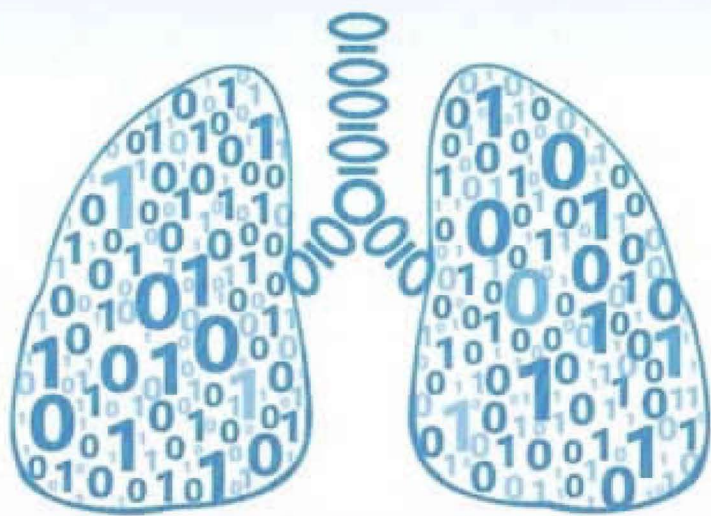


# 肺部肿瘤标准数据集

STANDARDIZED DATA SET FOR PULMONARY TUMORS

(2023版)

主审 / 白春学 主编 / 曾奕明 李 强 王向东



上海科学技术出版社

# 肺部肿瘤标准数据集(2023 版)

主 审 白春学  
主 编 曾奕明 李 强 王向东

上海科学技术出版社

## 图书在版编目 (CIP) 数据

肺部肿瘤标准数据集 : 2023版 / 曾奕明, 李强, 王向东主编. — 上海 : 上海科学技术出版社, 2023. 8  
ISBN 978-7-5478-6262-9

I. ①肺… II. ①曾… ②李… ③王… III. ①肺肿瘤—标准—数据集 IV. ①R734.2-65

中国国家版本馆CIP数据核字(2023)第134872号

### 肺部肿瘤标准数据集(2023版)

主编 曾奕明 李强 王向东

上海世纪出版(集团)有限公司 出版、发行  
上海科学技术出版社

(上海市闵行区号景路159弄A座9F-10F)

邮政编码 201101 www.sstp.cn

印刷

开本 787×1092 1/16 印张 4.75

字数 40千字

2023年8月第1版 2023年8月第1次印刷

ISBN 978-7-5478-6262-9/R·2803

定价: 58.00元

本书如有缺页、错装或坏损等严重质量问题, 请向印刷厂联系调换

## 内 容 提 要

本书参考了国内外最新肺部肿瘤诊疗指南, 以临床数据交换标准协会(CDISC)数据标准为基础, 对在肺部肿瘤临床诊断、治疗、随访等医疗活动过程中所产生的文本、数值等原始医疗资料通过结构化、归一、映射、逻辑计算等处理, 形成由模块名、序号、子模块、数据元、值域、数据类型和数据加工类型等构成的数据集。结合临床实践和临床研究的实际需求, 建立了肺部肿瘤真实世界研究领域统一、规范的数据元标准, 纳入包括人口学信息、就诊信息、既往史、个人史、家族史、诊疗史、现病史、体格检查、诊断评估、实验室检查、影像学检查、病理与分子检测、手术治疗、药物治疗、呼吸介入治疗、放射治疗、随访信息等17个数据模块。其中支气管镜和介入治疗是本书的一大特色模块, 对临床诊疗更具参考价值。

