

不同阅片者对肺亚实性结节 临床处理一致性的影响

李焕国 崔凤 杨勇 瞿华 钟琦 邱煥 杜亮

【摘要】目的 分析不同放射诊断医生对肺亚实性结节的临床处理方法的一致性及其影响因素。**方法** 回顾 49 个亚实性肺结节的影像学资料,共有 3 家医院 8 位阅片者,包括 A 单位 6 位(高级职称:A1、A2;中级职称:A3、A4;初级职称:A5、A6)、B 与 C 单位各 1 位(记为 B1、C1)。所有阅片者需要完成的内容包括:结节类型、结节大小、临床处理方法,所有阅片者独立完成阅片,并在 1 个月后再次进行阅片,并记录上述内容。阅片者间分析采用第一次记录结果。**结果** 临床处理方法的一致性:阅片者 A5 的内部一致性最低,为 0.385(95%CI:0.132~0.638),其次为阅片者 A2,为 0.502(95%CI:0.226~0.778),其余阅片者的内部一致性均为好;A 单位高级、中级、初级职称阅片者间的一致性逐渐降低,前两者各组间的临床处理方法均无统计学差异(均 $P > 0.05$),后者有统计学差异($P < 0.05$);3 个单位同年资阅片者间一致性为 0.5(95%CI:0.338~0.662),临床处理方法有统计学差异(均 $P < 0.05$);通过对临床处理方法出现分歧的结节分析发现,同单位不同年资、不同单位的同年资阅片者间出现分歧的结节分类大部分为部分实性结节患者,不同的是前者对分歧结节的分类一致性较好(Kappa=0.754,95%CI:0.677~0.831),后者对分歧结节的分类一致性较差(Kappa=0.171,95%CI:-0.131~0.473),对于实性大小的测量两者的一致性均较差(Kappa=0.0909,95%CI:-0.088~0.270;Kappa=-0.335,95%CI:-0.677~0.007)。在 A 单位相同年资的分歧结节均为部分实性结节,对于实性部分的大小,两位阅片者一致性较差(Kappa=-0.849,95%CI:-1.294~1.463)。**结论** 肺结节的临床处理方法、分类以及实性成分的测量在不同阅片者间存在一定的差异,年资、工作单位都可能会影响肺结节的分类、测量以及临床处理策略。

【关键词】 亚实性肺结节 临床处理 一致性

Discrepancy in film reading of pulmonary subsolid nodules among different readers LI Huanguo, CUI Feng, YANG Yong, et al.
Department of Radiology, Hangzhou Municipal Hospital of Traditional Chinese Medicine, Hangzhou 310007, China

【Abstract】Objective To analyze the discrepancy in film reading of pulmonary subsolid nodules and clinical management among different readers. **Methods** The results of film reading of 49 pulmonary subsolid nodules were retrospectively analyzed. There were 8 readers from 3 hospitals, including 6 readers in hospital A (2 senior level readers: A1, A2, 2 intermediate level readers: A3, A4, 2 junior level readers: A5, A6), 2 readers from hospitals B and C (both were intermediate level: B1, C1). All readers were asked to independently complete the film reading and to report the nodular type, nodule size and clinical management, and the reading was repeated 1 month later. The intra-reader and inter-reader consistency was analyzed. **Results** The intra-reader consistency of clinical management varied from 0.385 (95%CI:0.132~0.638) to 0.791 (95%CI: 0.511~1.071). The intra-reader consistency of clinical management in senior, intermediate and junior readers were decreased with the decrease of seniority in hospital A, and there was no significant difference between senior and intermediate level readers ($P > 0.05$), but there was significant difference between intermediate and junior readers ($P < 0.05$). The inter-reader consistency of the readers with same seniority from 3 different hospitals was 0.5 (95%CI:0.338~0.662) and there was significant difference among them ($P < 0.05$). The part-solid nodules were most discrepant in clinical management strategies among the readers in the same hospital with different seniority and in different hospital with the same seniority, and the inter-reader consistency of these nodules in clinical management strategies was good (Kappa=0.754, 95%CI:0.677~0.831) for readers in the same hospital with different seniority, but that was poor (Kappa=0.171, 95%CI:-0.131~0.473) for readers in the different hospital with the same seniority. The inter-reader consistency of size measurement for part solid was poor (Kappa=0.0909, 95%CI: -0.088 to 0.270; Kappa=-0.335,

doi:10.12056/j.issn.1006-2785.2018.40.10.2018-728

基金项目:浙江省重大科技专项重大社会发展项目(2013C03044-3);浙江省中医药科研基金项目(2015ZA1160)

