

《临床科研数据集技术规范》团体标准编制说明

一、任务来源

为深入贯彻习近平总书记关于构建以数据为关键要素的数字经济的重要指示，落实《国家数据标准体系建设指南》等重要文件要求，充分发挥数据在激活医疗要素潜能、做优数字医疗服务等方面的引领和规范作用，制定临床科研数据集技术规范标准显得尤为重要。当前，大数据、人工智能等技术的发展为医疗数据的处理和分析提供了强大的技术支持。然而，这些技术的应用需要标准化的数据作为支撑。制定临床科研数据集技术规范标准，可以为大数据、人工智能等技术在医疗领域的应用提供数据基础。本文件由中国人民解放军总医院、北京嘉和美康信息技术有限公司、陆军特色医学中心、浙江大学第一附属医院、北京软件和信息服务业协会等 5 家单位共同提出制订。

二、制定规范的必要性和意义

临床科研数据集技术规范的目的通过制定和实施这一标准，期望能够解决临床科研中数据不一致、信息孤岛、隐私保护不足等问题，推动医疗信息化和智能化的发展，提高医疗服务质量和科研水平。

其意义主要体现在以下四个方面：

1.提高数据标准化：在临床科研领域，数据的标准化是提高数据质量、促进数据共享和利用的关键。通过制定统一的数据模型，可以确保不同医疗机构、研究机构之间数据的一致性和可比性，从而提高科研效率和质量。

2.促进信息共享：标准化的数据模型有助于打破信息孤岛，促进医疗信息在不同系统和平台之间的流通，为临床决策支持、疾病监测和公共卫生研究提供更为丰富和准确的数据资源。

3.提升数字化水平：统一的数据模型为医疗数字化的开发提供了指导，有助于简化系统设计，降低系统集成的复杂性和成本，同时提高系统的互操作性和可扩展性。

4.完善信息安全保护：在数据共享和利用的过程中，标准化的数据模型可以更好地实现患者隐私的保护，通过规范数据的收集、存储和传输，确保患者信息的安全。

三、主要编制过程

第一阶段，成立规范研制起草工作组（以下简称工作组）编制项目立项材料和标准草案。工作组组由 11 名相关专家和专业人员组成，来自中国人民解放军总医院、北京嘉和美康信息技术有限公司、陆军特色医学中心、浙江大学第一附属医院、北京软件和信息服务业协会等单位。工作组通过查阅文献，形成理论框架和工作方法，并对已有相关内容基础进行了梳理，多轮研讨沟通，形成立项申请材料 and 标准草案。工作组于召开立项准备沟通会，为项目立项制定完整方案。

第二阶段，项目立项评审。工作组于 2024 年 11 月 8 日在线召开了项目立项专家评审会。评审专家包括（姓氏拼音排序）：中国标准化研究院研究员康健，中国信通院主任工程师、高级工程师邱艳娟，北京航空航天大学软件学院教授级高工王宝会，清华大学公共管理学院决策科学实验室副主任、高级工程师王理达，首都医科大学附属北京朝阳医院高级工程师杨京菊。标准立项通过专家评审后，信息在全国团体标准信息平台和协会官方网站和北京软协官网发布。

第三阶段，编写规范征求意见稿。围绕标准草案及各方意见，工作组开展了进一步调查研究，通过多轮内部研讨和标准内容文件修订，形成征求意见稿。

四、制定规范的原则和依据，与现行法律、法规、标准的关系

在制定临床科研数据集技术规范时，主要遵循以下原则：

科学性：规范应基于临床医学特点，确保数据的科学性、准确性和可靠性。数据元的选择和定义应经过严格论证，确保符合医学研究和临床实践的需求。

规范性：遵循国家关于数据标准制定的相关法律法规和标准，确保规范内容的一致性和规范性。通过统一的术语定义、数据元属性和值域代码表，提高数据集的可读性和互操作性。

实用性：规范应考虑实际应用的需求，确保数据集能够满足医学科研、数据处理等多个领域的需求。同时，规范应具有可操作性，便于数据集的采集、存储、共享和分析。

可扩展性：考虑到疾病研究和临床实践的不断进步，规范应具有可扩展性，能够容纳新的数据元和属性，确保数据集能够持续更新和完善。

本标准根据《中华人民共和国标准法》及其《实施细则》、《标准化工作导则第 1 部分：标准的结构和编写》GB/T 1.1-2020 进行编制。本规范的内容参考 GB/T

2659 世界各国和地区名称代码、GB/T 3304 中国各民族名称的罗马字母拼写法和代码、GB/T 6565-2015 职业分类与代码、GB/T 14396 疾病分类与代码、WS/T 364.12 卫生信息数据元值域代码 第12部分：计划与干预、WS445 电子病历基本数据集、《常用临床医学名词（2023年版）》、WS/T 370-2022 等内容。并结合医学研究和临床实践，确保规范能够真实反映疾病的病理特点和诊疗需求，考虑数据集在医学科研、数据处理等领域的应用需求，确保规范具有实用性和可操作性。

