

· 专题二：双清论坛“面向未来的中国医学—免疫视角下的中西医融合之道” ·

健康数据科学驱动的中西医融合研究^{*}

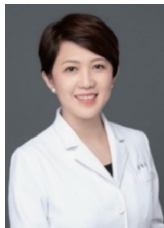
彭苏元^{1,2} 张 腊³ 杨 超^{4,5} 武之越¹
王福琳¹ 周 亮¹ 刘旭生^{3**} 张路霞^{1,4,5**}

1. 北京大学 健康医疗大数据国家研究院, 北京 100191
2. 北京大学 公共卫生学院, 北京 100191
3. 广州中医药大学 第二临床医学院/广东省中医院肾病科/
省部共建中医湿证国家重点实验室, 广州 510120
4. 北京大学第一医院 肾内科/北京大学肾脏疾病研究所/
卫生部肾脏疾病重点实验室/慢性肾脏病防治教育部重点实验室, 北京 100034
5. 浙江省北大信息技术高等研究院, 杭州 311215

[摘要] 中、西医作为两种不同的理论体系,在传统的学科融合研究中较难建立统一的知识与认识体系。随着大数据时代的到来,“健康数据科学”这一新兴交叉学科的蓬勃发展为变革中西医融合研究范式提供了新的契机。本文从数据密集型科学研究范式,健康数据科学、环境健康和中西医结合医学的跨学科背景出发,在宏观视角下运用“三因制宜”的观点,论述多疾病谱与多维度数据驱动的知识发现问题,为推动中西医结合的理论发展与知识融合提供新模式与新途径,也为健康数据科学在医学中的应用提供可推广范例。

[关键词] 中西医融合;健康数据科学;三因制宜;本体

中医是中国古代哲学思想的集中体现,而西医是从西方自然科学发展而来的;前者侧重于整体,而后者则注重“组织—细胞—分子”的机制解析。由于二者观念不同,使得他们的思维模式有较大的区别;在融合发展过程中,存在着难以沟通、“结而不合”等诸多问题。如果单纯地采用西方自然科学的语言和方法来诠释我国传统医学,生搬硬套地进行临床实践,“照葫芦画瓢”地解释临床现象,将会掩盖中医学独特的医学魅力,禁锢中医学特有的思维模式,继而



张路霞 北京大学健康医疗大数据国家研究院副院长、教授、主任医师、博士生导师,国家杰出青年科学基金获得者。主要研究方向为大数据驱动的重大疾病防控管理、健康医疗大数据资源治理与应用。



刘旭生 广东省中医院肾病大科主任、教授、主任医师、博士生导师,广东省名中医。国家中医临床研究基地重点病种(慢性肾脏病)负责人;国家中医药管理局区域诊疗中心(华南)负责人;全国第二批名老中医学学术继承人。主要研究方向为慢性肾脏病中西医结合防控与管理。



彭苏元 中医内科学博士,北京大学健康医疗大数据国家研究院博士后、助理研究员。长期从事健康数据科学与肾脏疾病负担、重大慢性病管理的交叉学科研究。

进一步减弱中西医结合医学的国际影响力。

目前,中西医结合诊疗模式尚未形成一个公认的国际标准,未来中西医可合作的领域仍存在大量

收稿日期:2022-05-20;修回日期:2023-02-01

* 本文根据第303期“双清论坛”讨论的内容整理。

** 通信作者,Email:zhanglx@bjmu.edu.cn;liuxu801@126.com

本文受到国家自然科学基金项目(72125009,81873261,82174308),省部共建中医湿证国家重点实验室开放课题(SZ2021KF10)的资助。

空白。因此,促进中西医结合事业发展的关键是要加强中医药的国际认可度,提高中医药诊治与评价的正确性、科学性和可重复性,增强人们对中医药临床治疗的安全性和有效性的认识,加快中医药“走出去”的步伐,进一步深化中西医结合的进程。

