

· 大型队列项目 ·

呼吸系统疾病专病队列研究的标准制定与数据共享

孙一鑫 裴正存 詹思延

100191 北京大学公共卫生学院流行病与卫生统计学系

通信作者: 詹思延, Email: siyan-zhan@bjmu.edu.cn

DOI: 10.3760/cma.j.issn.0254-6450.2018.02.019

【摘要】 目的 慢性阻塞性肺疾病、哮喘、间质性肺疾病和肺血栓栓塞症是重大呼吸系统疾病,严重危害我国居民健康,整合并开展大规模人群队列研究有助于观察疾病的暴露、发病与转归情况。本研究针对我国社区与临床队列资源的多源异构现状,制定呼吸系统疾病专病队列(呼吸专病队列)数据标准,为解决多源异构数据所致共享障碍,以及项目最大程度开展数据交换、整合、共享、储存与利用提供思路与方法。**方法** 呼吸专病队列数据标准制定思路:①学习、参考国际标准,包括临床数据交换标准协会(CDISC)的CDASH模型,观察性医疗结果合作组织(OMOP)的CDM通用数据模型;②整理、归纳所纳入的4个呼吸专病队列资源,评估各队列资源间的同质性与整合的可能性;③专家讨论,建立呼吸专病队列数据标准。**结果** 研究纳入的现有呼吸专病队列变量模块同质性较好,基本结构相似,具有数据整合的可行性。参考国际标准,经专家讨论,项目组构建呼吸专病队列的数据标准概念框架,由呼吸专病队列通用数据标准及疾病特异数据标准两部分构成,其中通用数据标准针对各专病队列中均有涉及、能够统一标准的问题或研究变量;特异数据标准则为各疾病特有的问题。经映射匹配,认为该标准与各现有专病队列的变量模块匹配良好,标准可行。**结论** 数据标准建立后,在回顾性整合现有队列资源的同时,使不同项目以相同的定义和标准开展长期随访,收集核心数据集,为未来开展多中心研究扫除因数据标准不一导致的数据共享障碍,更有利于多源的整合与共享。

【关键词】 呼吸系统疾病队列研究;通用数据标准;数据整合;数据共享

基金项目:国家重点研发计划项目(2016YFC0901100)

Data harmonization and sharing in study cohorts of respiratory diseases Sun Yixin, Pei Zhengcun, Zhan Siyan

Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China

Corresponding author: Zhan Siyan, Email: siyan-zhan@bjmu.edu.cn

【Abstract】 Objective Chronic obstructive pulmonary disease, asthma, interstitial lung disease and pulmonary thromboembolism are the most common and severe respiratory diseases, which seriously jeopardizing the health of the Chinese citizens. Large-scale prospective cohort studies are needed to explore the relationships between potential risk factors and respiratory disease outcomes and to observe disease prognoses through long-term follow-ups. We aimed to develop a common data model (CDM) for cohort studies on respiratory diseases, in order to harmonize and facilitate the exchange, pooling, sharing, and storing of data from multiple sources to serve the purpose of reusing or uniforming those follow-up data appeared in the cohorts. **Methods** The process of developing this CDM of respiratory diseases would follow the steps as: ① Reviewing the international standards, including the Clinical Data Interchange Standards Consortium (CDISC), Clinical Data Acquisition Standards Harmonization (CDASH) and the Observational Medical Outcomes Partnership (OMOP) CDM; ② Summarizing four cohort studies of respiratory diseases recruited in this research and assessing the data availability; ③ Developing a CDM related to respiratory diseases. **Results** Data on recruited cohorts shared a few similar domains but with various schema. The cohorts also shared homogeneous data collection purposes for future follow-up studies, making the harmonization of current and future data feasible. The derived CDM would include two parts: ① thirteen common domains for all the four cohorts and derived variables from disparate questions with a common

schema, ② additional domains designed upon disease-specific research needs, as well as additional variables that were disease-specific but not initially included in the common domains. **Conclusion** Data harmonization appeared essential for sharing, comparing and pooled analyses, both retrospectively and prospectively. CDM was needed to convert heterogeneous data from multiple studies into one harmonized dataset. The use of a CDM in multicenter respiratory cohort studies would make the constant collection of uniformed data possible, so to guarantee the data exchange and sharing in the future.

