

肺癌病理标准数据集

(2023版)

主编 韩昱晨

主审 潘常青 张杰

组织编写

上海市胸科医院

技术支持

厦门极元科技有限公司

肺癌病理标准数据集 (2023 版)

主 编 韩昱晨
主 审 潘常青 张 杰
组织编写 上海市胸科医院
技术支持 厦门极元科技有限公司

上海科学技术出版社

图书在版编目 (CIP) 数据

肺癌病理标准数据集 : 2023版 / 韩昱晨主编. --
上海 : 上海科学技术出版社, 2023.9
ISBN 978-7-5478-6289-6

I. ①肺… II. ①韩… III. ①肺肿瘤—病理—标准—
数据集 IV. ①R734.2-65

中国国家版本馆CIP数据核字(2023)第154662号

肺癌病理标准数据集(2023版)

主编 韩昱晨

主审 潘常青 张杰

组织编写 上海市胸科医院

技术支持 厦门极元科技有限公司

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

(上海市闵行区景景路159弄A座9F-10F)

邮政编码 201101 www.sstp.cn

当纳利(上海)信息技术有限公司印刷

开本 787×1092 1/16 印张 4

字数: 83千字

2023年9月第1版 2023年9月第1次印刷

ISBN 978-7-5478-6289-6/R·2820

定价: 25.00元

本书如有缺页、错装或损坏等严重质量问题, 请向工厂联系调换

内 容 提 要

肺癌是全世界癌症相关死亡的主要原因, 发病率和死亡率均居国人恶性肿瘤首位。《WHO 胸部肿瘤分类》(第5版)对肺腺癌的病理诊断与分级进行了更新, 提出了不典型腺瘤样增生(AAH)、原位腺癌(AIS)和微浸润性腺癌(MIA)等病理概念。本书基于《WHO 胸部肿瘤分类》(第5版)中的肺癌病理分型提供一个全面的肺癌病理标准数据字段集, 以实现肺癌病理数据库的标准化和结构化, 为临床研究提供有利用价值的科研数据, 并为后续大数据分析、人工智能等技术提供基础, 服务临床更高效地开展肺癌诊疗工作。此外, 本书还涵盖了分子信息、人口学信息、诊断、检查、检验、影像学、药物治疗、手术治疗、随访等多方面的信息, 以便读者全面了解肺癌病理相关知识, 并提供实用和标准的字段信息, 帮助读者更好地理解 and 收集相关数据。

本书适合对肺癌病理学有一定基础的读者, 包括医学生、病理学家、肿瘤科医生、生物医学工程师、生物信息学家以及从事肺癌研究和治疗的其他专业人员。此外, 需要收集和整理肺癌及病理相关数据的人员, 例如研究员、数据分析师和科研工作者等, 也可以从本书中获益。

编委会

主 编 韩昱晨

主 审 潘常青 张 杰

副主编 侯旭敏 于永春 方文涛 陆 舜 钟 华

编 委 (按姓氏笔画排序)

向 婵 汤钦华 杨文娴 李 榕 汪 刚 宓林晖 赵继开 赵瑞英 俞容山
施焱鑫 袁骏毅 商占仙 滕昊骅

感谢厦门极元科技有限公司对本数据集提供技术支持

专业标准数据集说明

《肺癌病理标准数据集(2023 版)》模块参考了国家电子病历及信息化行业标准,以及最新的肺癌病理诊断指南与共识,由上海市胸科医院韩昱晨教授牵头,联合厦门极元科技有限公司共同编写。

I 数据模块与子模块展示

一、病理检查信息表	二、人口学信息采集表	三、诊断信息表
<ul style="list-style-type: none">● 病理信息● 病理类型● 腺癌● 鳞状细胞癌● 神经内分泌癌● 唾液腺型肿瘤● 其他肿瘤类型● 细胞学或其他特征● 分级● 病灶● 肿瘤尺寸● 疗效● 手术标本● 淋巴结● 肿瘤性其他发现● 非肿瘤性其他发现● 分期● 免疫组化(IHC)● PCR● NGS● NGS - SNV● NGS - CNV● FISH● 原位杂交	<ul style="list-style-type: none">● 患者人口学信息	<ul style="list-style-type: none">● 诊断信息