当今全球医疗卫生系统面临着人口老龄化、医疗资源匮乏、医疗费用高昂等问题;电子医疗数据的高速增长,以及新兴技术方法的不断涌现,为解决这些问题提供了新的思路和路径。健康数据科学是应运而生的一门新兴交叉学科,旨在采用跨学科的思路和方法挖掘健康医疗数据的价值、结合医学领域知识形成洞察力、最终赋能健康医疗实践。因此,本文从中西医融合现有研究范式和数据密集型科学发现范式入手,紧密围绕健康数据科学的内涵、详细论述多疾病谱视角与多维度效应数据驱动知识发现,希望对构建中西医结合诊疗模式起到积极的推动作用。

1 我国中西医融合现有研究范式

中、西医的思维模式存在较大区别。在日常诊疗中,西医强调“病”,以解剖学、生理学和病理学等理论为基础,通过病史采集、系统查体以及实验室辅助检查,综合临床资料对疾病进行鉴别诊断。中医则更多强调“证”,诊断时以“整体观”与“辨证论治”思想为基础,通过“望、闻、问、切”四诊合参了解疾病的病因、病位和病性,探讨外在表征与内在脏腑经络的联系,充分阐述人体与疾病的相互关系。中医思维以整体观念为核心,后者是指事物与事物以及事物内部之间都存在紧密关联;是中医学对中国传统哲学思想的继承与发展,体现了中医运用宏观、联系的思维看待生命健康与疾病的特点。

我国中西医融合的发展是一个从实践到认识、再从认识到实践,不断循环往复的过程^[1]。当前中西医融合的基本模式可分为:医学教育层面的相互渗透、医疗实践层面的相互补充,以及医学理论方面的相互融合^[2]。在医疗实践中,往往多遵循西医辨病、中医辨证的方式,通过病证结合的模式,分别制定基于循证指南的西医治疗方案及个体化的中医辨证处方^[3]。在科学研究层面,中医界逐步走向循证化的道路,包括广泛开展各类临床试验、中药成分药理等分析,同时对各类疾病名称、诊断指标、疗效判定实施标准化处理,并适时地对传统中药剂型进行改良等等^[4]。时至今日,中西医融合的最优方式仍然在不断摸索。

但是,目前中西医融合存在两个主要问题。首先在认识论层面上,两个学科知识体系存在巨大的差异,难以实现知识间的直接关联推理^[5]。中西医诊疗术语的差异性,自西方医学传入中国以来就一直存在,而两者之间的差异性中西医结合及其发展的重要阻碍。中医学理论有着独特的体系架构,其知识背景,无论是哲学起源、还是中医药术语的语义和语境,都与西医的知识体系有着极大的差异。“坚持中西医并重”,实施诊疗术语的规范化、标准化成为解决问题的关键所在。其次,中西医知识本体存在差异,难以建立两个学科体系间的关系。强行利用西医理论去诠释中医药概念和术语、忽略其发展的规律和特殊性,在研究建立基于某个顶层本体的中医药本体知识库时将会产生语义表达二义性的问题。因此,中西医融合需要在认识论层面形成知识本体连接,在当前“生物—心理—社会医学”模式下,结合外部视角(如环境气候、社会学等因素)打破学科间藩篱,创新中西医融合的研究范式。

2 健康数据科学的兴起和发展

2.1 健康数据科学学科内涵

在当今的中国,人民生命和健康得到前所未有的关注,而健康守护也面临着巨大的挑战。与此同时,大数据、人工智能、区块链等技术正以更快的速度渗透于医学,其研究范式与实践模式也在不断创新与发展。在数字化驱动健康医疗模式变革之际,“健康数据科学(Health Data Science)”作为一门新的学科应运而生。健康数据科学是“数据科学”学科在医疗健康领域的分支,数据科学是一套从数据中提取信息和知识的指导原则,其内涵不仅限于数据挖掘算法,更重要的是如何从数据角度看待所要解决的业务问题^[6,7]。健康数据科学具有多学科深度交叉、随科技发展动态变化的特征;其学科内涵是“采用跨学科的思路和方法挖掘健康医疗数据的价值,结合领域知识形成洞察力,并最终赋能健康医疗实践”。健康数据科学不仅仅是关注数据本身,更加关注经由数据衍生的价值,以及通过数据促进医学的发展^[8]。