【Key words】 Cohort study of respiratory diseases; Common data model; Data harmonization; Data sharing

Fund program: National Key Research and Development Program of China (2016YFC0901100)

慢性呼吸系统疾病是严重危害我国居民健康的疾病之一,已经与心脑血管病、癌症并列成为我国居民的主要死因^[1],导致巨大的经济负担与疾病负担。其中,COPD、哮喘、间质性肺疾病和肺血栓栓塞症是常见的重大呼吸疾病,其发病机制、病因、个体化防治策略等尚不明确,亟需建立呼吸专病队列开展研究。

队列研究是目前国际上公认的研究暴露因素与疾病结局之间关联的流行病研究方法,是探索现代医学问题的重要研究手段。目前,研究者倾向于开展大型人群队列研究,大规模人群调查虽然有利于提升结果的可靠性,但却需要耗费大量人力、物力及时间成本,在探索特殊问题上也存在局限性,因此,对现有多个小样本数据整合为大样本研究的需求日益突出^[2]。充分利用现有研究资源,在对多源异构数据进行整合和规范的基础上,进一步开展科学研究,以获得准确可靠、外推性强的研究结果就显得很有必要。

在对多个项目资源之间进行数据整合与共享方面,国际上已经进行了多次尝试,但国内尚缺乏成功经验与模式。数据整合与共享方式按照时序性可分为回顾性、前瞻性与双向整合 3 种。①回顾性整合是在项目数据收集已经完成的情况下,开展的数据整合与共享工作,例如美国 Fred Hutchinson 癌症研究中心的 COMPASS 部门,针对多个癌症数据库开展的数据整合工作^[3],由于不同研究对于变量的定义、分类等方面不一致,因此这种数据整合方式最关键的步骤就是评估多源异构数据的相容性与一致性,从而识别、确定标准的通用数据元素,再将各项目数据分别进行映射与转化,形成最小数据集^[4]。②前瞻性整合是指在项目开始之初就有数据共享的意愿,因此针对某个特定领域定义最小数据集,使不同项目以相同的定义和标准来收集核心数据^[4],国际上开展的前瞻性多中心合作项目大多采用的是这种方式。③双向整合结合了前两者的优点,适用于数据收集已经开始但尚未完成的情况,基于现有资源建立标准化数据模型,一方面对已收集的数据进

行映射与转化,充分利用已有数据资源;另一方面在未来可采用统一标准收集数据,为数据整合扫除障碍,节约研究成本,更有利于多源数据的共享。

本文主要介绍的“呼吸系统疾病专病队列研究”已获国家科技部立项支持(2016YFC0901100),项目基于我国 5 个较为完备的大型社区队列和 4 个已有良好基础的呼吸疾病临床队列,将系统整合现有队列资源,建立呼吸系统疾病专病队列(呼吸专病队列),项目分课题“呼吸疾病专病队列标准化与质控的方法学研究”(2016YFC0901105)将在回顾性整合数据的同时,针对研究队列的多源异构现状,前瞻性确定最小数据集,建立呼吸系统疾病通用数据模型与各专病队列特异数据模型,以统一的标准收集数据,并在此基础上开展长期随访,实现各队列资源的数据统一与信息共享。呼吸专病队列项目有助于充分挖掘我国现有数据的科研价值,为今后多源异构的大规模队列数据整合与利用提供模式和方法上的重要参考,现将其数据标准制定的思路与阶段性成果介绍如下。

资料与方法

针对呼吸系统疾病领域建立通用数据标准的步骤:首先,学习与参考国际标准,其中临床数据交换标准协会(Clinical Data Interchange Standards Consortium, CDISC)的 clinical data acquisition standards harmonization(CDASH)模型,与观察性医疗结果合作组织(Observational Medical Outcomes Partnership, OMOP)的通用数据模型(common data model, CDM),对于呼吸专病队列研究的标准制定最具有借鉴意义;其次,整理、归类现有呼吸专病队列资源,评估各队列资源与国际标准的匹配与差异情况;最终建立呼吸专病队列数据标准。标准制定的思路与技术路线见图 1。

1. 学习、参考国际标准:在数据标准化和整合方面,国际上已有组织提出了针对不同领域的标准数据集,以及通用数据模型的概念,目前最受关注的是 CDISC^[5]针对临床试验提出的 CDASH 模型和