四、病史信息表	五、检验信息表	六、影像学检查信息表	七、其他检查信息表
<ul style="list-style-type: none"> ● 个人史 ● 家族史 	<ul style="list-style-type: none"> ● 血常规 ● 血生化 ● 尿常规 ● 凝血功能 ● 血清肿瘤标志物信息 	<ul style="list-style-type: none"> ● 影像学检查 	<ul style="list-style-type: none"> ● 心电图详情 ● 甲状腺功能检查 ● 体格检查 ● ECOG ● 生命体征
八、肺癌手术治疗信息表	九、肿瘤相关药物治疗信息表	十、放射治疗信息表	十一、随访信息表
<ul style="list-style-type: none"> ● 手术基本信息 ● 胸腔探查 ● 肺切除 ● 术中并发症 	<ul style="list-style-type: none"> ● 治疗信息 ● 治疗方案 	<ul style="list-style-type: none"> ● 治疗信息 	<ul style="list-style-type: none"> ● 随访信息 ● 血清肿瘤标志物检测

II 体例说明

本数据集共集成 11 个表单模块,446 个数据元。对部分术语解释如下。

1. 数据元

数据元是可以标识的数据集合,是通过一系列属性描述的数据单元,在特定条件下可被认为是不可再分的最小数据单位。此数据集中的数据元由编号、表单名称、子模块名称、关联项、字段名称、值域/数据类型和数据加工类型构成。因版面需要,表单名称只在该表单首页出现一次。

2. 值域/数据类型

参考主要指南及标准以及该疾病领域专家从实用性出发选定的值域/数据类型作为主要标准。

3. 单位名称

数据类型为“数值”的,必要时在其字段名称末端用括号标出单位名称,无单位的“数值”型数据则省略标注。

4. 内容层级关系说明

(1) “一、病理检查信息表”“二、人口学信息采集表”……中，“一”至“十一”为模块编号。

(2) “1.1”“1.2”……“2.1”“2.2”……中，“.”前的数字为全书 1~11 个表单的顺位，“.”后的数字为各表单内该数据元的顺位，完整的编号表明该数据元是第几个表单下第几个顺位的数据元，每个数据元在本数据集有唯一的编号。如“1.9”，字段名称是“腺癌组织学形态”，该数据元即“一、病理检查信息表”内第 9 个数据元。

(3) “关联项”指该字段与其他字段或其他字段下的值域/数据类型具有关联性，方便读者了解该字段的属性。例如：“1.13”数据元字段名称为“组织学形态亚型”，该关联项为“非黏液性腺癌”，表明该“组织学形态亚型”来自“非黏液性腺癌”。

III 数据集使用授权

使用《肺癌病理标准数据集(2023 版)》及其更新版，须经上海市胸科医院同意及授权。

前 言

医疗行业正快步迈入数字化时代,如何通过创新,提高医疗行业服务能力,缩小地域间医疗水平的差距,切实为医生和患者创造价值,将成为医疗行业数字化转型、形成智慧医疗的新型医疗服务模式的关键。数字病理将成为医疗行业在大数据时代的重要产物。数字病理的广泛应用,例如数字储存、远程病理诊断、基于人工智能的病理诊断、数字可视化判读分析等,将无限拓展医疗行业的未来。

数字病理的开展及人工智能的应用都离不开规范化、标准的数据集。目前在数据规范化、采集和存储方面存在很多问题。例如,真实世界数据缺乏规范的记录、采集和存储流程,数据存储分散、标准不一致,多为非结构化数据,通常需要通过繁琐的人工方式整理和分析数据,导致大量临床数据没有被充分挖掘,其价值未能得到充分利用。虽然国际上已有一些癌症数据库,如美国癌症基因组图谱(TCGA)、美国国家癌症数据库(NCDB)、国际癌症基因组联盟(ICGC),国内也有一些相对统一的标准数据集,但这些数据集主要针对多癌种或单一癌种,如肺癌的全科目标标准字段集,而专注于病理学的统一、标准的数据集非常少。