作者单位:310007 杭州市中医院放射科(李焕国、崔凤、瞿华、钟琦、邱煥、杜亮),胸外科(杨勇)

通信作者:杨勇,E-mail:yyyg1107@sina.com

95%CI: -0.677~0.007, respectively). In hospital A, nodules with discrepant clinical management strategies among same seniority were all part-solid nodules, and the inter-reader consistency of the size measurement for part solid was poor ($Kappa=-0.849$, 95%CI: -1.294 ~ 1.463). **Conclusion** There are differences in clinical management strategy, classification and measurement of solid components of pulmonary nodules among different readers. The seniority and hospital of readers may affect the results of pulmonary nodule reading.

【Key words】 Subsolid nodules Clinical management Agreement

随着低剂量 CT (low dose computed tomographic, LDCT) 在胸部检查中的广泛应用, 肺结节的检出率日益增加。根据 CT 图像上肺结节的不同表现分为实性结节和亚实性。目前针对肺结节的管理已有多个指南^[1-4], 均是通过肺结节在 CT 图像上的形态学特点进行良恶性的评估以及管理, 而目前对于肺结节的形态学评价主要依靠放射科诊断医生的主观判断。多项研究发现, 不同阅片者对于实性结节的评估具有较高的一致性, 而对于亚实性结节的分类及临床处理策略存在较大差异^[5-6]。在这些研究中, 阅片者来自不同的工作地点, 且为 10 年以上 (10~30 年) 高年资医生, 并未分析 10 年以下以及不同工作场所对结节分类及管理的影响情况。因此本研究通过不同医院的不同级别放射诊断医生对同一组亚实性结节进行评估, 以进一步分析影响阅片者间一致性的因素。

1 资料和方法

1.1 一般资料 回顾 2016 年 1 月至 2017 年 1 月在我院确诊为腺癌的亚实性肺结节患者 49 例, 共 49 个结节, 年龄 38~86 (55.57±11.67) 岁, 其中男 14 例, 年龄 40~86 (55.93±12.09) 岁, 女 35 例, 年龄 38~74 (55.63±11.16) 岁。

1.2 检查方法 检测前去除患者身上金属物品, 防止金属伪影, 并指导患者屏气训练。采用 Philips 256iCT、SIEMENS 16 排 CT (SOMATOM Perspective) 进行肺部扫描, 扫描范围从肺尖到肺底。扫描参数: (1) Philips 256iCT: 机架转速 0.27s/r, 探测器 128×0.625mm, 管电压 120kV, 自动管电流, 层厚 5mm, 螺距 1.0, 重建层厚 1.5mm, 重建层间距 1.5mm, 肺窗窗宽 1 500, 窗位-549; (2) SIEMENS 16 排 CT: 机架转速 0.35s/r, 探测器 16×0.75mm, 管电压 120kV, 自动管电流, 层厚 5mm, 螺距 1.0, 重建层厚 1.5mm, 重建层间距 1.5mm, 肺窗窗宽 1 500, 窗位-549; 扫描矩阵 512×512。

1.3 图像分析与处理 将所选取的患者进行随机编号, 将层厚为 1.5mm 的肺窗影像 (DICOM 格式) 导入相应的编号内, 在新的编号中滤掉患者的所有个人信息以