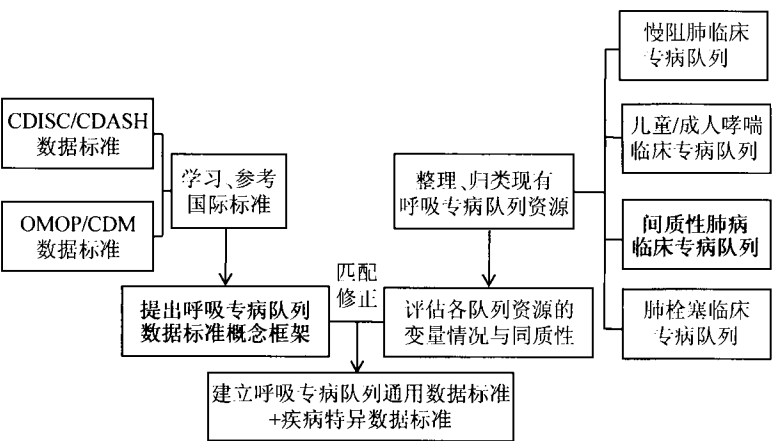


图 1 呼吸专病队列数据标准制定思路与技术路线图

OMOP^[6-7]针对观察性数据定义的 CDM。二者在其应用领域均制定了适用的数据标准,提出了标准化调查模块与调查条目,并对条目进行明确的定义与说明,对本项目数据标准制定有极高的参考价值,现进行简要介绍。

(1) CDISC/CDASH 标准:CDISC 是一个开放的、多学科参加的非盈利性组织,其针对临床试验的

各个环节,建立了一套涵盖试验方案设计、操作、数据收集、分析、提交等方面的完整 ODM(operational data model)数据标准^[8]。CDASH 模型作为该标准中的一部分,主要用于临床试验的数据收集,有助于建立标准的病例报告表(case report form, CRF),从而支持临床研究从数据收集到数据提交的标准化^[9]。

CDASH 模型^[10]包括 10 个主要模块(Domain),涵盖受试者基线特征、干预、事件等多个方面,定义及设定模块的目的是将 CRF 中的问题或研究变量归类,

统一制定适用的数据标准,具体介绍见表 1。

(2) OMOP/CDM 标准:OMOP 组织一直致力于合理利用多来源的观察性电子医疗数据库,开展医疗产品的安全性与有效性研究^[11]。由于多源数据往往结构各异,若电子医疗数据的结构化程度高则更有利于数据的提取与使用,因此,OMOP 组织创建了一套统一的框架,即 CDM,来标准化观察性数据的

表 1 CDASH 模型的主要模块及说明

模块英文名称	缩写	模块名称	模块内容
Adverse Events	AE	不良事件	对研究期间不良事件的信息收集,如是否发生不良事件、名称、开始时间、结束时间、严重程度等
Comments	CO	注释	对所收集数据的备注与详细说明,为自由编辑的文本
Prior and Concomitant Medications	CM	既往与伴随用药史	对药物/临床治疗情况的信息收集(包括既往用药、伴随用药与研究所关注药物的使用情况),如药物名称、剂量、使用时间等
Demographics	DM	人口统计学	对受试者基本人口统计学特征的信息收集,如年龄、性别、民族与种族等
Disposition	DS	试验完成状况	对研究中采取的措施与方案中重要事件的收集,如盲法、受试者的随访状态、试验完成或终止的时间等
Drug Accountability	DA	用药计数	对试验所用疗法/药物的分配与收回的信息收集,如疗法/药物的名称、分配或收回的时间、数量等
ECG Test Results	EG	心电图检测结果	对研究期间心电图检测结果的收集(包括中心读图结果、当地读图结果、对临床意义的总体评价与解释),如是否进行心电图检查、测量方法、检查日期等
Exposure	EX	试验干预	对受试者暴露于试验干预措施的信息收集,如干预开始时间、结束时间、剂量等
Inclusion and Exclusion Criteria	IE	纳入/排除标准	对研究所纳入的受试者资格的信息收集,如受试者是否符合纳入/排除标准等
Laboratory Test Results	LB	实验室检查结果	对研究期间实验室检查结果的收集(包括中心结果、当地结果、对临床意义的总体评价与解释),如样本收集时间、检查项目、结果、是否正常等
Medical History	MH	病史	对受试者病史的信息收集(包括既往与伴随疾病、临床事件),如受试者是否有病史、疾病处于进展期还是控制期,发生时间、结束时间等
Physical Examination	PE	体格检查	对研究期间体格检查结果的收集,如检查时间、是否正常等
Protocol Deviations	DV	方案偏离	对研究方案偏离情况的评价,如有无方案偏离、偏离内容、开始时间、结束时间等
Subject Characteristics	SC	受试者特征	具体研究所关注的受试者其他特征
Substance Use	SU	物质滥用	对物质滥用的信息收集,如有无物质滥用(烟、酒、咖啡因等)、数量、频率、持续时间等
Vital Signs	VS	生命体征	对受试者生命体征(身高、体重、血压等)的信息收集,如测量时间、方法等

