



2023

CHINA

INTERCHANGE

BEIJING | 25-26 AUGUST



CDISC学术研究现状、局限性及趋势

Current Status, Limitations and Trends of CDISC Academic Research

Li Mengya, Statistician, Medical Statistics Department
National Center for Cardiovascular Disease/ Fuwai hospital
limengya@mrbc-nccd.com



Meet the Speaker

Mengya Li

Title: Statistician

Organization: Department of Medical Statistics in National Center for Cardiovascular Disease/ Fuwai hospital

Education experience : Phd at Peking Union Medical College;
Master at University College London.

Work experience: Statistician manager at Sinocelltech company;
Statistical programmer at IQVIA.



Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*



Agenda

1. 概念和背景
2. 文献研究现状
3. 研究局限性
4. 研究趋势
5. 小结



概念和背景

- CDISC标准简介
- CDISC 文献发表情况

CDISC

临床数据交换标准协会（Clinical Data Interchange Standards Consortium, CDISC）是一个全球、开放、多学科的非盈利性组织，建立了涵盖研究方案设计、数据采集、分析、交换、递交、储存等环节的一系列标准，其目的是实现临床试验数据标准化，方便数据的交流与汇总分析，缩短研究时间和审批周期，提高临床试验数据质量以及统计分析的质量和效率。

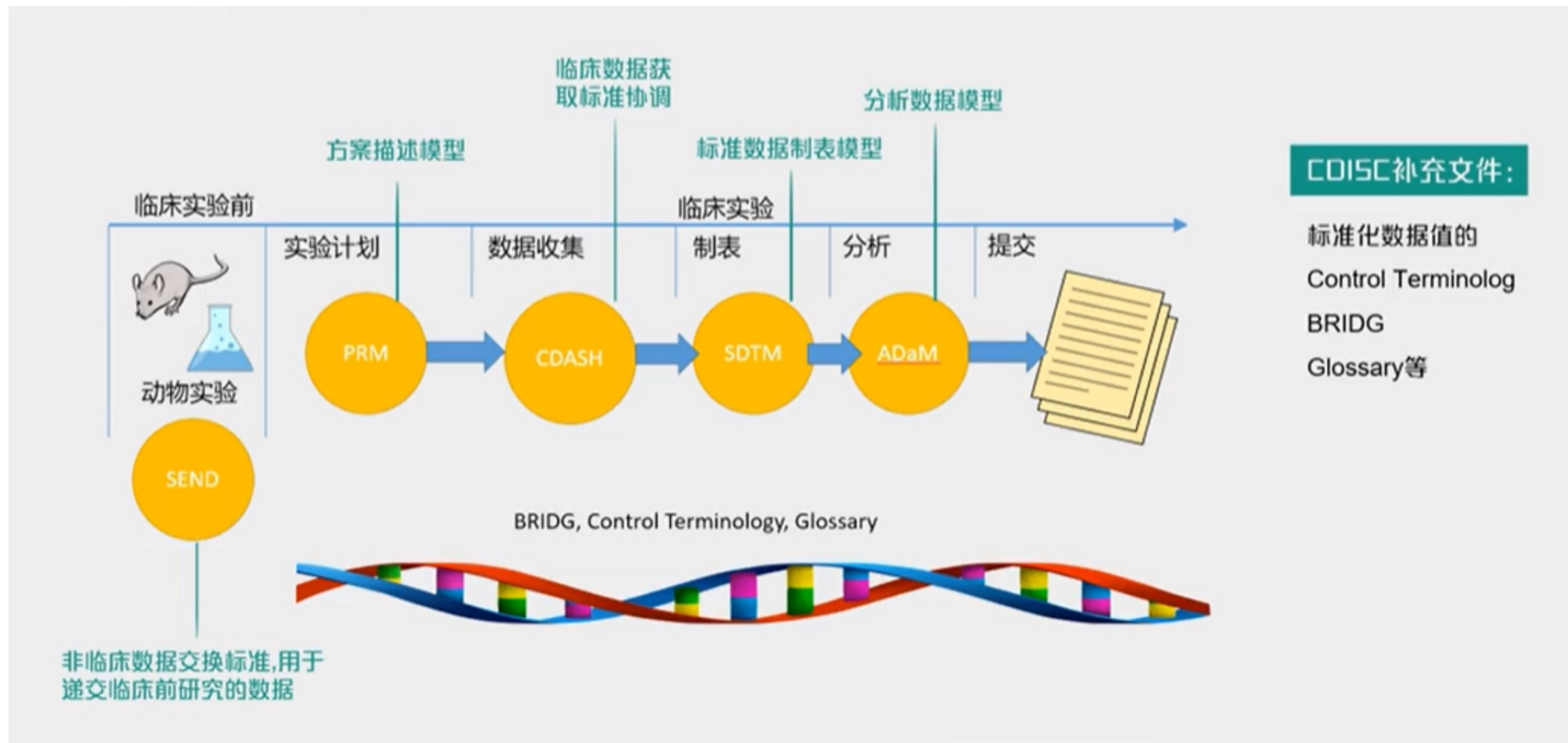
最早**成立于1997年**，经过20余年的发展壮大，CDISC已拥有超过**500家**跨国公司会员，超过**60个**国家的志愿者参与了标准的开发。CDISC标准已经成为目前在国际上应用最为广泛的临床试验标准。

