疗中心^[6],患者大部分来自欧美国家,所以治疗及护理过程是基于欧美人群体质进行的,如体液平衡、治疗选择、预后评估等,是否符合我国人群体质特点还有待商榷。

3.2 展望:利用 MIMIC 数据库进行急诊医学临床研究,尤其是预测性分析,可有效地帮助临床医生快速诊断、风险预警及预后分析。我国人口基数巨大,急重症患者数量多、病情复杂,使得急诊科工作较繁重,应用大数据分析的方法可使急诊医学临床实践更加快速有效。但是, MIMIC 数据库的应用也有其局限性,尽早建立符合我国人群特征的专属于急诊科的多源异构数据库就显得尤为重要。国外已有相关研究利用大数据对急诊医学的流行病学、早期预警及风险预测进行了分析^[16-18]。我国目前也有相关探索,如探讨利用网络技术将院前抢救措施、生命体征等信息与院内急诊科共享,实现不同救治阶梯的一体化,并进一步收集、储存数据,服务于科研、培训及临床实践^[19],或应用经典的、非特异性的预警评估系统对特定人群和疾病进行早期预警及风险评估^[20],但仍处于起步阶段,且缺乏大数据支撑。2016年5月,中共中央、国务院发布了《国家创新驱动发展战略纲要》,在战略任务中明确提出“促进组学和健康大数据研究”。利用大数据进行临床研究是大势所趋,也将为急诊医学的发展提供新的动力,必将在急诊医学的临床决策、流行病学研究、急诊分级诊疗等各个方面发挥重要作用。

参考文献

[1] Danner OK, Hendren S, Santiago E, et al. Physiologically-based, predictive analytics using the heart-rate-to-systolic-ratio significantly improves the timeliness and accuracy of sepsis prediction compared to SIRS [J]. *Am J Surg*, 2017, 213 (4): 617-621. DOI: 10.1016/j.amjsurg.2017.01.006.

[2] Ofli F, Meier P, Imran M, et al. Combining human computing and machine learning to make sense of big (aerial) data for disaster response [J]. *Big Data*, 2016, 4 (1): 47-59. DOI: 10.1089/big.2014.0064.

[3] Ryan J, Hendler J, Bennett KP. Understanding emergency department 72-hour revisits among medicaid patients using electronic healthcare records [J]. *Big Data*, 2015, 3 (4): 238-248. DOI: 10.1089/big.2015.0038.

[4] Brixner D, Ghate SR, McAdam-Marx C, et al. Association between cardiometabolic risk factors and body mass index based on diagnosis and treatment codes in an electronic medical record database [J]. *J Manag Care Pharm*, 2008, 14 (8): 756-767. DOI: 10.18553/jmcp.2008.14.8.756.

[5] Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database [J]. *Sci Data*, 2016, 3: 160035. DOI: 10.1038/sdata.2016.35.

[6] Saeed M, Villarreal M, Reisner AT, et al. Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database [J]. *Crit Care Med*, 2011, 39 (5): 952-960. DOI: 10.1097/CCM.0b013e31820a92c6.

[7] 李敏, 时景璞, 于慧. 真实世界研究与随机对照试验、单病例随机对照试验在临床治疗性研究中的关系比较 [J]. *中华流行病学杂志*, 2012, 33 (3): 342-345. DOI: 10.3760/cma.j.issn.0254-6450.2012.03.021.

Li M, Shi JP, Yu HH. Relationship between the "real world" research, randomized controlled trial and number of one randomized controlled trial in clinical therapeutic study [J]. *Chin J Epidemiol*, 2012, 33 (3): 342-345. DOI: 10.3760/cma.j.issn.0254-6450.2012.03.021.

[8] Hudis CA. Big data: are large prospective randomized trials obsolete in the future? [J]. *Breast*, 2015, 24 Suppl 2: S15-18. DOI: 10.1016/j.breast.2015.07.005.

