

临床研究

DOI: 10.13406/j.cnki.cyx.003227

肺实性结节半自动体积测量提高观察者间 Lung-RADS 评分一致性

刘会佳, 张 瑜

(空军军医大学西京医院放射科, 西安 710032)

【摘要】目的:研究 3D-Slicer 软件实性肺结节体积测量对不同观察者在肺部影像报告数据系统(lung CT screening reporting and data system, Lung-RADS)分类一致性中的影响。**方法:**纳入 76 例患者中的 76 个实性结节。由 3 位放射科医师分别采用手动和 3D-Slicer 软件半自动体积测量方法获得结节的直径与体积,并转化为相应的 Lung-RADS 评分,其中 2 分为阴性,3 分及以上为阳性。采用同类相关系数(intraclass correlation coefficient, ICC)及 Bland-Altman 指数来评价观察者间直径与体积测量的一致性, Kappa 分析评价观察者间 Lung-RADS 评分及阳性/阴性组间的一致性。**结果:**ICC 分析结果显示手动直径测量的一致性(0.994~0.996)明显低于半自动体积测量的一致性(0.997~0.998),同时 Bland-Altman 指数分析结果显示手动直径测量的偏倚高于半自动体积测量。采用半自动体积测量,能够比手动直径测量明显提高观察者间 Lung-RADS 评分及阳性/阴性之间的一致性(0.963~0.975 及 0.957~0.977 vs. 0.833~0.866 及 0.863~0.892)。**结论:**3D-Slicer 半自动体积测量能够提高实性肺结节观察者间大小测量的一致性,相应的 Lung-RADS 分类一致性也随之提高。

【关键词】肺部影像报告和数据系统;肺结节;癌症筛查;X 线计算机体层显像

【中图分类号】R445.3

【文献标志码】A

【收稿日期】2022-07-07

Semi-automatic volumetry of solid pulmonary nodules improves the interobserver consistency of Lung-RADS score

Liu Huijia, Zhang Yu

(Department of Radiology, Xijing Hospital of Air Force Medical University)

【Abstract】Objective: To study the effect of solid pulmonary nodules volume measurement by 3D-Slicer software on the interobserver consistency of lung CT screening reporting and data system (Lung-RADS) categories. **Methods:** Seventy-six solid nodules from 76 patients were included in this study. The diameter and volume of nodules were obtained by three radiologists using manual and 3D-Slicer software semi-automatic volume measurement methods, respectively, and the raw data was converted to the corresponding Lung-RADS score, of which 2 points was negative and 3 points or above was positive. Intraclass correlation coefficient (ICC) and Bland-Altman index were used to evaluate the interobserver consistency of diameter and volume measurements. Kappa analysis was used to evaluate the interobserver consistency of Lung-RADS score and the intergroup consistency of positive/negative. **Results:** ICC analysis showed that the consistency of manual diameter measurement (0.994~0.996) was significantly lower than that of semi-automatic volume measurement (0.997~0.998), and Bland-Altman index analysis showed that the bias of manual diameter measurement was higher than that of semi-automatic volume measurement. Compared with manual diameter measurement, the use of semi-automatic volume measurement can significantly improve the consistency of Lung-RADS score and positive/negative classifications (0.963~0.975 and 0.957~0.977 vs. 0.833~0.866 and 0.863~0.892). **Conclusion:** 3D-Slicer semi-automatic volume measurement can improve the interobserver consistency of the size measurement of solid pulmonary nodules, as well as the corresponding Lung-RADS categories.

【Key words】lung CT screening reporting and data system; pulmonary nodule; cancer screening; X-ray computed tomography

作者介绍: 刘会佳, Email: huijialiu@163.com,

研究方向: 呼吸系统、消化系统及泌尿生殖系影像诊断。

通信作者: 张 瑜, Email: zywcqq@163.com。

优先出版: <https://kns.cnki.net/kcms2/detail/50.1046.R.20230602.1413.012.html>

(2023-06-04)

肺癌是全球发病率第二、病死率第一的恶性肿瘤^[1]。大多数肺癌诊断时已到晚期,错过了根治机会。早期发现肺癌并积极干预,能够使患者获得最佳的生存获益。目前全球运用最广泛的 CT 肺癌筛查方法是根据美国放射学会提出的肺部影像报告数