美国和日本的监管机构要求企业采用CDISC标准递交数据。**中国目前则是鼓励**企业采用CDISC标准。

CDISC核心标准

标准	描述
非临床数据交换标准(SEND)	SDTM标准的拓展,用于递交临床前研究的数据.
方案表述模型(PRM)	包括研究方案的内容,如研究设计、资格标准和官方注册的要求,对研究方案提供的标准
试验设计模型(TDM)	定义了用于表述事件的计划顺序和试验处理计划的结构的内容标准.该部分是SDTM和方案表述(PRM)的子集.
临床数据获取协调标准(CDASH)	用于病例报告表中数据收集字段的内容标准,该标准基于SDTM
研究数据列表模型(SDTM)	数据组织和格式化的标准,用于向监管部门递交的内容标准.
分析数据模型(ADaM)	分析数据时所执行的标准,用于向监管部门递交的内容标准.
操作数据模型(ODM)	基于XML,用于获取,交换,递交和归档临床研究数据的内容和格式标准.
病例报告表格数据定义规范(define.xml)	基于ODM的元数据标准,用来描述SEND、SDTM和ADaM的元数据。以便向监管部门告知使用了哪些数据集,变量,受控术语,以及其他的特定的元数据.
Dataset-XML	用来描述SEND、SDTM和ADaM的数据集。
实验室数据模型(LAB)	用于在临床化验室和研究申办者/CRO间进行数据转移的内容和格式标准.
CDISC术语集(CT)	支持CDISC模型/标准所设计的标准词汇和编码集

CDISC 标准组成图示



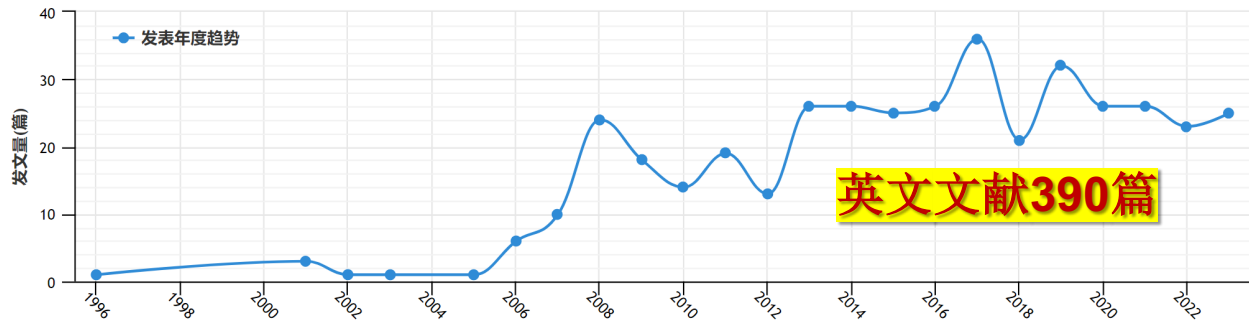
CDISC 文献发表情况

- 检索条件：主题='CDISC' OR 主题='Clinical Data Interchange Standards Consortium' OR 主题='临床数据交换标准' OR 篇关摘='CDISC' OR 篇关摘='Clinical Data Interchange Standards Consortium' OR 篇关摘='临床数据交换标准'
- 检索时间：2023.8