在国内,肺癌已成为最常见的癌症类型之一。国内的数字化医疗保健系统在肺癌的诊断和治疗方面取得了显著进展,但是在肺癌病理学方面仍存在字段信息不全面、不统一、信息缺失等问题。《肺癌病理标准数据集(2023版)》正是一本涵盖肺癌病理学的最新进展、提供全面的肺癌病理标准数据集的书。本书有助于读者了解肺癌的病理学特征,同时提供了一个包含了分子信息、人口学信息、诊断、检查、检验、影像学、药物治疗、手术治疗、随访等多方面的全面的标准字段集,为肺癌的研究和诊断提供了重要的数据采集依据。

本书的出版对肺癌病理学的数字化发展具有重要的意义,它将为病理医生、肿瘤学家、临床医生和医学研究人员等提供重要的参考资料,同时也将对医疗保健领域的数字化发展产生积极的推动作用。

主编 **韩昱晨**

2023年3月6日

目 录

- 一·病理检查信息表 /1
- 二·人口学信息采集表 /23
- 三·诊断信息表 /25
- 四·病史信息表 /27
- 五·检验信息表 /29
- 六·影像学检查信息表 /32
- 七·其他检查信息表 /34
- 八·肺癌手术治疗信息表 /37
- 九·肿瘤相关药物治疗信息表 /41
- 十·放射治疗信息表 /44
- 十一·随访信息表 /46
- 参考文献 /48

病理检查信息表

编号	子模块名称	关联项	字段名称	值域/数据类型	数据加工类型
1.1	病理信息		病理号	数值	结构化
1.2	病理信息		标本数量	数值	结构化
1.3	病理信息		编号	数值	结构化
1.4	病理信息		检查日期	YYYY-MM-DD	结构化
1.5	病理信息		标本类型	左全肺;左上叶;左下叶;右全肺;右上叶;右中叶;右下叶;左侧主支气管;右侧主支气管;右侧中间支气管;其他	结构化
1.6	病理信息		手术方式	楔形切除术;段切除术;肺叶切除术;双叶切除术;全肺切除术;主要气道切除术;袖状切除术;其他	结构化
1.7	病理信息		段切除术	尖段切除;后段切除;尖后段切除;前段切除;固有段切除;舌段切除;外侧段切除;内侧段切除;背段切除;基底段切除;内基底段切除;前基底段切除;内前基底段切除;外基底段切除;后基底段切除;星段切除;其他	结构化

(续表)

编号	子模块名称	关联项	字段名称	值域/数据类型	数据加工类型
1.8	病理类型		肿瘤病理类型	腺癌;鳞状细胞癌;神经内分泌肿瘤;唾液腺型肿瘤;大细胞癌;腺鳞癌;肉瘤样癌;NUT 癌;SMARCA4 缺失;良性肿瘤;异位肿瘤;转移性肿瘤;其他	结构化
1.9	腺癌		腺癌组织学形态	黏液性;非黏液性;肠型腺癌;胎儿型;胶状腺癌	结构化
1.10	腺癌		黏液性腺癌	浸润性黏液腺癌(IMA);混合性浸润性非黏液性和黏液性腺癌	结构化
1.11	腺癌		非黏液性腺癌	非典型腺瘤性增生(AAH);原位腺癌(AIS);微浸润性腺癌(MIA);浸润性腺癌(IA)	结构化
1.12	腺癌		胎儿型	高级别胎儿型;低级别胎儿型	结构化
1.13	腺癌	非黏液性腺癌	组织学形态亚型	贴壁;腺泡;复杂腺体结构;乳头状;微乳头状;实性	结构化
1.14	腺癌		贴壁成分占比(%)	数值	结构化
1.15	腺癌		腺泡成分占比(%)	数值	结构化
1.16	腺癌		复杂腺体结构占比(%)	数值	结构化
1.17	腺癌		乳头状成分占比(%)	数值	结构化
1.18	腺癌		微乳头状成分占比(%)	数值	结构化
1.19	腺癌		实性成分(%)	数值	结构化