本书可供肺部肿瘤领域各级别医疗机构的管理人员、科研人员和临床医师阅读参考。

# 编委会

**主 审** 白春学(复旦大学附属中山医院)

**主 编** 曾奕明(福建医科大学附属第二医院)  
王向东(复旦大学附属中山医院)

**副主编** 徐 源(福建医科大学附属第二医院)  
杨栋勇(福建医科大学附属第二医院)

**编 委** (以姓氏拼音为序)

蔡思清(福建医科大学附属第二医院)  
陈路阳(福建医科大学附属第二医院)  
胡成平(中南大学湘雅医院)  
李伯义(福建医科大学附属第二医院)  
苏应瑞(福建医科大学附属第二医院)  
杨建胜(福建医科大学附属第二医院)  
查金顺(福建医科大学附属第二医院)

**组织编写** 福建医科大学附属第二医院  
数字福建肿瘤大数据研究所

**技术支持** 上海森亿医疗科技有限公司

李 强(同济大学附属东方医院)

胡志坚(福建医科大学公共卫生学院)

陈 琳(福建医科大学附属第二医院)  
何 斐(福建医科大学公共卫生学院)  
赖清泉(福建医科大学附属第二医院)  
林贤斌(福建医科大学附属第二医院)  
吴炜景(福建医科大学附属第二医院)  
余 晶(上海森亿医疗科技有限公司)  
张华平(福建医科大学附属第二医院)

中国肺癌防治联盟

# 序 言

肺癌是全球发病率和死亡率较高的恶性肿瘤之一。近年来我国肺部恶性肿瘤的发病率呈明显上升趋势,国内外一直致力于通过筛查来实现肺部恶性肿瘤的早诊早治,以期降低其死亡率。肺部恶性肿瘤严重影响患者的生活质量,造成巨大的社会和家庭负担。建立标准化数据库是肺部肿瘤防治体系中重要的基础性工程,对提升临床研究水平、加快临床科研成果转化、提高肺部肿瘤疗效,具有重大的意义。

肺部肿瘤相关的流行病学、危险因素、诊疗措施、预防干预等方面的研究都离不开宝贵的卫生健康与临床医疗数据的支持,规范化的数据采集和关联对于推进肺部肿瘤疾病相关的基础及临床研究至关重要,但是数据的标准化问题值得关注。随着信息技术的高速发展,医疗大数据和人工智能已在许多疾病的诊断和治疗中发挥重要作用,帮助临床医师总结经验,提升疾病的诊治水平,达到精准治疗目的,进而实现赋能临床、科研和教学等多个环节。但目前全国医院信息化建设标准与规范尚不统一,不同医院,不同承建厂商信息系统的数据结构和标准存在差异,数据交互、整合和共享存在困难,导致信息“孤岛”现象严重。另外,由于缺乏数据整合、转化的平台,临床资源缺乏规范化管理,无法进行深层次研究分析,导致对疾病的风险评估与临床诊断的认识不足,进一步影响疾病的干预与临床研究成果的转化。

标准化的肺部肿瘤数据集是肺癌等肺部恶性肿瘤研究的重要信息资源,是系统推进标准化工作流程的重要环节,也是顺应我国当前医疗形势和发展的内在要求,能不断促进肺部肿瘤精准医学研究的发展、推进临床诊疗技术智能化的过程。因此,构建肺部肿瘤标准化数据库具有重要的现实意义。通过构建《肺部肿瘤标准数据集》对肺部肿瘤疾病相关的数据收集、整合和管理,打破数据壁垒,形成长期、连续、动态、多源、大规模的专病数据积累,将促进肺部肿瘤相关疾病的基础研究与临床研究,也对呼吸系统疾病信息资源的合理利用起到积极推动作用。

《肺部肿瘤标准数据集》的发布,将有效促进肺部肿瘤疾病的质量控制和临床数据资源的整合与利用,助力我国肺部肿瘤疾病的数据层面规范管理,促进丰富的临床病例资源向宝贵的医学研究资源转化,对我国肺部肿瘤疾病的标准化、规范化、同质化发展有着重要的意义。