据系统(lung CT screening reporting and data system, Lung-RADS)来对结节恶性可能性进行分层,评分越高,恶性可能性越高^[2]。根据密度高低,肺结节可分为实性结节与亚实性结节,体检发现的结节绝大多数是实性结节。亚实性结节的恶性程度明显低于实性结节,治疗时间窗可以很长,常采用积极的随访观察策略。但实性结节的恶性程度往往较高,很容易出现转移,及时诊断非常必要^[3]。

在 Lung-RADS 中,实性结节的大小,包括二维直径大小及三维体积大小是分类的一个重要因素。随着结节大小的增长,恶性可能性随之提高。二维直径大小用结节最大横断面长径与垂直于长径的短径平均值表示^[1]。然而,手动测量二维直径大小的观察者间一致性存在差异,造成相应的 Lung-RADS 评分不一致^[4]。

理论上,半自动三维体积测量能够比手动二维直径测量更好地代表结节的真实大小,因此,三维体积得出的 Lung-RADS 评分应该比二维直径更加准确,观察者间一致性更好。但是,相关的研究还比较少见。3D-Slicer 软件是一种成熟的、广泛运用的影像后处理软件,能够根据 CT 阈值测量一定范围内体素的容积。因此,本研究使用 3D-Slicer 软件测量实性结节体积,并与手动直径测量比较其对观察者间一致性的影响。

1 材料与方法

1.1 一般资料

搜集空军军医大学附属西京医院 2022 年 1 月至 5 月体检行胸部 CT 检查的患者临床及影像学资料,所有患者的报告均按照 Lung-RADS 评分由 2 名经验丰富的放射科主治医师审核签发。由另一名放射科主治医师按照报告结果进行病例筛选,排除 1 分(没有结节或者典型良性结节,即结节有特征性钙化及脂肪)结节;2 分中主危结节为裂旁结节(小于 10 mm 的裂旁结节均是良性结节,且大小不影响评分);4x 分(结节含恶性特征,如毛刺,此时结节大小相对不影响评分);CT 图像质量不能满足诊断要求的病例。最终共 76 例患者中的 76 例实性结节纳入本研究,其中男性 43 例,女性 33 例,40 例有吸烟史,患者平均年龄(53.1 ± 10.2)岁,范围 31~84 岁。76 例结节中 Lung-RADS 2 分 32 例,3 分 22 例,4A 分 15

例,4B 分 7 例,需要注意的是,此评分是原始报告的数据,并不作为金标准参考。

1.2 CT 检查

所有病例均采用 GE Light Speed VCT 64 排 CT 扫描仪在患者深吸气状态时进行低剂量扫描。管电压 120 kV,管电流设置自动毫安秒(范围 20~40 mA),螺距 0.948,扫描层厚 5 mm,标准算法重建,重建层厚及间距均为 0.625 mm,FOV 300~350 mm。

1.3 二维结节直径大小测量

由 3 名放射科医师(1 名主治医师,2 名住院医师)在 PACS 上独立阅片,阅片时屏蔽报告结果,仅告知结节的具体位置,在轴位最大面上测量结节的长径与垂直于长径的短径,取其平均值作为结节二维直径大小。3 名医生均熟练掌握 Lung-RADS 的评分标准。

1.4 三维结节体积大小测量

之前的研究报道以 -350 HU 作为软件测量部分实性结节中实性成分的阈值,与病理上实性成分大小的一致性较好^[5]。因此本研究中 3D-Slicer 软件测量实性结节时设置渲染阈值 > -350 HU,再调节渲染刷至合适粗细,选取结节任意一个层面进行渲染,范围要超过结节边界,剩余大部分层面软件可以自动识别边界,对于未能识别的部分需要手动修饰,渲染结束后可以一键计算所勾画结节体积。

1.5 统计学处理

结节二维直径与三维体积测量的观察者间一致性采用同类相关系数(intraclass correlation, ICC)及 Bland-Altman 指数评价。将 Lung-RADS 2 分归为阴性组,3 分、4A 分、4B 分归为阳性组, Lung-RADS 分类之间及阴性组与阳性组之间的观察者间一致性采用 Kappa 分析评价。统计软件采用 MedCalc 软件(版本:19.0.5)。