格式,支持多源异构的观察性医疗数据通过提取、转化与加载(extract transform load,ETL)来形成标准化数据仓库,从而进行数据的分析与利用^[12]。多中心合作研究可以通过使用所定义的CDM,从不同的数据中心获取研究所需的数据,实现数据共享^[13]。

CDM目前已经发展到第五版^[14],包括标准化术语表、标准化临床数据表、标准化健康系统数据表、标准化卫生经济数据表等6大模块,见图2。

其中,标准化临床数据模块(standardized clinical data table)对临床数据进行了归纳整理,并按数据的特点将具体的问题就研究变量归类至不同的表格中,即为不同的数据模块,主要内容见表2。

2. 整理、归类现有呼吸专病队列资源:本研究依托于4个已有良好基础的呼吸疾病临床队列:COPD、

专病队列、儿童/成人哮喘专病队列、间质性肺疾病专病队列和肺栓塞专病队列,将各队列现有的数据资源及CRF进行初步整理、归纳,评估现有队列资源的变量模块等基本情况,以及与国际标准的匹配与差异情况,从而评估各队列资源间的同质性与整合的可能性。

3. 建立呼吸专病队列数据标准:结合前述国际数据标准的模块与适用条件,以及各专病队列的数据特征及匹配情况,针对呼吸疾病领域定义最小数据集,建立呼吸专病队列数据标准,包括呼吸队列通用数据标准和疾病特异数据标准两部分。

结 果

1. 现有呼吸专病队列资源基本情况:对于COPD、

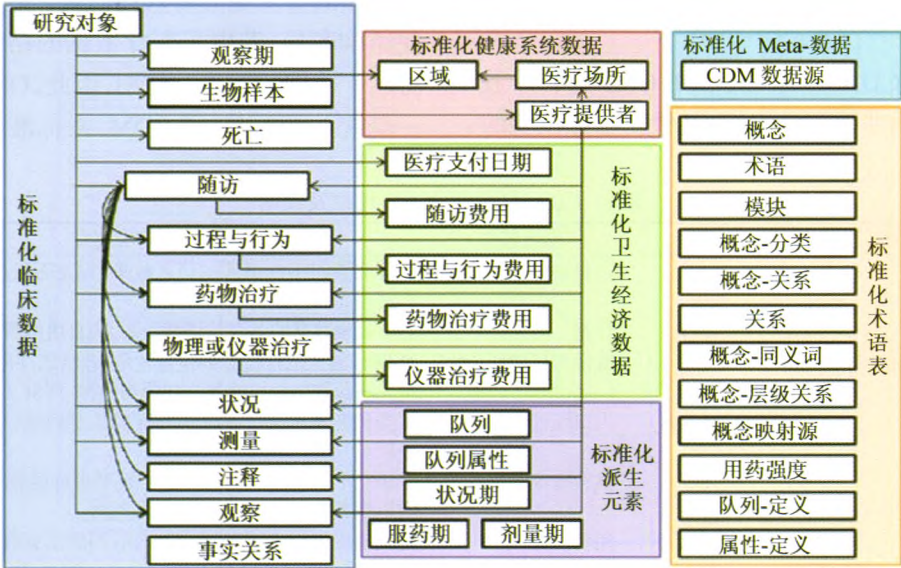


图2 观察性医疗结果合作组织通用数据模型(CDM)5.0版本^[14]