总体趋势分析

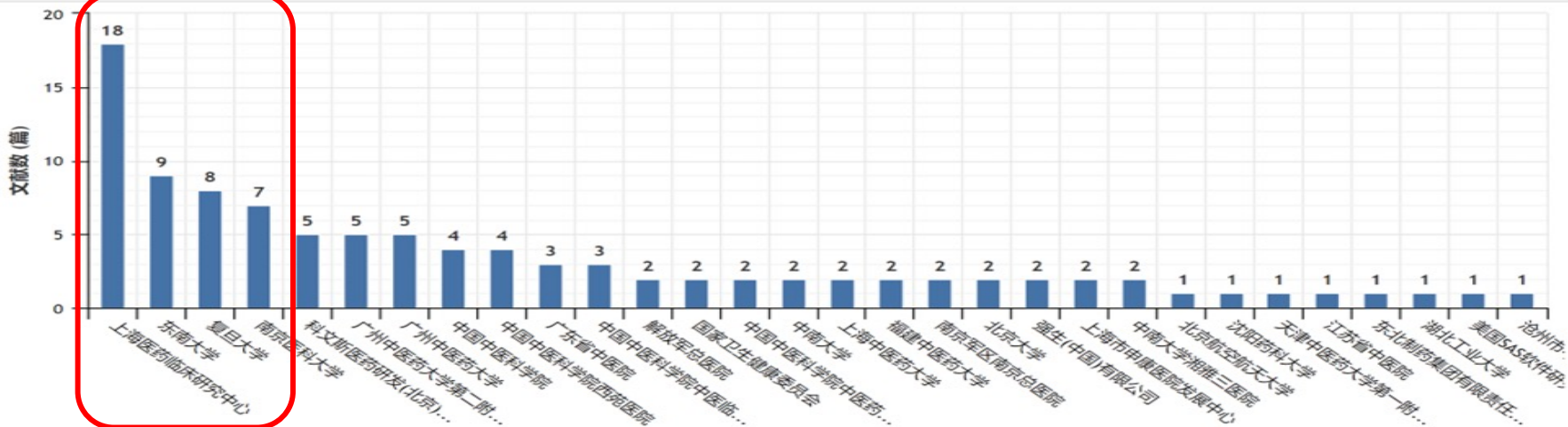


总体趋势分析

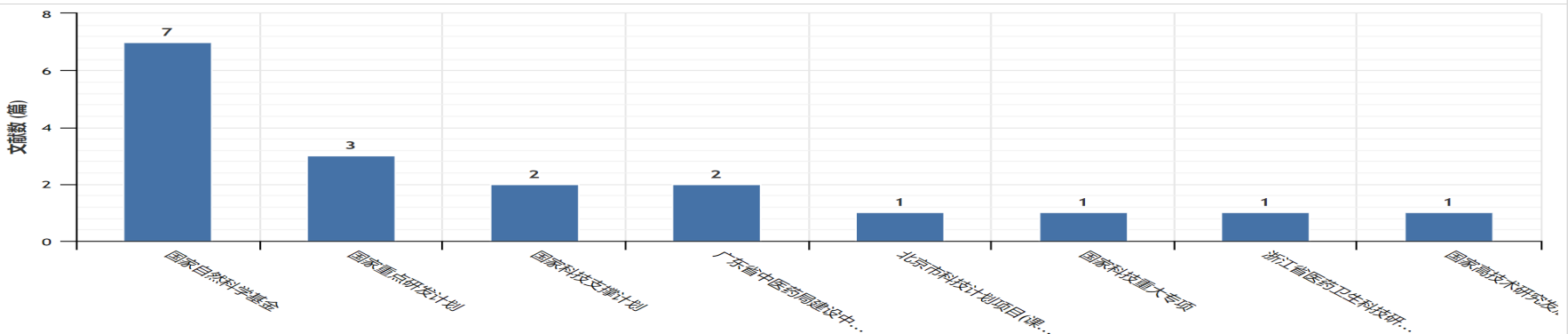


CDISC 中文文献发表—机构/基金分布

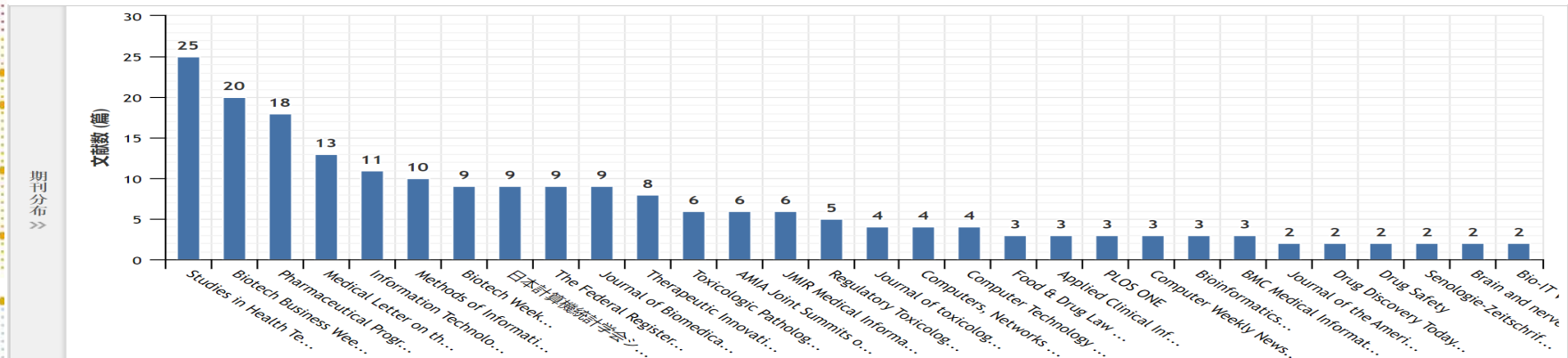
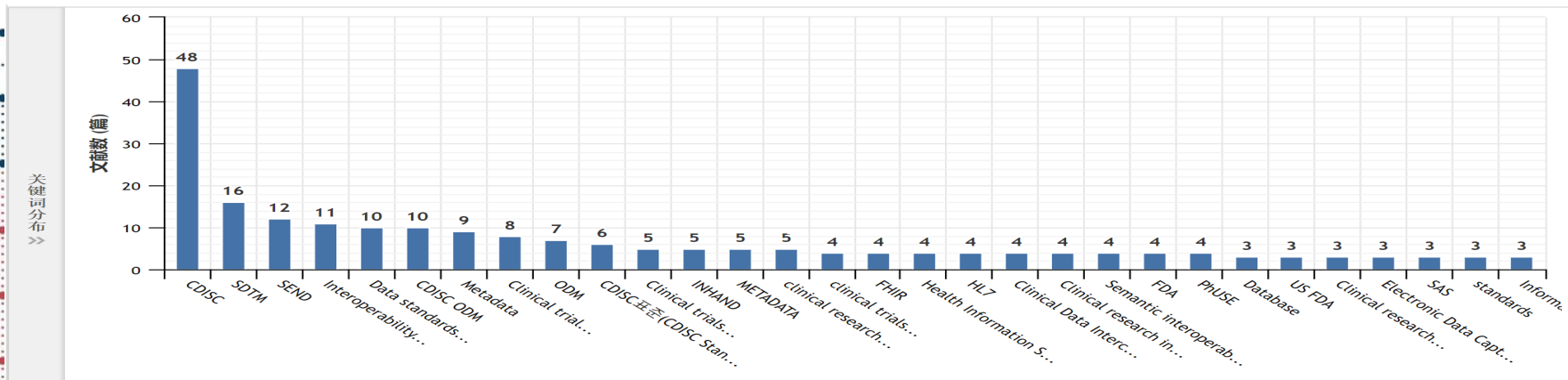
机构分布 >>



基金分布 >>



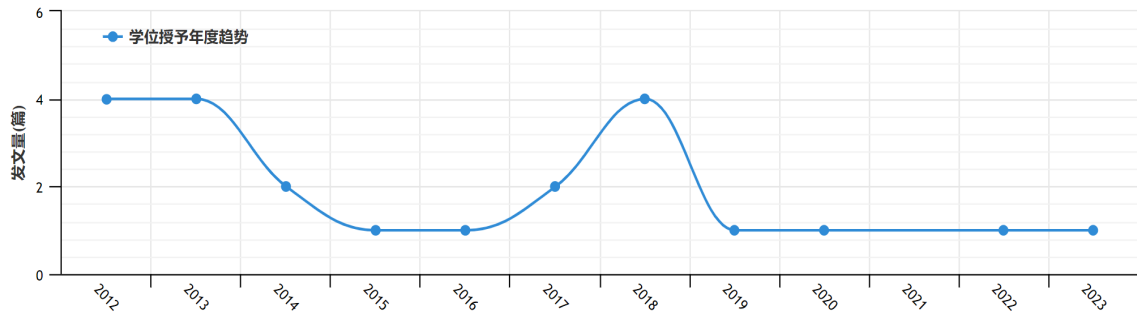
CDISC 英文文献发表—关键词/期刊分布



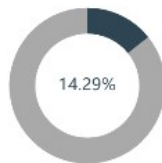
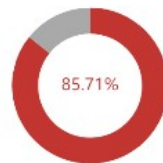
CDISC 21篇 博硕士学位论文