自2014年起,全球各高校相继设立“健康数据科学”专业,进行跨学科人才培养,将数据科学应用于健康医学,使其成为一门独立的学科。在我国,北京大学率先顺应科学发展趋势、设立“健康数据科学”这一前沿交叉学科。

2.2 健康数据科学推动医学研究进展

随着健康医疗大数据资源的积累与数据挖掘技

术的进展,从多源多模态的数据中挖掘价值成为可能^[9]。健康数据科学的发展也给医学研究带来了巨大契机,能够更加高效、深度释放健康医疗数据的价值,为现代医学研究提供了新的思路和方法^[10]。

除了临床医疗服务过程中产生的海量数据,多组学数据(前沿技术驱动)、行为数据(社交媒体和可穿戴设备产业驱动)、愈加精细化的环境数据(多学科发展驱动)、以及医疗运营管理数据等,也为医疗行为及疾病关联分析提供了全新的研究角度。医学既是一门自然科学,又具有显著的人文性和社会性。在很多情况下,人文和社会因素与健康结局之间的关系难以在个体层面精确量化、或者精准预测。通过外部视角的引入,结合人文社会学学科的研究方法,有助于发现总体人群的医疗行为与健康结局的关系。例如,在新型冠状病毒肺炎(Coronavirus Disease 2019, COVID-19)疫情防控之中,通过手机移动大数据与疾病数据进行关联,可以准确预测整体人口流动情况与社区传播风险,从而为疫情防控政策制定提供数据支撑^[11]。在全球范围内各个领域都在数字化的今天,如何运用“数据密集型科学发现”研究范式驱动学科间的融合,成为健康数据科学研究的巨大机遇^[8]。

3 健康数据科学驱动中西医融合研究

3.1 健康数据科学的从宏观视角创新科研范式

中医学的认识论来源于东方哲学思想,主要表现为中国传统文化与哲学思想对中医学理论形式、内在结构与基本框架构建的影响。中医学认为人体属于自然环境的一部分,且人与自然是不可分割的,大体可概括为“气一元论”的本体论,“天人合一”的认识论,“三因制宜”的整体观,以及“辨证施治”的诊疗特点。

中医学认为,疾病的发生、发展与预后除患者本身的人口学特点和生活方式外,还受节气时令、气候、地理位置、体质多方面因素的影响,治疗上需顺应结合疾病与季节、地理环境、患者体质特点,采取合适的综合治疗方案^[12]。中医学认为人体是一个有机整体,人与自然环境、社会环境都存在统一性。

中医“天人合一”的整体观与健康数据科学从宏观视角、结合多维度数据创新科研范式的思路高度契合。健康数据科学除关注人群流行病学特征、个体生活习惯等疾病危险因素外,还结合气候变化、人口流动、就医行为等变量综合分析环境因素与社会因素对健康的影响。*The Lancet* 杂志 2022 年 10 月

发布的《柳叶刀人群健康与气候变化倒计时 2022 年报告》指出^[13],过去一年间的极端气候、高温热浪对健康的影响日益加剧,与高温有关的死亡人数显著上升;气候变化在不断影响传染病传播及慢性非传染性疾病预防的同时,也通过多途径影响农业生产与粮食安全,从而进一步影响人类疾病谱。

而 COVID-19 大流行以及俄乌冲突过度依赖化石燃料所造成的能源和社会生活成本危机,进一步加重卫生系统的负担,对健康造成负面影响^[13, 14]。

健康数据科学也可用于分析空间尺度对健康结局的影响。我国医疗资源与城市人口的不均衡,本团队前期基于中国 400 多万患者的住院记录,调查了城市间患者流动的模式和影响异地就诊行为的因素,并对区域诊疗中心设置、公共卫生政策制定等提出建议^[15]。此外,综合时间与空间维度的分析,例如,有研究显示我国 PM_{2.5}, PM₁₀ 等细颗粒物暴露导致的脑卒中、缺血性心脏病和肺癌的死亡人数于 2004—2016 年间增加了 40 万人。而且在时间和地理上都表现出了显著的区域性,以长江三角洲、珠江三角洲和京津冀地区最为突出^[16];健康数据科学通过宏观视角,运用多维度数据从时间、空间、社会学因素等更宏观的角度看待疾病发展与健康结局,从数据科学的角度反映了中医学“天人合一”的认识论,“三因制宜”的整体观的合理性和科学性。