李强



2023年5月

# 前 言

肺癌是最常见、致死率较高的癌症之一。根据国际癌症研究机构 2020 年的数据统计,肺癌是全球癌症新发病例位居第二的癌种,约有 220 万新发病例;同时,肺癌是癌症死亡的首要原因,每年约有 180 万死亡病例。在我国,肺癌是男性发病率最高的癌种,也是癌症首位死因。根据国家癌症中心的数据估计,2022 年中国将约有 87 万肺癌新发病例,同时约有 77 万的死亡病例。

近年来,随着肺癌的诊断和治疗快速发展,以及早发现、早诊断、早治疗的不断普及,我国肺癌的 5 年生存率也从 16.1% 提高到 19.7%, 但与国际先进水平相比,我国在肺部肿瘤的防治水平方面仍有较大的差距。因此,紧跟肺部肿瘤防治学术技术前沿,进一步规范诊疗,不断推动肺部肿瘤的临床与基础研究,任重道远。

肺部肿瘤专病数据库是临床诊疗、科学研究、医学教育重要的资源平台,而肺部肿瘤标准数据集则是建设肺部肿瘤专病数据库的重要工具。

《肺部肿瘤标准数据集》参考国内外最新肺部肿瘤诊疗指南,结合临床实践和临床研究的实际需求,以临床数据交换标准协会(CDISC)数据标准为基础,对在肺部肿瘤临床诊断、治疗、随访等医疗活动过程中所产生的文本、数值等原始医疗资料,通过结构化、归一、映射、逻辑计算等处理,形成由模块名、序号、子模块、数据元、值域、数据类型和数据加工类型等构成的数据集。

随着导航、内镜超声、隧道活检等介入诊断技术,以及消融、冷冻、激光、光动力等介入治疗技术应用于临床,以经内镜介入、经皮介入、经血管介入为主要途径的一系列呼吸介入诊治技术在中晚期肺癌甚至在不可手术的早期肺癌的分子诊断与精准治疗中,发挥越来越重要的作用,切实提高了中晚期肺癌患者的生活质量、延长了生存期。

呼吸介入诊治技术已成为肺部肿瘤诊疗体系的重要组成部分,本数据集结合编撰专家在介入呼吸病学领域的丰富经验,将呼吸介入诊治相关的临床原始数据及资料作为特别关注的模块录入本数据集,最终形成《肺部肿瘤标准数据集》,该模块是本数据集最主要的特点。

本数据集主要包括：人口学信息、就诊信息、五项病史、手术、病理与分子检测、介入治疗、随访信息等 17 个模块、700 多个数据单元，数据集完整、准确。本书的读者对象主要为各级别医疗机构的管理人员、临床医师、科研人员。本数据集旨在指导临床医疗机构肺癌专病大数据平台的建设，基于数据统一以优化诊疗规范，并促进临床研究和转化，使不同区域的患者享受到均质化的医疗服务。

临床学术与技术的发展日新月异，专病数据集及专病数据库的建设也必须与时俱进，不断地在临床实践中加以维护、更新和升级，只有这样才能保证其为临床、教学、研究提供准确、优质、高效的数据服务。

由于专病数据集及专病数据库的建设仍需在实践中不断完善和发展，同时也由于笔者经验尚有不足，本数据集可能存在不完善之处，我们愿虚心接受同行的批评和指正。

本数据集的编撰得到了统计学、影像学、介入呼吸病学、分子诊治等相关领域专家的大力支持。在此，我们对所有参与编写的临床医学和相关领域专家的辛勤付出表示衷心感谢！

曾奕明



2023 年 6 月

# 目 录

---

一 数据集说明 / 1

二 肺部肿瘤标准数据集 / 3

1. 人口学信息 / 3

2. 就诊信息 / 5

3. 既往史 / 7

4. 个人史 / 9

5. 家族史 / 11

6. 诊疗史 / 12

7. 现病史 / 13

8. 体格检查 / 17

9. 诊断评估 / 19

10. 实验室检查 / 22

11. 影像学检查 / 30

12. 病理与分子检测 / 42

13. 手术治疗 / 48

14. 药物治疗 / 52

15. 呼吸介入治疗 / 58

16. 放射治疗 / 61

17. 随访信息 / 63

主要参考文献 / 64

## 数据集说明

本数据集参考国内外最新肺部肿瘤诊疗指南,结合临床研究的实际需求,以临床数据交换标准协会(CDISC)数据标准为基础,由福建医科大学附属第二医院、福建省呼吸医学中心专家共建而成。数据集由模块名、序号、子模块、数据元、值域、数据类型和数据加工类型(数据提取方式)等构成,总共包括 17 个数据模块、703 个数据元。