2 结果

2.1 结节二维直径与三维体积测量的一致性比较

尽管观察者间结节二维直径测量取得了非常高的一致性 (ICC=0.994~0.996),但仍然低于三维体积测量 (ICC=0.997~0.998),并且其 95% 可信区间没有重叠。Bland-Altman 检验示观察者 A 和观察者 B 二维体积测量的结果比观察者 C 小约 1.3%,而三维体积测量结果仅小约 0.4%,提示三维体积测量结果的一致性更高(表 1)。

2.2 Lung-RADS 分类

与二维直径测量相比,三维体积测量中阴性结节(2 分)更多,阳性结节(3 分、4A 分、4B 分)更少,但这种差异没有统计学意义(表 2)。

表 1 不同观察者间直径和体积测量的一致性分析

观察者	ICC		Bland-Altman 检验/%	
	直径测量	体积测量	直径测量	体积测量
B vs. A	0.996 (0.994~0.997)	0.998 (0.998~0.999)	-1.6 (-22.0~18.9)	-0.5 (-13.2~9.3)
C vs. B	0.994 (0.992~0.995)	0.997 (0.996~0.998)	-1.3 (-18.0~21.1)	-0.4 (-11.2~10.3)
C vs. A	0.994 (0.993~0.995)	0.997 (0.997~0.998)	-1.3 (-20.0~17.9)	-0.4 (-10.2~11.3)

表2 不同观察者间 Lung-RADS 分布

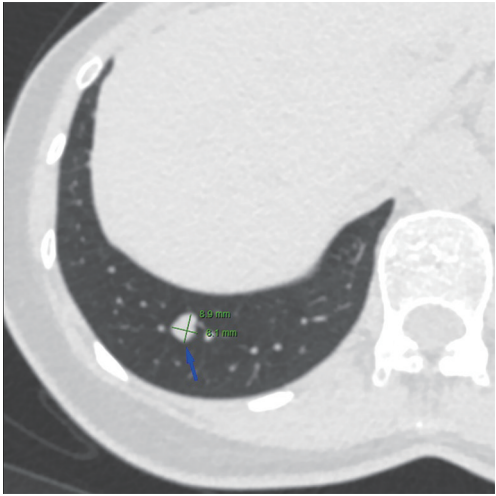
观察者	2分	3分	4A分	4B分	P值
A					
二维直径测量	32	21	14	9	0.631
三维体积测量	40	17	11	8	
B					
二维直径测量	32	20	18	6	0.611
三维体积测量	38	19	12	7	
C					
二维直径测量	29	25	15	7	0.501
三维体积测量	38	19	12	7	

2.3 Lung-RADS 评分一致性

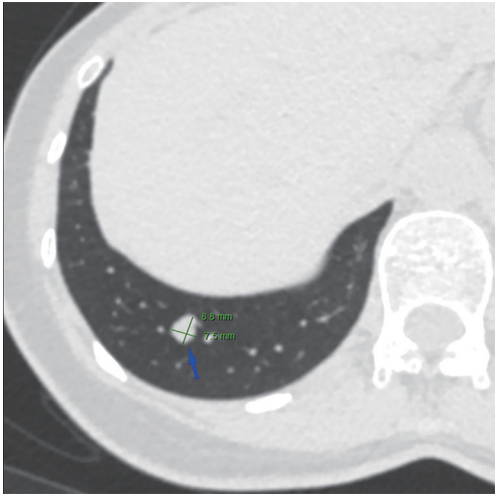
二维直径测量观察者间 Lung-RADS 评分间及阳性/阴性组间 Kappa 值范围为 0.833~0.866 及 0.863~0.892, 明显低于三维体积测量的 0.963~0.975 及 0.957~0.977, 并且其 95% 可信区间没有重叠(表3)。典型病例图像见图1, 患者女, 46岁, 体检发现右肺下叶实性结节(蓝色箭头), 3位观察者测量大小分别为图1A、B、C, 平均直径依次为 8.5、6.2、7.7 mm, 对应的 Lung-RDAS 依次为 4A、4A、3分。使用 3D-slicer 软件测量, 3位观察者测得结节体积分别为 200.9、202.3、195.7 mm³(后面2个数值未在此显示), 对应的 Lung-RDAS 均为 3分。