表2 通用数据模型标准化临床数据表的主要模块及说明

模块英文名称	模块名称	模块内容
PERSON	研究对象	研究对象的人口统计学特征,如性别、出生日期、民族与种族等
OBSERVATION_PERIOD	观察期	观察期的信息与记录,如观察开始时间、结束时间等
SPECIMEN	生物样本	生物样本的信息,如生物样本种类、采集时间、数量等
DEATH	死亡	研究对象的死亡情况,如死亡事件、原因等
VISIT_OCCURRENCE	随访	研究对象的随访信息,如随访方式、开始时间、结束时间等
PROCEDURE_OCCURRENCE	过程与行为	研究对象以诊断或治疗为目的的过程与行为信息,如行为种类、时间等
DRUG_EXPOSURE	药物治疗	研究对象的药物治疗情况,如药物名称、开始时间、结束时间、剂量等
DEVICE_EXPOSURE	物理或仪器治疗	研究对象的理疗或仪器治疗情况,如仪器种类、开始时间、结束时间、数量等
CONDITION_OCCURRENCE	状况	研究对象基于诊断与症状所反映的身体状况信息,如状态名称、时间等
MEASUREMENT	测量	研究对象或样本通过一系列检查所获得的信息,如检查时间、测量值、测量值正常范围等
NOTE	注释	对研究对象非结构化信息的备注与说明,为自由编辑的文本,如备注时间、内容等
OBSERVATION	观察	通过检查、问卷与活动获取的研究对象其他临床信息,包括其他模块未纳入的变量(病史、家族史等)
FACT_RELATIONSHIP	事实关系	各模块间的相互关系

哮喘、间质性肺疾病和肺栓塞临床专病队列,整理、归纳各队列现有的数据资源及 CRF,统计其变量模块的基本情况见表 3。经评估可以发现,各队列资源的变量模块同质性较好,基本结构相似,具有数据整合的可行性。

2. 呼吸专病队列数据标准:参考国际标准与各专病队列现有变量基本情况,经呼吸疾病临床专家、流行病学专家讨论,本研究提出呼吸专病队列数据标准概念框架,由呼吸队列通用数据标准及疾病特异数据标准构成。其中,通用数据标准针对各专病队列中均有涉及、且能够统一标准的问题或研究变量,如基本信息、社会人口学信息、危险因素等;疾病特异数据标准是针对各专病特有的问题或研究变量所设定的数据标准,如间质性肺疾病涉及的多个特异性检查、肺栓塞涉及的抗凝治疗情况调查等。

(1)呼吸队列通用数据标准:通过分析各专病队列的数据特征,同时匹配国际标准已有模块,研究组已初步拟定包含“基本信息”、“社会人口学信息”、“暴露因素/危险因素”等 13 个模块的变量表单,以此构成呼吸队列通用数据标准,各专病队列的 CRF 设计及优化应参照本标准,以实现数据标准统一及数据共享。通用数据标准具体内容见表 4。

(2)疾病特异数据标准:各专病队列特有的、无法列为通用数据标准的问题及研究变量,亦按照通用数据标准制定的流程,将特有数据匹配至国际标准模块或重新定义成适用于本项目的模块,最终汇总形成疾病特异数据标准。

(3)各专病队列变量模块与通用数据标准的匹配情况:将各专病队列的变量模块与构建的通用数据标准进行匹配,见表 5。结果显示匹配程度良好,说明研究构建的通用数据标准适用于以上 4 个呼吸专病队列,研究组将继续细化通用数据标准的变量,并使各专病队列参照本标准进行 CRF 设计与优化,在未来以相同的标准与定义收集数据,开展随访。

表 3 慢阻肺、哮喘、间质性肺病、肺栓塞专病队列变量基本情况

变 量	慢阻肺 专病队列	儿童哮喘 专病队列	成人哮喘 专病队列	间质性肺病 专病队列	肺栓塞 专病队列
基本信息、社会人口学	25	12	12	47	14
危险因素	80	41	74	45	16
症状、病史与用药史	188	40	40	198	78
检查信息	184	111	112	607	295
诊断与治疗	180	10	11	99	74
随访情况	96	120	117	30	226
附录-量表、问卷	146	35	29	81	-
特异性模块	34	-	-	101	431
变量总数	933	369	395	1 208	1 134