3.2 健康数据科学驱动中西医融合研究—以“三因制宜”理论为例

“三因制宜”是《黄帝内经》理论体系中的一个重要内容,它包括因时制宜、因地制宜和因人制宜,其目的在于阐释“天”“地”“人”三大因素与疾病特征之间的关系,以及上述因素对于治疗方案的影响^[17]。三因制宜理论认为人体与自然界具有相通相应的关系,无论四时气候、地理环境、昼夜晨昏的变化都会影响到人体的生理病理状态。《黄帝内经·素问》中首次提出了六淫邪气概念,即风、寒、暑、湿、燥、火;至南宋时期,陈无择所著的《三因极·病证方论》在六淫概念的基础上,发展总结出“三因学说”,详述了气候、地理环境、人体之间关系失衡的致病特点。

近年来多个临床研究发现,湿热证是多种肾系疾病的常见证型,贯穿于疾病发生、发展及转归的整个病程,可能是慢性肾脏病(Chronic Kidney Disease, CKD)缠绵难愈和反复发作的主要原因^[18]。

CKD 具有高患病率、高致残率、高医疗费用、低知晓率的特点,已成为危害人类健康的重要公共卫生问题。2012 年全国多中心横断面调查结果显示,

化学成分以及药物靶点与基因疾病关联的数据库^[31-33]。通过(DisGeNET)^[34]、人类表型本体(Human Phenotype Ontology, HPO)^[35]、基因本体(Gene Ontology, GO)^[36]、药物银行(Drug Bank)^[37]、一体化医学语言系统(Unified Medical Language System, UMLS)^[38]等本体,对药物靶点、转录组及标准术语进行映射,基于系统生物学的理论,用生物信息学和网络分析方法对生物系统进行分析,从系统水平研究药物作用机制、进行多靶点药物分子设计,从系统的角度和分子水平研究中药方剂的作用机理。

此外,随着开放生物与生物医学本体库(OBO Foundry)^[39]的发展,建立相互兼容和非重复的生物与生物医学本体体系成为可能,目前已建立200余个本体系统^[40, 41]。除了生物信息学、网络药理学常用的HPO、GO外,在中医药领域,目前已展开有通用形式化本体(General Formal Ontology, GFO)、推荐上层合并本体(Suggested Upper Merged Ontology, SUMO)等顶层本体的中医药本体构建研究^[42, 43]。在顶层本体的框架下,朱彦等对经络腧穴相关概念进行梳理、辨析、定义及归类,建立了包含子类关系、表里关系、映射关系等的腧穴本体架构^[44]。杨啸林等^[5]通过梳理中医药领域的重要概念,提出了统一的中医药基础语义框架。通过本体研究可以建立本体间的逻辑框架,有利于实现中医本体同其他领域本体之间的整合。本体构建根本目的是让计算机能理解人的知识,需要将复杂的概念进行辨析、形式化表达,从而达到语义层面的互操作,实现多源数据的整合,进而达到学科融合的目的。

4 总结

健康数据科学驱动的中西医融合研究,并非传统意义上的“中医药现代化”,而是聚焦宏观视角、重视整体观念,逐步建立多维格局的中西医融合数据支撑体系。

在数据层面上,我国已建立多个国家级数据平台、人群队列。首先,国家人口与健康科学数据中心^[45](原国家人口与健康科学数据共享服务平台)于2019年正式获批,该平台是国家科技基础条件平台组成部分,承担国家人口健康领域科学数据整合汇交、审核、加工、保存、挖掘、认证和共享服务任务,保障人口健康领域科学数据的保存和管理。目前已集成生物医学、基础医学、临床医学、药学、公共卫