表3 不同观察者间 Lung-RDAS 分级的一致性分析

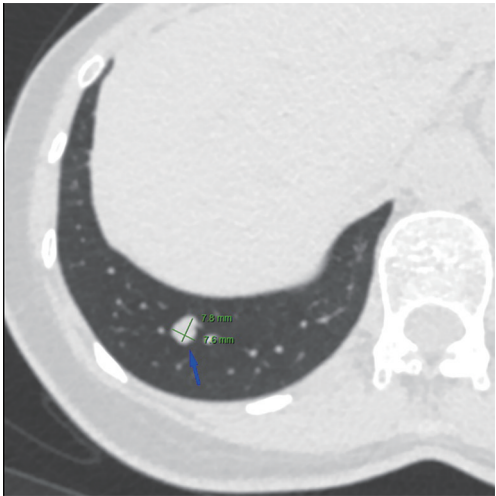
观察者	Lung-RADS 2分 vs. 3分 vs. 4A分 vs. 4B分		Lung-RADS 阳性 vs. 阴性	
	直径测量	体积测量	直径测量	体积测量
B vs. A	0.864(0.821~0.901)	0.963(0.921~1.000)	0.892(0.845~0.943)	0.967 (0.956~1.000)
C vs. B	0.866(0.831~0.920)	0.975(0.940~1.000)	0.889(0.864~0.933)	0.977 (0.976~1.000)
C vs. A	0.833(0.818~0.911)	0.963(0.921~1.000)	0.863(0.848~0.918)	0.957 (0.946~1.000)



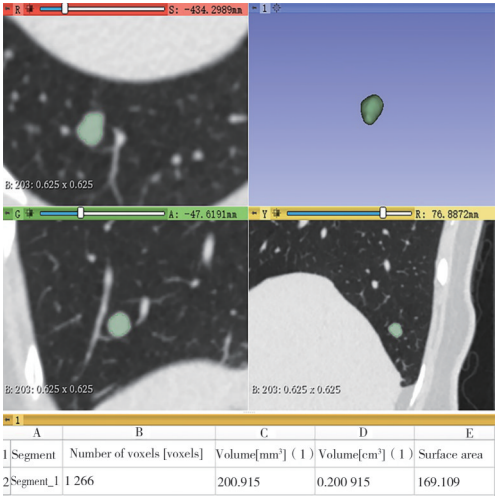
A. 测量大小为 8.9 mm × 8.1 mm



B. 测量大小为 8.8 mm × 7.5 mm



C. 测量大小为 7.8 mm × 7.6 mm



D. 3D-slicer 软件测量结节体积

图1 患者女, 46岁, 体检发现右肺下叶实性结节(蓝色箭头)

3 讨 论

肺结节的大小测量需要考虑 2 个非常重要的因素:首先是准确性,因为随着结节大小的增大,恶性可能性也越来越高。根据 Lung-RADS v1.1,小于 6 mm (113 mm^3) 的实性结节(2分)恶性可能性小于 1%,6~8 mm ($113\sim 268 \text{ mm}^3$) 的实性结节(3分)恶性可能性为 1%~2%,8~15 mm ($268\sim 1\,767 \text{ mm}^3$) 的实性结节(4A分)恶性可能性为 5%~15%,大于 15 mm ($1\,767 \text{ mm}^3$) 的实性结节(4B分)恶性可能性则大于 15%。不同的 Lung-RADS 评分对应不同的处理方法^[1]。因此,结节大小测量不准确有可能导致错误的评分及不同的处理方法。其次是观察者间的一致性,Lung-RADS v1.1 提出在随访中结节直径增长超过 1.5 mm,其评分会增高,恶性可能性会相应增大,这对测量方法一致性提出了更高的要求,一致性低的测量方法在前后对比时可能会产生假阳性或者假阴性结果。

由于大多数结节的形态不规则,三维体积能够比二维直径更好地代表结节的真实大小,对临床也更加有指导意义^[6]。另外,本研究表明,使用半自动肺结节三维体积测量法,能够比手动直径测量提高观察者间的一致性,相应的 Lung-RADS 评分一致性也有明显提高。并且不论对于大小测量的 ICC 值还是 Lung-RADS 评分的 Kappa 值,二维直径测量和三维体积测量的 95% 可信区间没有重叠,说明三维体积测量具有巨大优势。