硕士论文占比较高，多来自复旦大学和东南大学，流行病学与卫生统计专业。

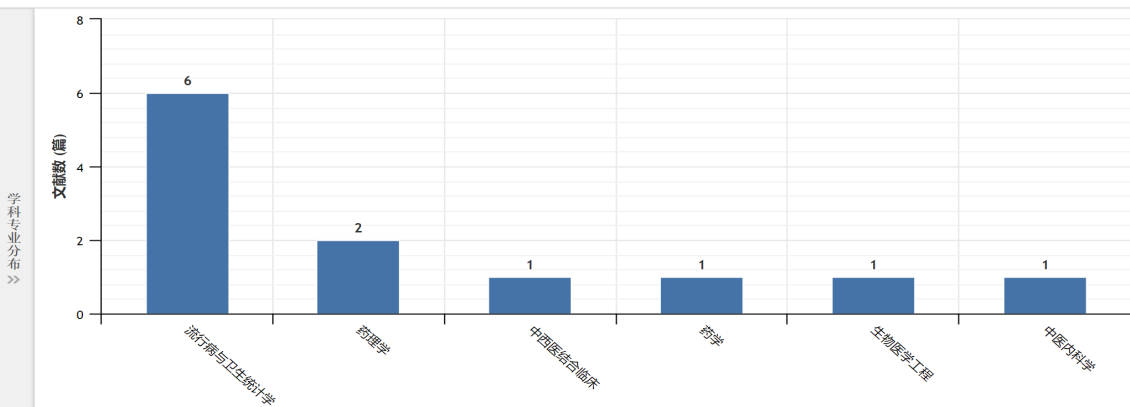
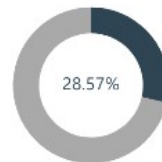
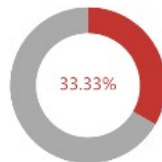
总体趋势分析



■ 硕士 ■ 其他 ■ 博士



■ 复旦大学 ■ 其他 ■ 东南大学





文献研究现状

- 中文文献
- 英文文献

CDISC 标准中文文章内容分布



CDISC 标准介绍



CDISC 数据库

- 多源临床研究数据整合关键技术与实现

数据来源：电子病历系统（EMRS）、临床检验系统（LIS）、医学影像管理系统（PACS）、放射信息管理系统（RIS）、输血管理系统（BTMIS）、护士工作站（NWS）等相关信息系统。

关键技术：CRF的标准化、电子病历到 CRF 的映射、CRF 到 SDTM 的映射

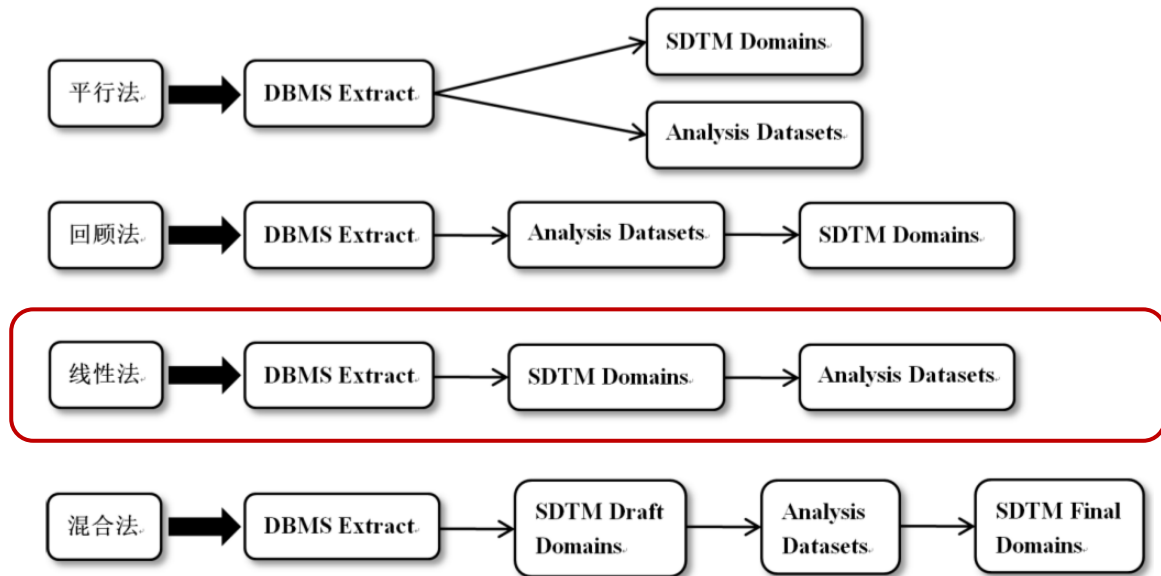
中国人民解放军总医院
围术期麻醉专科数据库

首都大学同仁医院
眼科研究所视功能数
据库

阜外医院
中国成人心血管外科
注册登记数据库

自动化实现

- 使用 EDC 系统收集和管理临床试验数据，实现临床试验数据管理电子化。并借助 SAS 宏和 Excel 变量映射文件生成通用的 CDISC 数据自动化系统，实现临床试验数据标准化，实现由 CDASH 数据生成 SDTM 和 ADaM 数据过程的自动化，提高生成 CDISC 数据效率，提高临床试验数据质量并缩短临床试验的时间。