(续表)

编号	子模块名称	关联项	字段名称	值域/数据类型	数据加工类型
1.20	腺癌	混合性浸润性非黏液性和黏液性腺癌	非黏液性占比(%)	数值	结构化
1.21	腺癌	混合性浸润性非黏液性和黏液性腺癌	非黏液性组织学形态亚型	贴壁;腺泡;复杂腺体结构;乳头状;微乳头;实性	结构化
1.22	腺癌	混合性浸润性非黏液性和黏液性腺癌	黏液性占比(%)	数值	结构化
1.23	腺癌	非黏液性腺癌	微浸润性腺癌(<3 cm 贴壁为主病变和≤5 mm 浸润)	贴壁生长;侵袭性成分	结构化
1.24	腺癌		侵袭性成分	腺泡;乳头状;微乳头状;实性;其他	结构化
1.25	鳞状细胞癌		鳞状细胞癌病理亚型	角化型鳞状细胞癌;非角化型鳞状细胞癌;基底细胞样鳞状细胞癌;原位鳞状细胞癌;淋巴上皮癌	结构化
1.26	鳞状细胞癌	鳞状细胞癌	肿瘤出芽	10 HPFs 中无出芽病灶;每 10 HPFs 中有<15 个出芽病灶;每 10 HPFs 中有≥15 个出芽病灶;无法评估	结构化
1.27	神经内分泌癌		神经内分泌癌病理亚型	小细胞癌;复合性小细胞癌;大细胞神经内分泌癌;复合性大细胞神经内分泌癌;典型类癌;非典型类癌	结构化
1.28	神经内分泌癌		复合性小细胞癌其他成分	腺癌;鳞状细胞癌;大细胞癌;大细胞神经内分泌癌;梭形细胞癌;巨细胞癌	结构化

(续表)

编号	子模块名称	关联项	字段名称	值域/数据类型	数据加工类型
1.29	神经内分泌癌		复合性大细胞神经内分泌癌其他成分	腺癌;鳞状细胞癌;巨细胞癌;梭形细胞癌	结构化
1.30	神经内分泌癌		典型类癌	核分裂象 $<2/2\text{ mm}^2$;无坏死;Ki67 增殖指数	结构化
1.31	神经内分泌癌	典型类癌	核分裂象($<2/2\text{ mm}^2$)	数值	结构化
1.32	神经内分泌癌	典型类癌	Ki67 增殖指数	数值	结构化
1.33	神经内分泌癌		非典型类癌	核分裂象 $>2/2\text{ mm}^2$ & $\leq 10/2\text{ mm}^2$;坏死;无坏死;Ki67 增殖指数	结构化
1.34	神经内分泌癌	非典型类癌	核分裂象($>2/2\text{ mm}^2$ & $\leq 10/2\text{ mm}^2$)	数值	结构化
1.35	神经内分泌癌	非典型类癌	Ki67 增殖指数	数值	结构化
1.36	唾液腺型肿瘤		唾液腺型肿瘤病理亚型	腺样囊性癌;上皮-肌上皮癌;黏液表皮样癌;玻璃样变透明细胞癌;肌上皮瘤肌上皮癌	结构化
1.37	其他肿瘤类型		其他肿瘤类型	腺鳞癌;肉瘤样癌;大细胞癌;转移癌;NUT 癌;胸部 SMARCA4 缺失未分化肿瘤	结构化
1.38	其他肿瘤类型	腺鳞癌	腺鳞癌成分	腺癌;鳞状细胞癌	结构化
1.39	其他肿瘤类型		腺癌占比	数值	结构化
1.40	其他肿瘤类型		鳞状细胞癌占比	数值	结构化