生、中医学、人口与生殖健康等多方面的科学数据资源^[46, 47]。截至发稿时间,共汇交重点研发计划项目1100余个,约15000个数据集。其次,在人群队列和生物样本库建设方面,“十三五”精准医学专项资助建设了7个百万级区域自然人群队列以及15个专病队列;生物样本库建设则主要是以医院患者资源为主体,截至2022年8月,获得科技部中国人类遗传资源保藏行政许可的临床生物样本库共156个^[48]。当前,已初步具备基于“数据密集型科研范式”开展多学科融合研究的条件。

中西医结合是一项独具特色的创新性研究,除数据基础外,还需要研究人员具备较高的科学素养:首先,需要了解中、西医两个截然不同医学体系对于事物的认识论以及研究的方法论,只有深刻理解他们的本质内涵,才能创新新思路,产生新方法,既要有创新思维,又不能牵强附会;第二,要积极运用多学科的科学理论与方法,借鉴信息技术、地球科学、人文社会科学、哲学等不同学科的前沿理论和技术手段,促使中西医结合研究工作快速发展。

医学既是自然科学又具备社会性与人文性,中、西医交叉融合趋势愈发明显,知识体系间的差异不应只从自然科学角度进行阐释。通过“数据密集型科学发现”研究范式,利用跨学科的思路和方法挖掘健康医疗数据的价值、结合领域知识形成洞察力,可以从社会性与人文性的角度构建知识与健康本质关联,发挥健康数据科学应有的作用。

参 考 文 献

- [1] 丛斌,陈香美. 中西医结合的认识论和方法论. 中国中西医结合杂志, 2021, 41(6): 742—747.
- [2] 白长川,王续琨,王硕. 试析中西医结合医学的基本模式. 中西医结合研究, 2018, 10(6): 327—330.
- [3] 佚名. 张伯礼院士:循证医学助力中医药高质量发展. 天津中医药, 2023, 40(1): 3.
- [4] Cheng CW, Wu TX, Shang HC, et al. CONSORT extension for Chinese herbal medicine formulas 2017: recommendations, explanation, and elaboration. *Annals of Internal Medicine*, 2017, 167(2): 112—121.
- [5] Zhu Y, Yao KY, Peng SY, et al. Traditional Chinese medicine (TCM) domain ontology: current status and rethinking for the future development. *Chung-Kuo i Hsueh K'o Hsueh Tsa Chih*, 2022, 37(3): 228—233.
- [6] Provost F, Fawcett T. Data science and its relationship to big data and data-driven decision making. *Big Data*, 2013, 1(1): 51—59.
- [7] Dhar V. Data science and prediction. *Communications of the ACM*, 2013, 56(12): 64—73.