**模块名:** 用于区别数据元的一级分类,由一个或多个词构成。

**子模块:** 子模块从属于模块名,同一个模块名下可有多个并列的子模块,由一个或多个词构成。

**参考标准:** 主要参考美国卫生信息传输标准(Health Level 7, HL7)、临床试验数据采集标准(CDASHIG v2. 2 2021. 9. 28)及中华人民共和国卫生行业标准等。

**数据元:** 每个模块下所包含详细的字段,如“人口学信息”数据模块包含姓名、性别、出生日期、国籍等多个字段。

**值域:** 参考指南和文献,兼顾本领域专家认可的实用性,列举出字段中可能出现的内容。

**数据类型:** 用于标识数据元的数据库存储方式。

**数据加工类型:** 根据数据来源及数据加工处理方式,数据加工类型可分为:① 直接映射存储规范的数据,如:医嘱与检验数据。② 需要将来源不同的数据或多份病历进行数据计算与逻辑判断,如:术后住院天数通过计算手术时间与出院时间得出。③ 需要通过结构化和归一算法,将自然语言处理成可进行统计分析的标准字段与阈值。

**数据集更新机制:** 本数据集拟根据疾病诊疗指南标准和实际临床、科研需求定期进行更新,包括版本、修订内容、修订原因和修订时间。

**数据集模板使用权限(著作权):** 著作权及相关商标归福建省呼吸医学中心所有,使用本模板须经过著作权所有方同意,违者必究。

**数据集模块:**

<b>人口学信息</b> <ul style="list-style-type: none"> <li>姓名</li> <li>性别</li> <li>出生日期</li> <li>国籍</li> <li>.....</li> </ul>	<b>就诊信息</b> <ul style="list-style-type: none"> <li>就诊方式</li> <li>健康卡号</li> <li>病案号</li> <li>入院日期</li> <li>.....</li> </ul>	<b>既往史</b> <ul style="list-style-type: none"> <li>肺部疾病史</li> <li>肿瘤患病史</li> <li>其他疾病史</li> <li>.....</li> </ul>	<b>个人史</b> <ul style="list-style-type: none"> <li>吸烟史</li> <li>饮酒史</li> <li>过敏史</li> <li>接触史</li> <li>.....</li> </ul>	<b>家族史</b> <ul style="list-style-type: none"> <li>肿瘤家族史</li> <li>亲属关系</li> <li>肿瘤名称</li> <li>亲属是否健在</li> <li>.....</li> </ul>	<b>诊疗史</b> <ul style="list-style-type: none"> <li>就诊时间</li> <li>就诊机构</li> <li>治疗情况</li> <li>治疗方案</li> <li>.....</li> </ul>	<b>现病史</b> <ul style="list-style-type: none"> <li>有无症状</li> <li>肺部症状</li> <li>肺外症状</li> <li>诊疗经过</li> <li>.....</li> </ul>	<b>体格检查</b> <ul style="list-style-type: none"> <li>体温</li> <li>呼吸频率</li> <li>脉搏</li> <li>血压</li> <li>.....</li> </ul>	<b>诊断评估</b> <ul style="list-style-type: none"> <li>临床诊断</li> <li>主诊断名称</li> <li>主诊断时间</li> <li>ICD编码</li> <li>.....</li> </ul>
<b>实验室检查</b> <ul style="list-style-type: none"> <li>肿瘤标志物</li> <li>血常规</li> <li>生化检查</li> <li>凝血功能</li> <li>.....</li> </ul>	<b>影像学检查</b> <ul style="list-style-type: none"> <li>胸部CT检查</li> <li>PET-CT检查</li> <li>骨扫描</li> <li>磁共振检查</li> <li>.....</li> </ul>	<b>病理与分子检测</b> <ul style="list-style-type: none"> <li>术中冰冻病理</li> <li>病理检查</li> <li>免疫组化</li> <li>基因检测</li> <li>.....</li> </ul>	<b>手术治疗</b> <ul style="list-style-type: none"> <li>手术信息</li> <li>麻醉相关信息</li> <li>术后并发症</li> <li>.....</li> </ul>	<b>药物治疗</b> <ul style="list-style-type: none"> <li>化疗药物</li> <li>靶向药物</li> <li>抗血管生成药物</li> <li>免疫药物</li> <li>.....</li> </ul>	<b>呼吸介入治疗</b> <ul style="list-style-type: none"> <li>支气管镜治疗</li> <li>血管介入治疗</li> <li>经皮肺穿术</li> <li>经皮消融术</li> <li>.....</li> </ul>	<b>放射治疗</b> <ul style="list-style-type: none"> <li>粒子治疗</li> <li>放射治疗</li> <li>.....</li> </ul>	<b>随访信息</b> <ul style="list-style-type: none"> <li>随访日期</li> <li>随访次数</li> <li>随访评价</li> <li>随访计划</li> <li>.....</li> </ul>	