及原有的报告, 所有图像均可以调节窗宽窗位。共有 3 家医院 8 位阅片者, 包括 A 单位 6 位 (高级职称 A1、A2, 分别为 20 年、16 年工作经验; 中级职称: A3、A4, 均为 9 年工作经验; 初级职称: A5、A6, 5 年、3 年工作经验各 1 位)、B 与 C 单位各 1 位 (工作经验为 10 年, 记为 B1、C1)。所有阅片者需要完成的内容包括: 判断结节的类型 (实性、部分实性、纯磨玻璃)、测量大小 [大小=(长径+短径)/2, 包括整个结节的大小、部分实性结节的实性部分大小, 精确到 mm], 所有阅片者独立完成阅片, 并在 1 个月后再次进行阅片, 并记录上述内容。

亚实性结节的定义: 是指不能完全掩盖其内走行的血管、支气管影的直径介于 0.3~3.0cm 的圆形或不规则形状密度影, 包括纯磨玻璃密度结节和部分实性结节^[7]。

依据影像学结果, 结节的下一步临床处理分为 2 种: CT 随访 (纯磨玻璃影结节、实性成分 < 5mm 的亚实性结节及 ≤ 8mm 的实性结节) 和临床干预 (实性成分 ≥ 5mm 的亚实性结节, > 8mm 的实性结节)^[8]。阅片者中至少有一位与其它阅片者意见不一样, 即为分歧结节。

1.4 统计学处理 采用 R 统计软件的 Kappa 检验, 其中 Multirater Fleiss Kappa 检验用于分析阅片者之间的一致性, Cohen Kappa 检验用于分析阅片者自身的一致性, 并计算 95% 可信区间 (95%CI), 阅片者之间的一致性均采用第一次阅片结果进行比较分析; 一致性评价原则: 差 ($Kappa \leq 0.20$), 一般 ($Kappa = 0.21 \sim 0.40$), 中度 ($Kappa = 0.41 \sim 0.60$), 好 ($Kappa = 0.61 \sim 0.80$), 非常好 ($Kappa = 0.81 \sim 1.00$)。采用 SPSS 17.0 统计软件, 计数资料两组非参数配对检验采用 Wilcoxon 秩检验, 多组非参数配对检验采用 Friedman 检验。

2 结果

2.1 不同阅片者内部对肺结节临床处理方法的一致性比较 阅片者 A5 的内部一致性最低, 属于一般, 其次为阅片者 A2, 一致性为中等, 其余阅片者均为好, 详见表 1。

2.2 不同阅片者之间对肺结节临床处理方法的一致性比较 A 单位高级、中级、初级职称阅片者间的一致性

表 1 不同阅片者内部对肺结节临床处理方法的一致性比较

阅片者	Kappa 值	95%CI
A1	0.745	0.465 ~ 1.025
A2	0.502	0.226 ~ 0.778
A3	0.791	0.511 ~ 1.071
A4	0.68	0.401 ~ 0.959
A5	0.385	0.132 ~ 0.638
A6	0.674	0.395 ~ 0.953
B1	0.67	0.403 ~ 0.937
C1	0.68	0.423 ~ 0.937

逐渐降低,其中高级职称、中级职称组间的临床处理方法均无统计学差异(均 $P > 0.05$),而初级职称组间的临床处理方法有统计学差异($P < 0.05$);因 A5 阅片者的内部一致性一般,故在评价 A 单位所有阅片者之间的一致性时未参与,其余 A1~4、A6 间的临床处理方法均有统计学差异(均 $P < 0.05$);ABC 3 个单位同年资阅片者间均为中等,临床处理方法有统计学差异($P < 0.05$),详见表 2。

表 2 不同阅片者间对肺结节临床处理方法的一致性比较

阅片者	Kappa 值	95%CI	Z/ χ^2 值	P 值
A1 与 A2	0.582	0.315 ~ 0.849	-1.890	0.059
A3 与 A4	0.464	0.187 ~ 0.741	-1.069	0.285
A5 与 A6	0.304	0.047 ~ 0.561	-2.023	0.019
A1 ~ 4、A6 间	0.492	0.404 ~ 0.580	11.304	0.023
A3、B1、C1 间	0.500	0.338 ~ 0.662	10.234	0.034

2.3 不同阅片者内部对肺结节分类的一致性比较 8 位阅片者对 49 个肺结节的分类结果中,仅 1 位阅片者 A5 的内部一致性为一般,其余 7 位阅片者的内部一致性为好,详见表 3。