讨 论

目前,国际上开展了多项多中心参与的队列资源整合与共享项目,例如 BioSHaRE 项目(Biobank Standardization and Harmonization for Research Excellence in the European Union),该项目基于 8 项欧洲人群研究,旨在为欧洲现有队列及生物样本库建立数据整合工具与标准化系统^[15];再例如德国开展的 LCPD 项目(Lung Cancer Phenotype Database),确定肺癌通用数据集,进行数据整合,从而建立关于肺癌的中心数据库^[16]等。多源数据的整合能够获得大样本量,提高研究结果的外推性,多中心合作的开展与现有数据资源的高效利用,具有单项研究不可比拟的优势^[15]。但是,大量数据的整合与管理过程也存在很多问题与挑战,其中最关键的在于多源数据由于设计、收集等方面的不同导致的异质性问题^[17-18]。因此,在项目开展之前就建立标准化数据模型,在变量定义、数据收集、平台搭建等多个方面均考虑后期数据共享的需要,这种前瞻性的数据整合方式更有意义。

本研究建立呼吸专病队列数据标准,所参考的现有数据标准为 CDISC/CDASH 模型与 OMOP/CDM 模型。CDASH 模型是针对临床试验所建立的标准,定义了数据收集阶段应包括的基本模块^[10],参考该标准有助于本项目建立标准的 CRF。但是本项目为观察性队列研究而非临床试验研究,与该标准存在一定程度的偏差与不适用性,因此不能直接运用该标准。OMOP 的 CDM 模型虽然是针对观察性数据库所建立的标准,但其关注的更多是药品、疫苗等医疗产品上市后的安全性与有效性研究^[11],与本项目临床队列研究也不完全匹配。

项目所建立的呼吸专病队列数据标准特点:首先,针对 COPD、哮喘、间质性肺疾病和肺栓塞的疾病特征与临床诊治规律,确定了呼吸系统疾病研究所需的特异性变量模块。为挖掘呼吸疾病相关信息,探究暴露与疾病的发生、发展关系,以及开展精准化诊治研究奠定坚实的数据基础。第二,针对队列研究的特点,确定了无人干预及随机分组的观察性研究所需的变量模块。有利于观察疾病的自然进程与演变规律,为后期观察性研究开展分析奠定基础。第三,针对本项目双向性研究的特征,确定了与现有队列资源和未来研究所需数据均匹配良好的变量模块。既能充分

表 4 呼吸专病队列通用数据标准基本情况和与 CHASH、CDM 标准的匹配情况

模块名称	模块内容	可匹配的 CDASH 模块	可匹配的 CDM 模块
入选排除标准	研究对象入选资格的信息,如受试者是否符合纳入/排除标准,是否签署知情同意书等	IE	-
基本信息	研究对象的基本信息,如纳入研究的时间、患者 ID/病历号、通讯电话/地址等	DM	PERSON OBSERVATION_PERIOD
社会人口学信息	研究对象的基本人口统计学特征,如年龄、性别、出生日期、民族等	DM	PERSON
暴露因素/危险因素	研究对象对疾病可能危险因素的暴露情况,分为 4 个具体模块: ①职业与环境暴露 ②生活行为方式(吸烟史、饮食与体力活动等) ③过敏史 ④家族史	SU SC	OBSERVATION
既往史/现病史	研究对象的既往/伴随症状、疾病与用药信息,分为 4 个具体模块: ①疾病史与基础疾病(合并症) ②症状 ③用药史 ④非药物治疗史	CM MH	OBSERVATION
诊断信息	研究对象的诊断信息(包括门诊/入院/出院诊断)	-	PROCEDURE_OCCURRENCE OBSERVATION
检查信息	在研究期间,研究对象的检查信息,分为 4 个具体模块: ①体格检查 ②实验室检查,如血生化、血常规、痰细胞学检查等 ③影像检查,如胸片、心电图检查等 ④辅助检查,如肺功能检查等	EG LB PE	MEASUREMENT
治疗情况	在研究期间,研究对象的治疗信息,分为 3 个具体模块: ①治疗概括 ②药物治疗 ③非药物治疗	EX	DRUG_EXPOSURE DEVICE_EXPOSURE
疾病进展与转归	研究对象基于诊断与症状所反映的疾病进展与转归情况	AE	CONDITION_OCCURRENCE DEATH
卫生经济学指标	在研究期间,研究对象由于诊断、治疗的费用信息(包括检查费用、治疗费用与药品费用)	-	-
随访情况	研究对象随访期的相关信息	DS	VISIT_OCCURRENCE
生物标本	生物标本的采集与存储情况	-	SPECIMEN
量表与标准化问卷	反映研究对象呼吸系统状态、心理、睡眠等信息的量表与标准化问卷	-	-