临床科研数据集技术规范与现行法律、法规、标准的关系主要体现在以下几个方面：

符合法律法规要求：规范内容严格遵守国家关于数据管理和标准化的相关法律法规，确保数据集的合法性和合规性。

与国家标准相衔接：规范在制定过程中，参照了《国家数据标准体系建设指南》等相关标准，确保数据集的标准体系框架与国家标准相一致。

补充完善行业规范：规范针对临床科研数据集的特点，制定了详细的数据元目录、数据元值域代码表和数据元索引，为行业内的数据管理和标准化提供了有益的补充和完善。

促进数据共享与互操作：通过制定统一的数据集技术规范，有助于提高数据集的可读性和互操作性，促进数据在医学科研、数据处理等领域的共享与应用。

五、主要条款的说明，主要技术指标、参数、试验验证的论述

本标准适用于指导临床科研数据集中患者基本信息的数据采集、存储、共享以及信息系统的开发。它适用于医疗保健机构、科研机构、信息系统开发商等在进行临床科研活动时的数据管理和交换，为实现医疗数据的标准化处理提供了框架和指导。本标准的制定结合了国内医疗解决方案提供商、医院和医疗研究机构的多方经验，相关内容已经在多种医疗研究场景下得到应用验证，具有普遍适用性。

主要内容：本团体标准的主要技术内容围绕着临床科研活动由15个部分组成，包括患者主记录信息表、就诊信息表、诊断信息表、用药医嘱信息表、处置医嘱信息表、手术信息表、费用信息表、体征信息表、护理评估信息表、病历文书信息表、检查报告信息表、检验报告信息表、临床结局信息表、病史和

风险因素信息表、数据集说明等。

每个部分内容涵盖了数据元属性种类、属性名称和属性值等公用属性，内部标识符、数据元标识符、名称、定义、数据类型、标识格式和允许值等数据元专用属性，以及数据元值域代码表。

六、重大意见分歧的处理依据和结果的说明

本规范研究制定过程中未出现重大分歧。

七、采用国际标准或国外先进标准程度的说明，以及国内、外同类产品或标准的对比情况

临床科研数据集的研究和标准制定是全球医疗数字化领域的重要课题。

(1) 国内研究情况

在中国，随着医疗数字化的快速发展，临床科研数据集技术规范的研究和标准制定受到了高度重视。国内的研究和实践主要集中在如何构建统一的医疗数据平台、如何确保数据的安全性和隐私保护、以及如何提高数据的互操作性和利用效率等方面。

相关部门和机构积极开展标准化工作，制定了一系列与医疗数据相关的标准，如 WS 370 卫生信息基本数据集编制规范，WS 445 电子病历基本数据集；GB/T 14396 疾病分类与代码等。这些标准都促进了医疗数据的标准化和共享。

但目前的相关标准主要面向临床医疗活动，其中申请、排队等大量业务流程数据不在临床科研应用场景中，且数据组织形式依据业务场景，区分体检、门急诊和住院，本标准的数据组织形式依据数据结构构建；此外，临床医疗活动结果数据不可直接应用于临床科研，存在部分内容颗粒度过粗的情况，在既有标准基础上新增相关数据元，例如科学研究样本的吸烟饮酒情况隐匿于个人史病历文书中，新增是否吸烟、吸烟年限、每日支数等数据元。

(2) 国际研究情况

在国际上，临床科研数据模型的研究同样活跃。例如，美国通过“健康信息技术促进临床研究”（Health Information Technology for Economic and Clinical Health, HITECH）法案推动了电子健康记录（EHR）的标准化和互操作性。国际标准化组织（ISO）和 HL7 国际等机构也在积极推动医疗信息标准化的工作，发

布了多项国际标准，如 HL7 的医疗信息交换标准、ISO 的医疗信息管理系统标准等。

(3) 技术进程及未来发展

目前，临床科研数据模型的技术正在不断发展和完善中。随着大数据、人工智能等技术的应用，数据模型的研究正朝着更加智能化、个性化的方向发展。未来，预计随着技术的成熟和医疗信息化的深入推进，临床科研数据模型将更加稳定和完美。这将有助于实现更高效的数据采集、处理和分析，为临床决策和科研提供更强有力的支持。

八、实施标准的措施建议

标准正在征求意见阶段，鼓励相关企业积极参与本标准的意见反馈工作。

九、其他说明

无。

《临床科研数据集技术规范》起草工作组

2024 年 11 月 21 日