表 3 不同阅片者内部对肺结节分类的一致性比较

阅片者	Kappa 值	95%CI
A1	0.696	0.448 ~ 0.944
A2	0.627	0.374 ~ 0.880
A3	0.804	0.540 ~ 1.068
A4	0.728	0.482 ~ 0.974
A5	0.392	0.303 ~ 0.721
A6	0.8	0.544 ~ 1.056
B1	0.64	0.373 ~ 0.907
C1	0.67	0.417 ~ 0.923

2.4 不同阅片者之间对肺结节分类的一致性比较 A 单位高级、中级职称各组间的结节分类均无统计学差异($P > 0.05$),初级职称者间的结节分类有统计学差异($P < 0.05$);去除阅片者 A5 的其余 5 位阅片者的一致性

为中等,该 5 位阅片者的结节分类一致性有统计学差异($P < 0.05$),而前 4 位(A1~4)阅片者的结节分类无统计学差异($P > 0.05$),这是因为 A1 与 A6、A4 与 A6 间的结节分类均有统计学差异(均 $P < 0.05$);3 个单位间的一致性为中度,其结节分类有统计学差异($P < 0.05$),详见表 4。

表 4 不同阅片者间对肺结节分类的一致性比较。

阅片者	Kappa 值	95%CI	Z/ χ^2 值	P 值
A1 与 A2	0.653	0.395 ~ 0.911	-0.905	0.366
A3 与 A4	0.676	0.404 ~ 0.984	-1.414	0.157
A5 与 A6	0.282	0.107 ~ 0.457	-2.223	0.011
A1~4	0.64	0.532 ~ 0.748	4.049	0.256
A1~4、A6 间	0.539	0.556 ~ 0.722	10.445	0.034
A3、B1、C1 间	0.541	0.389 ~ 0.693	16.000	0.000

2.5 临床处理方法出现分歧结节的比较分析 通过对阅片者在临床处理方法出现分歧结节进行分析发现,同一单位不同年资阅片者间、不同单位的同年资阅片者间出现分歧的结节分类均为部分实性结节患者,不同的是前者对分歧结节的分类一致性较好,后者对分歧结节的分类一致性较差,通过对部分实性结节的大小分析,对于实性大小的测量两者的一致性均较差。在 A 单位相同年资的分歧结节均为部分实性结节,对于实性部分的大小,两位阅片者一致性较差,详见表 5。

3 讨论

随着 LDCT 检查的广泛应用,肺结节的检出率日益增加,但对肺结节的管理在不同诊断医生间存在差异。通过对不同阅片者的结果分析,49 例结节的临床处理的内部一致性不同阅片者间的一致性存在一定差异,所有阅片者的内部一致性为 0.385~0.791,A 单位的阅片者 A1~4、A6 间的一致性为 0.492(95%CI:0.404~0.580),临床处理方法有统计学差异(均 $P < 0.05$);但进一步分析发现,9 年以上工作经验的阅片者(A1~4)间的临床处理方法无统计学差异($P > 0.05$);不同单位间同年资阅片者间有 30.61%(15/49 例)的分歧结节,其一致性属于为中等(Kappa=0.5,95%CI:0.338~0.662),临床处理方法有统计学差异($P < 0.05$)。本研究临床处理由结节的分类和部分实性结节的实性成分的大小两方面决定。通过临床处理方法出现分歧的结节研究发现,不同阅片者间出现分歧的原因也存在差异,同单位不同年资阅片者间、同单位同年资阅片者间主要是部分实性结节的实性部分大小的测量为造成分歧的主要原因(Kappa=0.091,95%CI:-0.088~0.270),而不同单位相同年资间

表 5 临床处理方法出现分歧结节的比较分析

阅片者	分类(个)			Kappa 值	95%CI	实性大小(个)		Kappa 值	95%CI
	pGGN	PSN	SN			≥5mm	<5mm		
A 单位									
A1	4	15	1	0.754	0.677 ~ 0.831	9	3	0.091	-0.088 ~ 0.270
A2	5	15				3	9		
A3	4	16				6	6		
A4	1	19				9	3		
A6	7	13				6	6		
A 单位同年资									
A3		14				5	9	-0.849	-1.294 ~ 1.463
A4		14				9	5		
不同单位间									
A3	2	13	0	0.171	-0.131 ~ 0.473	4	7	0.335	-0.677 ~ 0.007
B1	3	12	0			7	4		
C1	0	15	0			5	6		