表 5 各专病队列变量模块与通用数据标准匹配情况

模块名称	慢阻肺 专病队列	儿童哮喘 专病队列	成人哮喘 专病队列	间质性肺病 专病队列	肺栓塞 专病队列
入选排除标准	×	×	×	√	×
基本信息	√	√	√	√	√
社会人口学信息	√	√	√	√	√
暴露因素/危险因素	√	√	√	√	√
既往史/现病史	√	√	√	√	√
诊断信息	×	√	√	√	√
检查信息	√	√	√	√	√
治疗情况	√	√	√	×	√
疾病进展与转归	×	×	×	√	√
卫生经济学指标	×	×	×	√	×
随访情况	√	√	√	√	√
生物标本	√	×	×	×	×
量表与标准化问卷	√	√	√	√	√

统疾病队列研究建立的数据标准,因此,本研究所建立的通用数据标准具有一定意义。

本项呼吸疾病专病队列研究,在充分利用现有队列资源的同时,针对呼吸系统疾病领域建立通用数据标准,下一步,研究组将结合社区队列的变量情况,进一步细化、优化通用数据标准,使不同专病队列未来可以采用相同的定义和标准来收集核心数据,为开展数据共享,建立中国呼吸系统疾病专病队列中央数据库奠定基础;同时也为多地区、多中心数据资源的整合和利用提供重要参考。

利益冲突 无

参 考 文 献

- [1] 顾景范.《中国居民营养与慢性病状况报告(2015)》解读[J]. 营养学报, 2016 (6): 525-529. DOI: 10.3969/j.issn.0512-7955.2016.06.005.
Gu JF. 2015 report on Chinese nutrition and chronic disease[J]. J