- [8] 王迈, 王福琳, 张路霞. 重视提升临床医生数据素养 面向未来培养临床数据科学家. *中华内科杂志*, 2022, 61(3): 243—245.
- [9] Abraham R, Schneider J, vom Brocke J. Data governance: a conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 2019, 49: 424—438.
- [10] Dash S, Shakyawar SK, Sharma M, et al. Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, 2019, 6(1): 1—25.
- [11] Jia JS, Lu X, Yuan Y, et al. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*, 2020, 582(7812): 389—394.
- [12] 李志更. 历代中医学家对“三因制宜”学术思想的认识. *中国中医基础医学杂志*, 2010, 16(2): 98—100.
- [13] Romanello M, di Napoli C, Drummond P, et al. The 2022 report of the Lancet Countdown on health and climate change: health at the mercy of fossil fuels. *Lancet (London, England)*, 2022, 400(10363): 1619—1654.
- [14] Jay O, Capon A, Berry P, et al. Reducing the health effects of hot weather and heat extremes: from personal cooling strategies to green cities. *Lancet (London, England)*, 2021, 398(10301): 709—724.
- [15] Ding JQ, Yang C, Wang YY, et al. Influential factors of intercity patient mobility and its network structure in China. *Cities*, 2023, 132: 103975.
- [16] Liu MM, Huang YN, Ma ZW, et al. Spatial and temporal trends in the mortality burden of air pollution in China: 2004—2012. *Environment International*, 2017, 98: 75—81.
- [17] 姜青松, 王庆其. 三才思想与《黄帝内经》的三因制宜. *中华中医药学刊*, 2015, 33(9): 2065—2067.
- [18] 戴恩来, 李小会, 余仁欢, 等. 从湿热论治肾脏病的思路与方法. *中国中西医结合肾病杂志*, 2021, 22(5): 468—470.
- [19] Zhang LX, Wang F, Wang L, et al. Prevalence of chronic kidney disease in China: a cross-sectional survey. *The Lancet*, 2012, 379(9818): 815—822.
- [20] Wang FL, Wan WZ, Peng SY, et al. Effects of ambient temperature on hospital admissions for obstructive nephropathy in Wuhan, China: a time-series analysis. *Ecotoxicology and Environmental Safety*, 2022, 242: 113876.
- [21] 周润津, 庞博, 刘兴华, 等. 岭南地区湿热证研究进展. *广州中医药大学学报*, 2021, 38(5): 1071—1076.
- [22] 彭立, 黄杨, 吕崇山, 等. 基于湿邪致痰瘀互结探讨冠心病的防治思路. *中医杂志*, 2017, 58(22): 1978—1980.
- [23] 国务院联防联控机制综合组. 关于在新型冠状病毒感染医疗救治中进一步发挥中医药特色优势的通知. (2023-01-02)/[2023-02-10]. http://www.gov.cn/xinwen/2023-01/05/content_5735114.htm.
- [24] 北京市中医管理局, 李秀惠, 刘清泉, 等. 北京市新型冠状病毒肺炎中医药防治方案(试行第六版). *北京中医药*, 2022, 41(5): 509—511.
- [25] 四川省中医药局. 四川省新型冠状病毒肺炎中医药防控技术指南(第十一版). (2022-11-22)/[2023-02-01]. <https://www.sc.gov.cn/10462/10778/10876/2022/11/22/0662b513f12f49f2862113e2140d6154.shtml>.
- [26] 陕西省中医药管理局. 关于印发陕西省新冠病毒感染中医药预防方案(第三版)和治疗方案(试行第四版)的通知. (2022-12-13)/[2023-02-01]. http://atcm.shaanxi.gov.cn/sy/tzgg/202212/t20221213_2268825.html?mobile=80GJ8Cgd
- [27] 广东省中医药局. 广东省新冠肺炎中医药防治方案. (2022-12-09)/[2023-02-01]. http://szyyj.gd.gov.cn/zwgk/gsgg/content/post_4060696.html.
- [28] 中华人民共和国国家卫生健康委员会. 关于印发新型冠状病毒感染防控方案(第十版)的通知. (2023-01-07)/[2023-02-01]. <http://www.nhc.gov.cn/xcs/zhengcwj/202301/bdc1ff75feb94934ae1dade176d30936.shtml>.
- [29] Wang YH, Cheng D, Wang ZR, et al. F2-bubbles: faithful bubble set construction and flexible editing. *IEEE Transactions on Visualization and Computer Graphics*, 2022, 28(1): 422—432.
- [30] 徐晓花, 翁衡, 何泽慧, 等. 基于“三因制宜”理论探讨两地 COVID-19 的中医证治特点. *中国实验方剂学杂志*, 2022, 28(4): 172—180.
- [31] Wu Y, Zhang FL, Yang K, et al. SymMap: an integrative database of traditional Chinese medicine enhanced by symptom mapping. *Nucleic Acids Research*, 2019, 47(D1): D1110—D1117.
- [32] Xue RC, Fang Z, Zhang MX, et al. TCMID: traditional Chinese medicine integrative database for herb molecular mechanism analysis. *Nucleic Acids Research*, 2013, 41(D1): D1089—D1095.
- [33] Ru JL, Li P, Wang JN, et al. TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *Journal of Cheminformatics*, 2014, 6: 13.
- [34] Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research*, 2020, 48(D1): D845—D855.
- [35] Köhler S, Gargano M, Matentzoglou N, et al. The human phenotype ontology in 2021. *Nucleic Acids Research*, 2021, 49(D1): D1207—D1217.
- [36] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nature Genetics*, 2000, 25(1): 25—29.
- [37] Wishart DS, Knox C, Guo AC, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 2006, 34(suppl_1): D668—D672.
- [38] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 2004, 32(S1): D267—D270.
- [39] Jackson R, Matentzoglou N, Overton JA, et al. OBO Foundry in 2021: operationalizing open data principles to evaluate ontologies. *Database*, 2021, 2021: baab069.