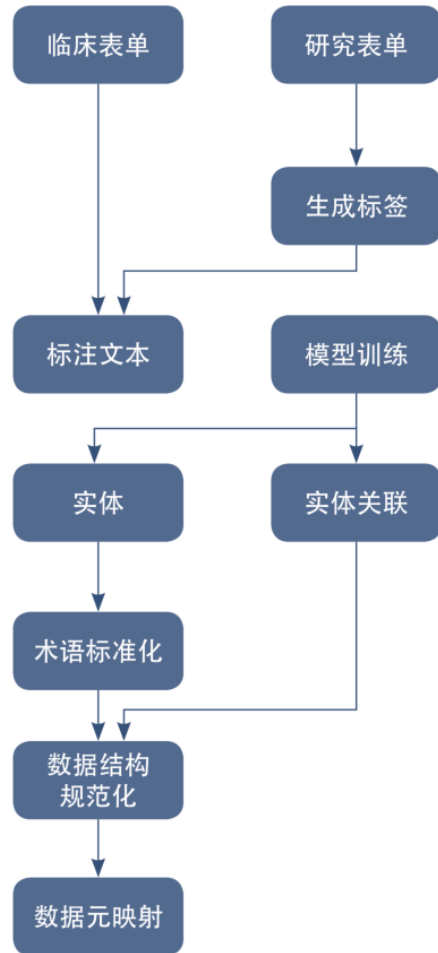


产生 SDTM数据集和ADaM数据集的四种方法

真实世界研究

从真实世界数据到临床研究数据标准转化

- 从医疗机构中的电子病历捕获源数据并将其传输到临床研究电子病例报告表，实现电子数据采集系统。
- 研发工具：根据临床研究方案中有关研究数据采集要求,在真实医疗环境中完成电子源数据记录。
- EDC 和电子病历分别发送CRF和患者临床表单至电子病例报告系统
- 研究数据集的建模及标签生成
- 模型训练和实体及实体间关系的提取
- 生成研究专用术语库
- 实体提取后，在填充CRF之前的规范化规则



CDISC标准在不同领域应用及推广

新冠

- COVID-19 治疗领域标准用户指南

肿瘤

- 临床试验中的应用；数据管理一般考虑

器械

- 项目中应用

中药

- 病例报告表设计；数据管理及标准化；

疫苗

- 数据管理软件开发；数据标准化

CDISC COVID-19--应对紧急事件，及时颁布相关指南

截至2021.4.7，ClinicalTrials 登记涉及 COVID-19 的临床研究 5 273 项；

截至2023.7.13，涉及 COVID-19 的临床研究 9 182项。

- 2020.4.21，CDISC发布了COVID-19 临时用户指南（COVID-19-IUG）、因 COVID-19 暴发而中断运行的研究指南。
- 基于WHO和国际严重急性呼吸道与新兴感染联合会联合开发的用来收集 COVID-19 相关临床特征的注释病例报告表（aCRF）和数据相关映射表格（MS）。
- 2021.7.8，CDISC更新并发布了COVID-19 TA标准（TAUGCOVID-19 v1.0），并增加了针对因 COVID-19 流行而中断运行的研究的临时分析数据模型（ADaM）指南和疫苗管理指南。

CDISC COVID-19

表 1 TAUG-COVID-19 主要内容

章节	是否有 SDTM 示例	SDTM 域	使用说明
危险因素	是		在发布 TAUG-COVID-19 时, 相关危险因素信息比较有限, 故只包括了目前尚属普遍金域大学发布的几个危险因素(肥胖、肥胖、年老和既往疾病, 其中既往疾病包括心脏病、肺病和糖尿病)。同时, 本节还讨论了个人防护装备、旅行史、接触史、嗜好品使用(如吸烟)和动物接触史等内容。
既往病史	是	既往病史域 (MH)	通过设置责任因素类别变量 MHCAT="RISK FACTOR FOR SEVERE ILLNESS FROM COVID-19" 来标记可能导致 COVID-19 的严重疾病类的危险因素。ER 域还处于草稿开发阶段, 预计它将在出现在接下来发布的 SDTMIG 的第 6.2 节。
个人防护装备 (PPE)	是	环境与社会因素域 (ER)	通过设置环境与社会因素类别变量 ERCAT="PPE" 来归类 PPE 信息。
旅行史	是	环境与社会因素域 (ER)	前往有 COVID-19 确诊病例的地区旅行可能会增加个人被感染的风险。为评估暴露于 COVID-19 的风险, 研究可能会收集特定时间内的旅行信息。通过设置环境与社会因素类别变量 ERCAT="COVID-19 RISK FACTOR" (COVID-19 危险因素) 和环境与社会因素类变量 ERSCAT="TRAVEL" 来归类旅行信息。
接触史	是	环境与社会因素域 (ER)	当受试者与确诊或可能患有 COVID-19 的病例有密切接触时, 可在相关人员 (AP) 域中收集该受试者密切接触的 COVID-19 或潜在 COVID-19 病例相关信息。
吸烟	是	嗜好品使用域 (SU)	举例说明了如何使用 SU 域收集吸烟相关信息, 包括仍在吸烟、已成烟和从不吸烟等状态。
动物暴露情况	是	环境与社会因素域 (ER)	在极少数情况下, 冠状病毒可能会通过受感染的动物传播给人类(人畜共患病)。可通过设置环境与社会因素类别变量 ERCAT="COVID-19 RISK FACTOR" (COVID-19 危险因素) 和环境与社会因素类变量 ERSCAT="ANIMAL EXPOSURE" 来归类动物暴露信息。
发病	是	既往病史域 (MH)	举例说明了在 MH 域中如何收集 COVID-19 症状出现时间和诊断时间。