既往也有一些报道用不同的方法来提高结节分类的一致性,如用纵隔窗上能够观察到实性成分这样的条件,明显提高了识别部分实性结节的观察者间一致性^[7]。Gierada DS 等^[4]的研究表明使用西门子 Syngo.via 软件半自动测量实性结节体积,能够提高实性结节观察者间 Lung-RADS 评分一致性。与 Gierada DS 等^[4]的报道相比,本研究有以下几点不同。首先,本研究使用的是 3D-Slicer 软件,在测量方法上相对于 Syngo.via 软件略复杂,需要对结节逐层渲染,但是作为一个开源且可以免费获取的软件,3D-Slicer 较 Syngo.via 更加普及,在肺结节的勾画及测量方面已有多篇报道^[8-9]。另外,虽然逐层渲染需要更多的操作,但是并不需要精确勾画结节,软件会根据阈值自动识别结节边缘,肺结节体积一般较小,所占层面不多,熟练后亦可非常迅速完成结节体积计算。其次,Syngo.via 软件并不能勾画出全部实性结节,经常会将结节周围高密度组织一同计算,从而造成勾画失败,对于这样的结节,之前的研究做了删除处理^[4],而 3D-Slicer 软件则相对不受影响。

三维体积测量降低了肺结节评分,使更多的检查被归为阴性结果,这与之前的研究基本一致^[4]。

笔者认为其原因在于体积阈值是根据直径阈值所对应的球形体积大小来定的,如直径 6 mm 的球体体积是 113 mm^3 ,分别为 2 分和 3 分的直径阈值和体积阈值。由于结节形态的不规则性,直径测量往往会过高估计结节的大小^[10]。

本研究有以下不足:首先,本研究是单中心回顾性研究,难免会出现一些偏倚,需要大样本前瞻性研究来验证;其次,3D-Slicer 软件需要手动测量结节体积,会被更加先进的全自动肺结节体积测量软件取代,然而全自动软件的勾画结果仍然需要医生确认,当出现识别不准确时修改勾画往往会比较麻烦;再次,二维直径测量受到结节形态的影响,形态越不规则测量可能越不准确,观察者间一致性越低^[11],但是形态学特点受主观评判影响很大,很难量化分析。

总而言之,用 3D-Slicer 软件对实性肺结节进行体积测量,能够比直径测量有更高的观察者间一致性,相应的 Lung-RADS 分类一致性也明显提高,这在肺癌筛查中有非常重要的临床实用价值。

参 考 文 献

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA A Cancer J Clin, 2018, 68 (6): 394-424.
- [2] American College of Radiology Web Site. Lung CT screening reporting and data system (lung-RADS) version 1.1[EB/OL]. (2019-05-31) [2022-03-01]. <http://www.acr.org/Quality-Safety/Resources/Lung-RADS>.
- [3] Wood DE, Kazerooni EA, Baum SL, et al. Lung cancer screening, version 3.2018, NCCN clinical practice guidelines in oncology[J]. J Natl Compr Canc Netw, 2018, 16(4):412-441.
- [4] Gierada DS, Rydzak CE, Zei M, et al. Improved interobserver agreement on lung-RADS classification of solid nodules using semiautomated CT volumetry[J]. Radiology, 2020, 297(3):675-684.
- [5] Cohen JG, Goo JM, Yoo RE, et al. Software performance in segmenting ground-glass and solid components of subsolid nodules in pulmonary adenocarcinomas[J]. Eur Radiol, 2016, 26(12):4465-4474.
- [6] Revel MP, Bissery A, Bienvenu M, et al. Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable?[J]. Radiology, 2004, 231(2):453-458.
- [7] Revel MP, Mannes I, Benzakoun J, et al. Subsolid lung nodule classification: a CT criterion for improving interobserver agreement[J]. Radiology, 2018, 286(1):316-325.
- [8] Yip SSF, Parmar C, Blezek D, et al. Application of the 3D slicer chest imaging platform segmentation algorithm for large lung nodule delineation[J]. PLoS One, 2017, 12(6):e0178944.
- [9] Xu FY, Zhu WC, Shen Y, et al. Radiomic-based quantitative CT analysis of pure ground-glass nodules to predict the invasiveness of lung adenocarcinoma[J]. Front Oncol, 2020, 10:872.
- [10] Heuvelmans MA, Walter JE, Vliegenthart R, et al. Disagreement of diameter and volume measurements for pulmonary nodule size estimation in CT lung cancer screening[J]. Thorax, 2018, 73(8):779-781.
- [11] Han DW, Heuvelmans MA, Vliegenthart R, et al. Influence of lung nodule margin on volume- and diameter-based reader variability in CT lung cancer screening[J]. Br J Radiol, 2018, 91(1090):20170405.

(责任编辑:周一青)