- [40] Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 2007, 25 (11): 1251—1255.
- [41] 何勇群, 余红, 杨啸林, 等. 本体: 生物医学大数据与精准医学研究的基础. *生物信息学*, 2018, 16(1): 7—14.
- [42] 龙海, 贾李蓉, 朱玲, 等. 顶层本体 GFO 在中医药领域本体构建中的应用展望. *中国中医药图书情报杂志*, 2016, 40 (1): 15—20.
- [43] 高成勉, 包含飞, 周强. 本体构建原则及其在中医顶层本体构建中的应用. *医学信息*, 2008, 21(5): 581—583.
- [44] 付璐, 李宝金, 姚克宇, 等. 顶层本体指导下的经络腧穴语义框架构建探索研究. *中国针灸*, 2022, 42 (9): 1064—1072.
- [45] 中华人民共和国科学技术部. 科技部 财政部关于发布国家科技资源共享服务平台优化调整名单的通知. (2019-06-05)/[2023-02-01]. https://www.most.gov.cn/xxgk/xinxifenlei/fdzdgnr/qtwj/qtwj2019/201906/t20190610_147031.html
- [46] 国家人口健康科学数据中心. 国家人口健康科学数据中心年度发展大事记. [2023-02-01]. <https://www.ncmi.cn/phda/support.html?type=aboutus>.
- [47] 刘德培. 人口与健康科学数据共享//中华医学会医学工程学分会第十次学术年会暨2009中华临床医学工程及数字医学大会论文集. 南京:中华医学会中国生物医学工程学会, 2009:14—23.
- [48] 科学技术部政务服务平台. [2023-02-01]. <https://fuwu.most.gov.cn/html/jgcx/xzxxks/>.

Research on Integration of Chinese and Western Medicine Driven by Health Data Science

Suyuan Peng^{1,2}

La Zhang³

Chao Yang^{4,5}

Zhiyue Wu¹

Fulin Wang¹

Liang Zhou¹

Xusheng Liu^{3*}

Luxia Zhang^{1,4,5*}

1. *National Institute of Health Data Science, Peking University, Beijing 100191*

2. *School of Public Health, Peking University, Beijing 100191*

3. *The Second Clinical Medical College/Department of Nephrology of Guangdong Provincial Hospital of Chinese Medicine/State Key Laboratory of Dampness Syndrome of Chinese Medicine, Guangzhou University of Chinese Medicine, Guangzhou 510120*

4. *Renal Division/Peking University Institute of Nephrology/Key Laboratory of Renal Disease, Ministry of Health of China/Key Laboratory of Chronic Kidney Disease Prevention and Treatment (Peking University), Ministry of Education, Peking University First Hospital, Beijing 100034*

5. *Advanced Institute of Information Technology, Peking University, Hangzhou 311215*

Abstract Traditional Chinese medicine (TCM) and western medicine are two different theoretical systems, and it's difficult to establish a unified system of knowledge and cognition between them. With the advent of the big data era, a novel cross-disciplinary - health data science emerges and provides an opportunity to reform the research paradigm of the integration of TCM and western medicine. This paper briefly introduced the interdisciplinary background of health data science, including the data-intensive scientific paradigm, health data science, environmental health, and integrative medicine. Moreover, we applied the “Three Categories of Etiological Factors System” perspective of TCM to discuss the multi-disease spectrum perspective and multi-dimensional data-driven knowledge discovery paradigm. We aim to provide a new model and a new way to promote the theoretical development and knowledge integration of integrated Chinese and western medicine. In addition, it provides a generalizable example for the application of health data science in medicine.

Keywords integration of TCM and western medicine; health data science; three categories of etiological factors system; ontology

(责任编辑 魏鹏飞 姜钧译)