症状与体征	是	既往病史域 (MH)、发现域 (FA)	本节中的示例包括变量 _DY (研究日期)、研究日期与“人口学特征”DM 域中的受试者参照开始日期 (RFSTDCI) 相关。在干预类研究中, 参照开始日期通常指治疗的开始日期。 如果有关症状和体征的发生是作为一个整体事件的话, 可将数据在事件类域中展示; 如果有有关症状和体征的发生只是对某事件的评估(比如某个时间点(快照)或某个时间段(片段)的评估), 则可以在发现 FA 域中展示。 在研究开始收集的有关症状发生的数据通常在 MH 域或 FA 域中展示, 比如 factexp 数据集。但如果症状数据在研究开始及研究期间均会重复采集, 则需在只在 factexp 数据集收集。无论哪种情况, 都可以通过设置分类变量 CAT 将两类症状和体征, 比如设为“COVID-19 SYMPTOMS” (COVID-19 症状)。
实验室检测	是	实验室检查域 (LB)	此域收集微生物学、病毒学或药代动力学数据, 这些数据均有其他域来存储。有关 COVID-19 的病毒识别和病毒载量相关数据标准信息, 可参阅“病毒识别”(https://www.cdisc.org/standards/foundational/cdash)。
诊断与病毒学	是		
病毒识别	是	微生物学域 (MB)	举例说明了如何在 MB 域中收集 COVID-19 病毒检测相关结果。
抗体检测	是	微生物学域 (MB)	举例说明了如何在 MB 域中收集 COVID-19 抗体检测相关结果。
病毒载量	是	微生物学域 (MB)	举例说明了如何在 MB 域中收集病毒载量检测相关信息。
体征和体征	是	VS 域	VS 域存储生命体征; LB 域存储检测结果。
实验室检查	是	实验室检查域 (LB)	
伴随用药	是	伴随和既往用药域 (CM)	举例说明了如何在 CM 域中收集伴随用药相关信息。
呼吸相关的发现	是		
影像	是	诊疗操作域 (PR)、呼吸系统检查域 (RE)	PR 域存储影像检查过程相关数据; RE 域存储影像检查结果相关数据。
肺功能测试	否	呼吸系统检查域 (RE)	肺功能测试在 RE 中展示, 具体可参见哮喘治疗领域标准用户指南 (TAUG-Asthma, https://www.cdisc.org/standards/therapeutic-areas/asthma) 和 COPD 治疗领域标准用户指南 (TAUG-COPD, https://www.cdisc.org/standards/therapeutic-areas/copd)。
心电图/QT 间期的测量	否		心电图和 QT 间期的测量可能对于本节数据没有用, 可参见 QT 治疗领域标准用户指南 (https://www.cdisc.org/standards/therapeutic-areas/qts-studies)。
住院	是	就诊事件域 (HO)	举例说明了因 COVID-19 导致住院相关数据如何在 HO 域中收集。
住院过程相关	是		
辅助通气和氧气治疗	是	诊疗操作域 (PR)、伴随和既往用药域 (CM)	PR 域可能会用变量“中央设备识别码 SPDEVID”, 以与设计标识符 DI (Device Identifiers) 域相联系。变量 PRDECOD (诊疗操作标准化名称) 的编码列表可参见 MedDRA (国际医学用语词典)。CM 域可用于记录氧气治疗的具体信息, 比如适应症、剂量、频率等。
疫苗接种	是	诊疗操作域 (PR)	举例说明了如果发生任何疫苗接种恶化的管理指南, 如何进行数据收集。 可参见疫苗治疗领域标准用户指南 (TAUG-Vax, https://www.cdisc.org/standards/therapeutic-areas/vaccine)。此外, 本节也对 TAUG-Vax 存在的一些问题提出了解决方案。
问卷、评分和量表	否		CDISC 将问卷、评分和量表 (QRS) 作为独立指南来维护 (https://www.cdisc.org/foundational/qrs)。本指南仅所有者的明确许可, CDISC 无法为受试者提供保护的工具。CDISC QRS 本指南和标准制订了如何将问卷和量表测量工具与 MedDRA (国际医学用语词典)、CM 域可用于记录氧气治疗的具体信息, 比如适应症、剂量、频率等。