## 肺部肿瘤标准数据集

### 1. 人口学信息

模块名	内 容	参 考 标 准
人口学信息	姓名、性别、出生日期、年龄、国籍、民族、证件类型、证件号码、婚姻状况、文化程度、职业类型、工作单位、ABO血型、RH血型、出生地、籍贯、现住址、本人电话、联系人姓名、联系人与患者关系、联系人电话、其他联系方式	① 中华人民共和国卫生行业标准 WS 445.10 - 2014 电子病历基本数据集 第 10 部分：住院病案首页。② 中国卫生信息数据元值域代码。③ 临床试验数据采集标准(CDASHIG v2.2 2021.9.28)

序号	子 模 块	数 据 元	值 域	数据类型	数据加工类型
1.1	基本信息	姓名		文本	映射
1.2	基本信息	性别	男,女	文本	映射
1.3	基本信息	出生日期	YYYY-MM-DD	日期	映射
1.4	基本信息	年龄(岁)		数值	逻辑计算
1.5	基本信息	国籍	中国,其他	文本	映射
1.6	基本信息	民族	汉族,其他	文本	映射

续 表

序号	子模块	数据元	值 域	数据类型	数据加工类型
1.7	基本信息	证件类型	身份证,护照	文本	映射
1.8	基本信息	证件号码		文本	映射
1.9	基本信息	婚姻状况	未婚,已婚,离婚,丧偶,不详	文本	映射
1.10	基本信息	文化程度	文盲,小学,初中,高中,中专,大专,大学,研究生及以上	文本	映射
1.11	基本信息	职业类型	无业人员,退(离)休人员,农民,企业管理人员,自由职业者,国家公务员,专业技术人员,职员,现役军人,工人,学生,个体经营者,其他	文本	映射
1.12	基本信息	工作单位		文本	映射
1.13	基本信息	ABO 血型	A,B,AB,O,不详	文本	映射
1.14	基本信息	RH 血型	阴性,阳性,不祥	文本	映射
1.15	基本信息	出生地		文本	映射
1.16	基本信息	籍贯		文本	映射
1.17	基本信息	现住址		文本	映射
1.18	基本信息	本人电话		文本	映射
1.19	联系人信息	联系人姓名		文本	映射
1.20	联系人信息	联系人与患者关系	父,母,兄弟,姐妹,子女,夫妻,其他	文本	映射
1.21	联系人信息	联系人电话		文本	映射
1.22	联系人信息	其他联系方式		文本	映射