利用现有队列数据,进行映射分析,又能以统一标准随访收集数据,整合与共享不同队列之间的数据,发挥数据的巨大作用。国际上缺乏特异性针对呼吸系

- Nutri, 2016 (6) : 525-529. DOI: 10.3969/j.issn.0512-7955.2016.06.005.
- [2] Griffith LE, van den Heuvel E, Raina P, et al. Comparison of standardization methods for the harmonization of phenotype data: an application to cognitive measures [J]. Am J Epidemiol, 2016, 184(10): 770-778. DOI: 10.1093/aje/kww098.
- [3] Rolland B, Reid S, Stelling D, et al. Toward rigorous data harmonization in cancer epidemiology research: one approach [J]. Am J Epidemiol, 2015, 182(12): 1033-1038. DOI: 10.1093/aje/kwv133.
- [4] 周光迪, 吴美琴, 赵丽, 等. 中国和加拿大合作出生队列研究数据统一及共享方法 [J]. 中国医药生物技术, 2015, 10(6): 494-497. DOI: 10.3969/j.issn.1673-713X.2015.06.004. Zhou GD, Wu MQ, Zhao L, et al. Data harmonization and sharing of birth cohort research cooperated between China and Canada [J]. Chin Med Biotechnol, 2015, 10(6): 494-497. DOI: 10.3969/j.issn.1673-713X.2015.06.004.
- [5] Clinical Data Interchange Standards Consortium (CDISC) [EB/OL]. (2017-01-01) [2017-06-10]. <http://www.cdisc.org>.
- [6] Observational Medical Outcomes Partnership (OMOP) [EB/OL]. (2017-01-01) [2017-06-10]. <http://omop.org/>.
- [7] OMOP. About Us [EB/OL]. (2017-01-01) [2017-06-10]. <http://omop.org/node/22>.
- [8] Hume S, Aerts J, Sarnikar S, et al. Current applications and future directions for the CDISC Operational Data Model standard: A methodological review [J]. J Biomed Inform, 2016, 60: 352-362. DOI: 10.1016/j.jbi.2016.02.016.
- [9] 黎燕兰, 胡镜清, 刘保延. CDISC 标准与中医药临床研究数据标准化 [J]. 世界科学技术-中医药现代化, 2012, 14(6): 2109-2114. DOI: 10.3969/j.issn.1674-3849.2012.06.001. Li YL, Hu JQ, Liu BY. CDISC standards and clinical research data standardization of traditional Chinese medicine [J]. World Sci Technol Modernizat Tradit Chin Med, 2012, 14(6): 2109-2114. DOI: 10.3969/j.issn.1674-3849.2012.06.001.
- [10] Clinical Data Acquisition Standards Harmonization (CDASH version 1.1) [EB/OL]. (2017-01-01) [2017-06-10]. <https://www.cdisc.org/standards/foundational/cdash>.
- [11] Trifirò G, Coloma PM, Rijnbeek PR, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? [J]. J Intern Med, 2014, 275(6): 551-561. DOI: 10.1111/joim.12159.
- [12] Rijnbeek PR. Converting to a common data model: what is lost in translation? Commentary on "fidelity assessment of a clinical practice research datalink conversion to the OMOP common data model" [J]. Drug Saf, 2014, 37(11): 893-896. DOI: 10.1007/s40264-014-0221-4.
- [13] 王玲. 美国观察医疗结果合作项目中数据组织及通用数据模型的应用研究 [J]. 中国药物警戒, 2015, 12(6): 341-346. Wang L. Study on application of data organization and common data model of the observational medical outcomes partnership in US [J]. Chin J Pharm, 2015, 12(6): 341-346.
- [14] OMOP Common Data Model Specifications Version 5.0 [EB/OL]. (2017-01-01) [2017-06-10]. <http://omop.org/CDM>.
- [15] Doiron D, Burton P, Marcon Y, et al. Data harmonization and federated analysis of population-based studies: the BioSHaRE project [J]. Emerg Themes Epidemiol, 2013, 10(1): 12. DOI: 10.1186/1742-7622-10-12.
- [16] Firmkorn D, Ganzinger M, Muley T, et al. A generic data harmonization process for cross-linked research and network interaction. Construction and application for the lung cancer phenotype database of the German center for lung research [J]. Methods Inf Med, 2015, 54(5): 455-460. DOI: 10.3414/ME14-02-0030.
- [17] Fortier I, Doiron D, Little J, et al. Is rigorous retrospective harmonization possible? Application of the DataSHaPER approach across 53 large studies [J]. Int J Epidemiol, 2011, 40(5): 1314-1328. DOI: 10.1093/ije/dyr106.
- [18] Bennett SN, Caporaso N, Fitzpatrick AL, et al. Phenotype harmonization and cross-study collaboration in GWAS consortia: the GENEVA experience [J]. Genet Epidemiol, 2011, 35(3): 159-173. DOI: 10.1002/gepi.20564.

(收稿日期: 2017-07-10)

(本文编辑: 王岚)

读者·作者·编者

本刊常用缩略语

本刊对以下较为熟悉的一些常用医学词汇将允许直接用缩写, 即在文章中第一次出现时, 可以不标注中文和英文全称。

OR	比值比	HBcAg	乙型肝炎核心抗原
RR	相对危险度	HBsAg	乙型肝炎e抗原
CI	可信区间	抗-HBs	乙型肝炎表面抗体
P_n	第n百分位数	抗-HBc	乙型肝炎核心抗体
AIDS	艾滋病	抗-HBe	乙型肝炎e抗体
HIV	艾滋病病毒	ALT	丙氨酸氨基转移酶
MSM	男男性行为者	AST	天冬氨酸氨基转移酶
STD	性传播疾病	HPV	人乳头瘤病毒
DNA	脱氧核糖核酸	DBP	舒张压
RNA	核糖核酸	SBP	收缩压
PCR	聚合酶链式反应	BMI	体质指数
RT-PCR	反转录聚合酶链式反应	MS	代谢综合征
Ct值	每个反应管内荧光信号达到设定的阈值时所经历的循环数	FPG	空腹血糖
PAGE	聚丙烯酰胺凝胶电泳	HDL-C	高密度脂蛋白胆固醇
PFGE	脉冲场凝胶电泳	LDL-C	低密度脂蛋白胆固醇
ELISA	酶联免疫吸附试验	TC	总胆固醇
A值	吸光度值	TG	甘油三酯
GMT	几何平均滴度	COPD	慢性阻塞性肺疾病
HBV	乙型肝炎病毒	CDC	疾病预防控制中心
HCV	丙型肝炎病毒	WHO	世界卫生组织
HEV	戊型肝炎病毒		