[9] Chatterjee S, Davies MJ, Khunti K. What have we learnt from "real world" data, observational studies and meta-analyses [J]. *Diabetes Obes Metab*, 2018, 20 Suppl 1: 47-58. DOI: 10.1111/dom.13178.

[10] Janke AT, Overbeek DL, Kocher KE, et al. Exploring the potential of predictive analytics and big data in emergency care [J]. *Ann Emerg Med*, 2016, 67 (2): 227-236. DOI: 10.1016/j.annemergmed.2015.06.024.

[11] Dunitz M, Verghese G, Heldt T. Predicting hyperlactatemia in the MIMIC II database [J]. *Conf Proc IEEE Eng Med Biol Soc*, 2015, 2015: 985-988. DOI: 10.1109/EMBC.2015.7318529.

[12] Mandelbaum T, Lee J, Scott DJ, et al. Empirical relationships among oliguria, creatinine, mortality, and renal replacement therapy in the critically ill [J]. *Intensive Care Med*, 2013, 39 (3): 414-419. DOI: 10.1007/s00134-012-2767-x.

[13] Henry KE, Hager DN, Pronovost PJ, et al. A targeted real-time early warning score (TREWScore) for septic shock [J]. *Sci Transl Med*, 2015, 7 (299): 299ra122. DOI: 10.1126/scitranslmed.aab3719.

[14] Lee J, de Louw E, Niemi M, et al. Association between fluid balance and survival in critically ill patients [J]. *J Intern Med*, 2015, 277 (4): 468-477. DOI: 10.1111/joim.12274.

[15] 王剑, 张政波, 王卫东, 等. 基于重症监护数据库 MIMIC-II 的临床数据挖掘研究 [J]. *中国医疗器械杂志*, 2014, 38 (6): 402-406. DOI: 10.3969/j.issn.1671-7104.2014.06.002.

Wang J, Zhang ZB, Wang WD, et al. Clinical data mining by exploring public MIMIC-II intensive care database [J]. *Chin J Med Inst*, 2014, 38 (6): 402-406. DOI: 10.3969/j.issn.1671-7104.2014.06.002.

[16] Ram S, Zhang W, Williams M, et al. Predicting asthma-related emergency department visits using big data [J]. *IEEE J Biomed Health Inform*, 2015, 19 (4): 1216-1223. DOI: 10.1109/JBHI.2015.2404829.

[17] Taylor RA, Pare JR, Venkatesh AK, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach [J]. *Acad Emerg Med*, 2016, 23 (3): 269-278. DOI: 10.1111/acem.12876.

[18] Frischknecht CE, Berlac PA, Nielsen H, et al. The Danish quality database for prehospital emergency medical services [J]. *Clin Epidemiol*, 2016, 8: 667-671. DOI: 10.2147/CLEP.S100919.

[19] 陈意飞, 张劲松. 院前院内急诊信息共享系统的构建与探讨 [J]. *中华危重病急救医学*, 2017, 29 (3): 268-269. DOI: 10.3760/cma.j.issn.2095-4352.2017.03.015.

Chen YF, Zhang JS. Construction and outlook of sharing information in pre-hospital emergency and emergency disposal in hospital [J]. *Chin Crit Care Med*, 2017, 29 (3): 268-269. DOI: 10.3760/cma.j.issn.2095-4352.2017.03.015.

[20] 于云鹏, 司君利, 刘冠群, 等. 国家早期预警评分系统评估急诊老年危重症患者死亡风险的有效性研究 [J]. *中华危重病急救医学*, 2016, 28 (5): 387-390. DOI: 10.3760/cma.j.issn.2095-4352.2016.05.002.

Yu YP, Si JL, Liu GQ, et al. A validation study of national early warning score in evaluation of death risk in elderly patients with critical illness [J]. *Chin Crit Care Med*, 2016, 28 (5): 387-390. DOI: 10.3760/cma.j.issn.2095-4352.2016.05.002.