(续表)

编号	子模块名称	关联项	字段名称	值域/数据类型	数据加工类型
1.41	其他肿瘤类型	肉瘤样癌	肉瘤样癌亚型	多形性癌;肺母细胞瘤;癌肉瘤	结构化
1.42	其他肿瘤类型	多形性癌	多形性癌亚型	梭形细胞癌;巨细胞癌;多形性癌	结构化
1.43	其他肿瘤类型		肺母细胞瘤	文本	非结构化
1.44	其他肿瘤类型	肉瘤样癌	癌肉瘤	文本	非结构化
1.45	其他肿瘤类型	转移癌	转移癌来源	文本	非结构化
1.46	细胞学或其他特征		细胞学或其他特征	黏液细胞;印戒细胞;透明细胞;梭形细胞;巨细胞;桑葚体;淋巴间质	结构化
1.47	细胞学或其他特征		黏液细胞占比	数值	结构化
1.48	细胞学或其他特征		印戒细胞特征占比	数值	结构化
1.49	细胞学或其他特征		透明细胞特征占比	数值	结构化
1.50	细胞学或其他特征		梭形细胞特征占比	数值	结构化
1.51	细胞学或其他特征		巨细胞特征占比	数值	结构化
1.52	细胞学或其他特征		淋巴间质	无;中等;显著	结构化

(续表)

编号	子模块名称	关联项	字段名称	值域/数据类型	数据加工类型
1.53	分级		浸润性非黏液腺癌 分级	高分化(G1):附壁为主,高级别成分<20%;中分化(G2):腺泡或乳头为主,高级别成分<20%;低分化(G3):高级别成分≥20%(实性、微乳头状、筛状、复杂腺体结构等)	结构化
1.54	分级		其他病理类型组织学 分级	高分化;中分化;低分化;无法评估;不适用	结构化
1.55	病灶		病灶类型	原发;转移(肺内转移)	结构化
1.56	病灶		转移灶部位	伴有同一肺叶的肺内转移(pT3);同侧不同肺叶的肺内转移(pT4);对侧肺内转移(pM1a)	结构化
1.57	病灶		转移灶数量	数值	结构化
1.58	病灶		转移灶大小	数值	结构化
1.59	肿瘤尺寸		肿瘤大小测量方式	肉眼;镜下测量;影像学/CT测量	结构化
1.60	肿瘤尺寸		肿块最长径	数值	结构化
1.61	肿瘤尺寸		肿块最短径	数值	结构化
1.62	肿瘤尺寸		肿块最厚径	数值	结构化
1.63	肿瘤尺寸		肿瘤总大小	数值	结构化
1.64	肿瘤尺寸		其他尺寸	数值	结构化
1.65	肿瘤尺寸		浸润灶大小	数值	结构化

(续表)

编号	子模块名称	关联项	字段名称	值域/数据类型	数据加工类型
1.66	肿瘤尺寸		浸润灶/肿瘤(%)	数值	结构化
1.67	疗效		已知的术前治疗	有;无	结构化
1.68	疗效		肿瘤细胞	有存活;无存活	结构化
1.69	疗效	肿瘤细胞	存活肿瘤细胞的占比 (每10%递增)	数值	结构化
1.70	疗效		肿瘤坏死	有;无	结构化
1.71	疗效	肿瘤坏死	坏死的占比(每10% 递增)	数值	结构化
1.72	疗效		间质	有;无	结构化
1.73	疗效		间质的占比	数值	结构化
1.74	疗效		炎症	有;无	结构化
1.75	疗效	炎症	炎症程度	轻微;中度;重度	结构化
1.76	疗效		其他特征	文本	非结构化
1.77	疗效		肿瘤床大小	数值	结构化
1.78	疗效		存活肿瘤大小	明确尺寸;明确百分比	结构化
1.79	疗效		存活肿瘤大小的尺寸	数值	结构化
1.80	疗效		存活肿瘤大小的百 分比	数值	结构化