SDTM: 研究数据集表结构; TAUG-COVID-19: COVID-19 治疗领域标准用户指南; MHCAT: 既往病史类属; ERCAT: 环境与社会因素类别; ERSCAT: 环境与社会因素类别变量; PRDECOD: 诊疗操作标准化名称; MedDRA: 国际医学用语词典; QRS: 问卷、评分和量表; -TESTCD: 发现类检测、检查或条目名称的缩写; -TEST: 发现类检测、检查或条目名称的缩写; -CAT: 类别。

表 2 COVID-19 中断指南相关领域使用说明

章节	是否有 SDTM 示例	SDTM 域	使用说明
方案偏离	是	方案偏离域 (DV)	COVID-19 可能导致分中心级别的方案违背, 但目前 SDTMIG DV 域只给出了受试者水平的方案偏离的示例说明。故本节给出了两种用示例展示分中心水平方案偏离的方法。当分中心发生影响受试者的方案偏离时, 申办者需确定并分别记录分中心中受到影响的每个受试者偏离情况, 并通过设置方案偏离类别变量 DVCAT="SITE DEVIATION" 来识别分中心级别的方案偏离。二、使用整合定义域 POOLDEF 来定义分中心内每个发生方案偏离的受试者集合, 然后通过方案偏离类别变量 DVCAT="SITE DEVIATION" 来识别分中心级别的方案违背。DV 域与 POOLDEF 域通过整合识别码 (POOLID) 链接。
处置	是	处置域 (DS)	不建议修改处置事件类别变量 DSCAT, 处置事件类变量 DSSCAT 的受控术语, 同时也不建议为处置事件或试验方案里非标准名称变量 DSDECOD 的受控术语(完成或未完成情况, NCOMPLT)添加新的术语。建议使用非标准变量 (NSV) 来表示处置是由于大流行而引发的。
错过的访视	是	方案偏离域 (DV)、自定义访问事件域 (VE)	可使用 DV 域来展示错过的访视相关数据; 尽管一般用 SV 展示有关访视数据, 但 SV 域仅包括已发生的访视。且在 SDTM 模型中无法将计划作为标准名称来补充记录符合到 SV 域。故本节中, CDISC 提供了一种临时解决方案来收集错过的访视信息, 即使用 VE、VE 域可记录新增或改访视计划或允许远程访视的方案修订, 允许任何方案修改之前或之后的访视信息在单个域中展示。
错过的评估	否		协调有关错过访视和否评估的数据
结果	否		如果完全遗漏了某次访视, 则无需单独记录该次访视应该进行但被遗漏的评估。但如果收集了计划访视中的某些数据, 则通常应该在相关域中记录预期但在访视中错过的评估。 在临床数据获取标准 CDASH 中, 使用-PERF 记录某次访视中的域中的所有检测均未采集(比如 LPERFE="N"); 在 SDTM 中测试的不执行状态用-STAT="NOT DONE" 表示。使用-ALL (例如, 用域中的 "LBALL" 检测) 来表示特定域中所有检测均未进行。-REASND 表示未完成、一个或一组检测的原因。在具有多个调查表的研究中, 可使用 QSCAT 调查表的缩写代码和 OSALL 来表示该调查表中的所有问题均未完成。 药物分发与回收记录 DA 域用来展示研究治疗药物的分配和回收, 可通过设置变量 DASTAT="NOT DONE" 来表示分配或回收研究治疗药物的失败。有关如何将-PERF 中的 "N" 值映射到 SDTM 模型中变量-STAT 的值 "NOT DONE", 以参阅 CDASH 实施指南 (https://www.cdisc.org/standards/foundational/cdash)。
发生预先定义的事件或干预	否	发生预先定义的事件或干预	参见未完成检测, 但未发生预先定义的事件或未发生事件/次表示所有事件或干预均未发生。
自发事件或干预记录	否	自发事件或干预记录	需注意的是, 目前没有方法能表明在特定访视中未收集自发事件/次 (例如, 不良事件、伴随药物等)。
未实施定期干预	否		预先安排好的干预措施可通过设置-PRESP 和-OCCLUR 即使在 CRE 上明确收取值为 "N" 来表示整个计划内的干预措施未执行。对于-OCCLUR="N" 的干预措施, 使用非标准变量 (NSV)-REASOC (发生原因) 记录未实施干预措施的原因。对于制定给药剂量的干预措施, 可用一些标准变量来表示治疗剂量调整或治疗中止的原因。可使用非标准变量 (NSV) 表示中断的原因。
药物收发记录的改变	是	药物收发记录域 (DA)	在流行期间, 受试者可能无法亲自到访研究分中心接收或退还研究产品(比如药物/输液等)。在某些情况下, 申办方可能采取更安全的送货上门方式(例如, 直接送货给受试者)。
不良事件数据收集的变更	是	不良事件域 (AE)	注意 COVID-19 相关不良事件的 MedDRA 编码。
暴露数据	是	暴露域 (EX)、处置域 (DS)	可使用以下域和位置: 暴露 (EX) 和处置 (DS) 域, 用永久终止研究治疗; EX 域, 用于剂量调整和中断; 变量 EXADJ 和 EXRSDSC, 分别用于收集暴露剂量调整和终止的原因; 非标准变量 (NSV) 代表中断的原因。
不良事件数据	是	暴露域 (EX)、处置域 (DS)	本指南提供了两种能识别一个中心 ID 迁移到另一个中心的方案。申办者可根据他的系统流程和及相关研究选择可行的解决方案。 ① 使用自定义事件 ST 域, 它遵循医疗器械类 SDTMIG (SDTMIG-MD, https://www.cdisc.org/standards/foundational/dtm-ig-md) 中的 DT 进行模块化变量-TERM 用于收集跟踪的运动类型(比如本节中的 STTERM 可以是“受试者从一个中心到另一个中心 B”)、变量-PARTY 用于表示受试者迁移的主体类型(比如本节为“SITE 分中心”)、并用变量-PRTYID 来表示受试者主体标识符(比如本节为“分中心 ID”)。 ② 使用 DS 域, 设置处置类别变量 DSCAT="OTHER EVENT" (其他事件)、事件类变量-PARTY 和-PRTYID, 最初是为 SDTMIG-MD 中的 DT 域开发的, 用于表示从一个中心到另一个中心 B”)、变量-PARTY 用于表示受试者迁移的主体类型(比如本节为“SITE 分中心”)、迁移的参与方(比如本节为“SITE 分中心”)、迁移参与方的标识符(本节为分中心 ID) 用-PRTYID 表示。 设置组合 ID 变量 TSGRPID="COVID-19"。
提供大流行相关的试验	是	试验概要域 (TS)	提供大流行相关的试验摘要



肿瘤领域

- 肿瘤药物临床试验本身特点和复杂性

- ✓ 复杂研究设计（适应性无缝设计、富集设计及主方案设计）对遵循CDASH设计的CRF造成影响
- ✓ 数据多样性对数据标准化提出挑战
- ✓ 滚动递交对数据管理和统计分析提出挑战

临床数据的标准化

提高临床数据的质量

优化数据交流、递交的流程；

- 缩短研究周期、节省费用

提高临床试验的效率

加快药物上市速度，延长患者的生存期

肿瘤域

肿瘤/病灶标识 (TU)

肿瘤/病灶测量 (TR)

疾病效应/状态评估 (RS)

医疗器械临床试验数据递交要求注册审查指导原则（2021年第91号）

发布时间：2021-11-25

- ✓ 鼓励注册申请人参照临床数据交换标准协会（Clinical Data Interchange Standards Consortium, CDISC）标准递交数据。
- ✓ 递交形式：原始数据库、分析数据库、说明性文件和程序代码分别放置于四个文件夹中。

SDTMIG for Medical Devices与医疗器械相关的域名总共有7个

研究参考域

设备标识符（DI）

干预类域

设备暴露（DX）

关系域

设备-对象关系（DR）

事件类域

设备事件（DE）

设备跟踪和处理（DT）

结果类域

设备使用情况（DU）

设备属性（DO）

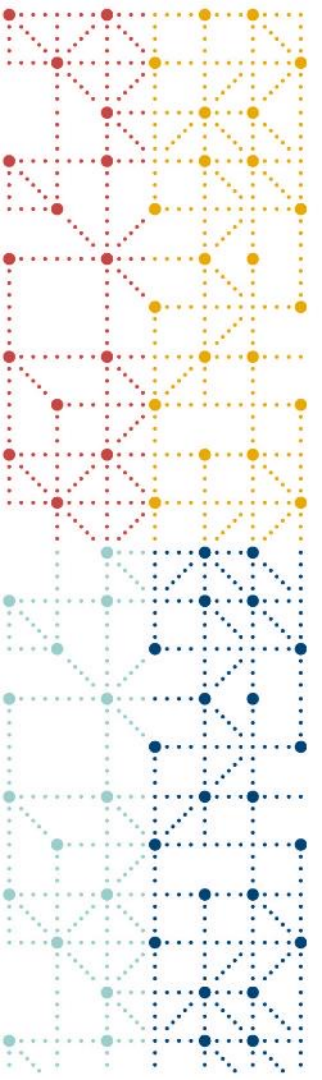
中医药领域

- 相关规范和指南中未对中药相关产品进行规定；
- 中医药研究团队推广CDISC标准在该领域的应用：
 - ✓ 中医特殊症状划分域
 - ✓ 创新具有中医药临床研究特点的新域
 - ✓ 中医临床试验术语体系构建
 - ✓ 符合CDASH的CRF设计