## 2. 就诊信息

模块名	内 容	参 考 标 准
就诊信息	就诊记录、住院信息、门诊信息	① 中华人民共和国卫生行业标准 WS 445.10 - 2014 电子病历基本数据集 第 10 部分：住院病案首页。 ② 中国卫生信息数据元值域代码

序号	子 模 块	数 据 元	值 域	数据类型	数据加工类型
2.1	就诊记录	就诊方式	住院,门诊,急诊	文本	映射
2.2	就诊记录	健康卡号		文本	映射
2.3	住院信息	病案号		文本	映射
2.4	住院信息	入院日期	YYYY-MM-DD	日期	映射
2.5	住院信息	出院日期	YYYY-MM-DD	日期	映射
2.6	住院信息	总住院天数(天)		数值	逻辑计算
2.7	住院信息	入院科室		文本	映射
2.8	住院信息	出院科室		文本	映射
2.9	住院信息	转科时间	YYYY-MM-DD hh:mm:ss	日期	映射
2.10	住院信息	住院次数(次)		数值	逻辑计算
2.11	住院信息	术后住院天数(天)		数值	逻辑计算
2.12	住院信息	是否转入 ICU/RICU	是,否	文本	逻辑计算

续 表

序号	子模块	数据元	值 域	数据类型	数据加工类型
2.13	住院信息	ICU/RICU 住院天数(天)		数值	逻辑计算
2.14	住院信息	总费用(元)		数值	映射
2.15	住院信息	自付金额(元)		数值	映射
2.16	住院信息	付费方式	无(自费),城镇职工基本医疗保险,城镇居民基本医疗保险,新型农村合作医疗,商业保险,其他	文本	映射
2.17	住院信息	离院方式	医嘱离院,医嘱转院,医嘱转社区卫生服务机构或乡镇卫生院,非医嘱离院,死亡,其他	文本	映射
2.18	住院信息	转归情况	治愈,好转,未愈,死亡,其他	文本	映射
2.19	门诊信息	门诊号		文本	映射
2.20	门诊信息	就诊科室		文本	映射
2.21	门诊信息	就诊日期	YYYY-MM-DD	日期	映射
2.22	门诊信息	就诊医师		文本	映射

## 3. 既往史

模块名	内 容	参 考 标 准
既往史	其他肺部疾病、肿瘤患病史、其他疾病史	① 中华人民共和国卫生行业标准 WS 445.12 - 2014 电子病历基本数据集 第12部分：入院记录。② 美国卫生信息传输标准(Health Level 7, HL7)。③ 临床试验数据采集标准(CDASHIG v2.2 2021.9.28)

序号	子 模 块	数 据 元	值 域	数据类型	数据加工类型
3.1	其他肺部疾病史	其他肺部疾病	无,有,不详	文本	结构化
3.2	其他肺部疾病史	其他肺部疾病	慢性阻塞性肺疾病,肺结核,支气管扩张,哮喘,肺间质纤维化,气胸,硅肺,其他	文本	结构化+归一
3.3	肿瘤患病史	肿瘤患病史	无,有,不详	文本	结构化
3.4	肿瘤患病史	肿瘤名称		文本	结构化+归一
3.5	肿瘤患病史	诊断日期	YYYY-MM-DD	日期	结构化
3.6	肿瘤患病史	治疗方案		文本	结构化+归一
3.7	其他疾病史	其他疾病史	无,有,不详	文本	结构化+归一
3.8	其他疾病史	心血管系统疾病		文本	结构化+归一
3.9	其他疾病史	内分泌系统疾病		文本	结构化+归一
3.10	其他疾病史	消化系统疾病		文本	结构化+归一
3.11	其他疾病史	肾脏疾病		文本	结构化+归一

续 表

序号	子模块	数据元	值 域	数据类型	数据加工类型
3.12	其他疾病史	血液系统疾病		文本	结构化+归一
3.13	其他疾病史	神经系统疾病		文本	结构化+归一
3.14	其他疾病史	免疫系统疾病		文本	结构化+归一