注:pGGN:纯磨玻璃结节;PSN:部分实性结节;SN:实性结节

结节的分类、实性部分的测量一致性较差 (Kappa=0.171, 95% CI: -0.131 ~ 0.473; Kappa = -0.335, 95% CI: -0.677 ~ 0.007)。

对于结节的分类,不同阅片者间的一致性存在差异。研究发现,同单位 8 位阅片者在结节分类中一致性为 0.392~0.804(95% CI: 0.540~0.721),其中 1 位阅片者的内部一致性为一般,其余的内部一致性为好;A 单位的所有阅片者的一致性位于中等 (Kappa=0.539, 95% CI: 0.556~0.722),这与 Ridge 等^[5]研究结果大致相同,为中等,这 5 位同单位不同年资阅片者对于结节的分类仍存在差异 ($\chi^2=10.95, P=0.009$),造成差异可能的主要原因:(1)目前对于亚实性结节定义缺乏客观的量化参考^[5],本研究将穿行其内的血管、支气管作为对比,若能观察到这些结构就定义为亚实性,但位于肺野外带肺纹理较中内带稀疏、纤细,所以同样的结节位于不同的肺野时,可能会出现不同分类结果;(2)低年资可能对结节的分类可能不够准确。本组研究中,A 单位 5 位阅片者的分类结果存在统计学差异,其原因是低年资阅片者 A6 与 A1、A4 2 位相对高年资阅片者间存在差异,而 A1~4 4 位 9 年以上工作经验阅片者间无统计学差异;(3)不同工作单位也可能会影响结节的分类。在本研究中,A 单位中同年资的阅片者 A3 与 A4 间的一致性非常好,为 0.676 (95% CI: 0.404~0.984),对结节的分类结果无统计学差异 ($Z=-1.414, P=0.157$),而不同单位同年资阅片者间的一致性仅为中度水平,对结节的分类结果也存在差异 ($\chi^2=16.000, P=0.000$),这可能与不同单位间平时诊断工作中采用的诊断标准存在差异有关;尽管我们统一制定了

诊断标准,但由于诊断标准自身的缺陷及阅片者长时间养成的诊断习惯,最终造成差异。此外,结节周围正常肺组织的模糊度也可能会影响结节的分类^[27]。

对结节临床处理出现分歧的病例分析发现,该类结节主要是部分实性结节,无论是阅片者内部还是不同阅片者间,尤其是同单位同年资阅片者均为部分实性结节,尽管结节的分类相同,但是临床处理方法却不同,这是因为部分实性大小的测量出现差异 (Kappa<0.2),而造成测量差异的主要原因可能是目前对于显示部分实性结节的窗宽窗位没有达成共识,Fleischner 学会建议在纵隔窗对部分实性结节进行分析测量^[9],但是纵隔窗常规的设为 400/40(窗宽/窗位),只有当实性部分的 CT 值超过 -160Hu 时才会显示,而在一项研究中发现浸润性成份的 CT 值大部分介于 -261 ~ -160 Hu^[8]。Schohen 等^[9]与顾亚峰等^[10]的分别应用结节分析软件对部分实性结节进行分析发现,-300Hu 为显示与测量实性部分的最佳阈值。在本研究中部分实性结节的分析测量是在肺窗上进行的,并且阅片者可以根据自己的需要调节窗宽窗位,这可能是造成临床处理方法差异的主要原因。其次部分结节的实性成分形态不规则,会对测量造成了一定地影响。

本研究存在局限性。第一,本研究样本较小且仅对亚实性结节进行了分析,可能会对结果造成偏倚。第二,本研究均采用低剂量扫描,这可能使得影像噪声增加,信噪比降低,最终图像质量下降,对结节的分析造成影响;第三,本研究并没有与病理结果进行对比分析,仅对

(下转第 1090 页)