- 1) 中医诊断：证候信息，
- 2) 中医治疗：中药、针灸以及推拿；
- 3) 其他方面：气候因素、二十四节气、生活环境等。

Tab.2 Traditional Chinese medicine characteristic information collection supplemented in various fields

类别	领域	增加中医特色信息
干预类型	伴随用药	中药信息
事件类型	病史	中医证候诊断
发现类型	体格检查	望诊、闻诊、切诊

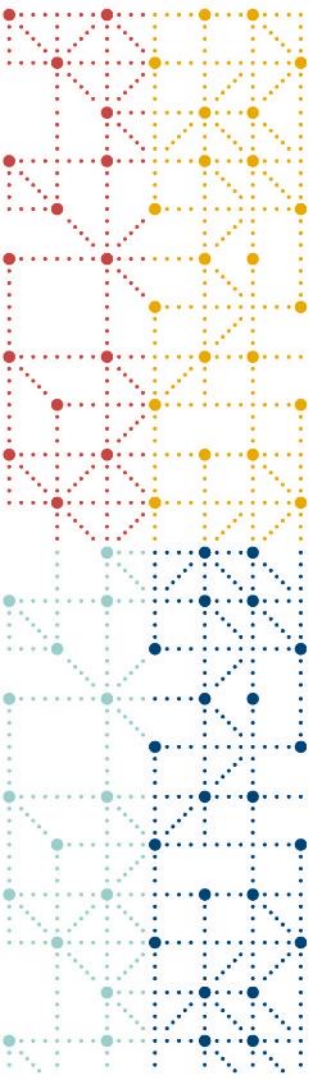


研究局限性

Slides Using Images

研究缺陷

- SDTM、ADaM、CDASH是目前应用最为广泛的三个CDISC标准，其他标准还有待研究；
- 目前相关研究多数是在校生或硕博论文，经验有限，提到的临床试验相关内容可能不全面，无法涵盖所有情况，不能满足所有试验的要求，需要进一步学习了解相关内容，结合实际工作，以便后续拓展；
- 自动化程度受限，软件功能进一步扩展，优化运算方法，以及对各种试验要求的适用性；
- 在CDISC标准本土化的过程中，中药、民族药由于其特殊性难以直接运用CDISC标准，尚未建立统一的数据标准，难以与国际通行的术语集相融合；
- 中医领域缺乏相关指导性文件；
- 翻译问题，优化翻译。

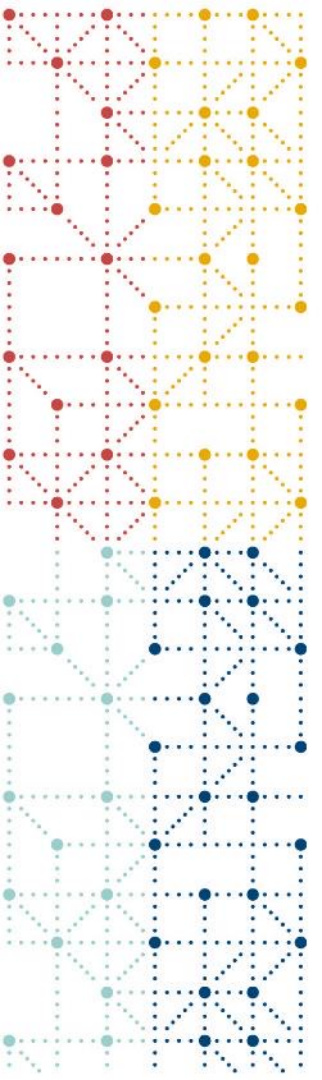


研究趋势

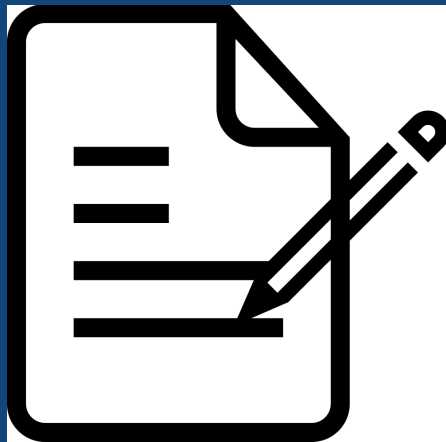
Slides Using Images

研究趋势

- 临床前SEND
- SAS宏程序的自动化程度、通用性进一步验证有待完善
- 有企业、监管部门的协力参与，要在业内广泛的进行讨论与培训，达成共识；
- 广泛讨论、达成共识、硬性要求
- 开发和支持能够实现信息互联互通的全球性、跨平台数据标准，使得不同临床研究间的数据可以方便地进行交换与共享，从而促进医药研究和相关卫生保健领域研究的发展。
- 尽快就中医药临床研究建立统一的数据交换标准以及控制术语规范，解决中医药临床研究数据交换与共享的问题



Summary 小结



Summary

- 早些年相关学术研究引领CDISC标准在行业内应用；
- 近年来学术研究相对滞后于工业界应用；
- CDISC标准结合突发事件和特殊领域进行完善；
- 工业界、学术界以及监管部门可以合作共同推进CDISC标准的发展和应用。



Thank You!



cdisc