

## 临床论著

## GATA 脊柱结核分型的可信度与可重复性研究

陈洪涛<sup>1</sup>, 金格勒<sup>1</sup>, 买尔旦<sup>1</sup>, 魏继虎<sup>1</sup>, 盛伟斌<sup>1</sup>, 王 俭<sup>2</sup>

(1 新疆医科大学第一附属医院骨科中心; 2 影像中心 830054 乌鲁木齐市)

**【摘要】目的:**对 GATA 脊柱结核分型的可信度及可重复性进行分析,为临床应用提供参考。**方法:**让 4 名脊柱外科医师以 GATA 脊柱结核分型为标准独立对 100 例脊柱结核患者进行两次分型(间隔 1 个月),通过计算两两医师之间及同一医师前后两次分型对最后分型、各指标判断相同的平均百分比,确定分型一致性,应用 Kappa 值和 Spearman 系数判定其可信度和可重复性,当 Kappa 值 $>0.5$ 时有中度或中度以上的一致性, Spearman 相关系数值 $>0.40$ 时有中度或中度以上正性相关。**结果:**4 名医师对各指标(椎体塌陷、椎旁脓肿、脊柱稳定性、脊柱后凸畸形、矢状面指数、椎间盘破坏)及最后分型判断相同的平均百分比分别为 61.92%、91.83%、63.33%、60.08%、78.25%、82.00%、52.33%;可信度的 Kappa 值分别为 0.430、0.519、0.317、0.380、0.536、0.531、0.343; Spearman 相关系数分别为 0.392、0.526、0.392、0.503、0.569、0.582、0.576。同一医师前后两次对各指标(椎体塌陷、椎旁脓肿、脊柱稳定性、脊柱后凸畸形、矢状面指数、椎间盘破坏)及最后分型判断相同的平均百分比分别为 72.75%、91.00%、85.25%、70.75%、88.75%、81.75%、73.00%;可重复性的 Kappa 值分别为 0.574、0.453、0.627、0.523、0.724、0.584、0.591; Spearman 相关系数分别为 0.748、0.473、0.632、0.660、0.730、0.613、0.762。**结论:**GATA 脊柱结核分型具有较高的可重复性,但可信度较差。

**【关键词】** GATA 脊柱结核分型;可信度;可重复性

doi: 10.3969/j.issn.1004-406X.2011.02.03

中图分类号: R529.2 文献标识码: A 文章编号: 1004-406X(2011)-02-0098-05

Reliability and inter-observer reproducibility of GATA classification for spinal tuberculosis/CHEN Hongtao, JIN Gele, MAI Erdan, et al//Chinese Journal of Spine and Spinal Cord, 2011, 21(2): 98-102

**【Abstract】 Objective:** To determine the reliability and reproducibility of GATA classification for spinal tuberculosis. **Method:** A total of 100 cases with spinal tuberculosis were distributed to 4 attending spine surgeons. All surgeons received files including clinical/radiographic data, and an evaluating sheet in terms of GATA classification system. The results were determined by calculating the average percentage of intra-observer and inter-observer agreement to induce the last classifications and the indicators. Reliability and reproducibility were quantified using Kappa coefficients and Spearman rank-order correlation. Moderate consistency was defined as a Kappa value greater than 0.5, moderate positive correlation was defined as Spearman correlation greater than 0.4. **Result:** Inter-reliability assessed by mean percentage and Kappa coefficient was 61.92%, 0.43 respectively for vertebral collapse, 91.83%, 0.519 respectively for abscess formation, 63.3%, 0.317 respectively for instability, 60.08%, 0.381 respectively for kyphosis, 78.25%, 0.536 respectively for sagittal index, 82%, 0.531 respectively for disc degeneration, 52.33%, 0.343 respectively for classification standard. Respective results using Spearman correlation were 0.392, 0.526, 0.392, 0.503, 0.569, 0.582, and 0.576. Intra-rater mean percentage and Kappa coefficient were 72.75%, 0.574 for vertebral collapse, 91%, 0.453 for abscess formation, 85.25%, 0.627 for instability, 70.75%, 0.523 for kyphosis, 88.75%, 0.724 for sagittal index, 81.75%, 0.584 for disc degeneration, and 73%, 0.591 for classification standard. Respective results using Spearman correlation were 0.748, 0.473, 0.632, 0.66, 0.730, 0.613, and 0.762. **Conclusion:** The GATA classification for spinal tuberculosis has good inter-observer reproducibility and low intra-observer reliability.

**【Key words】** GATA classification for spinal tuberculosis; Reliability; Reproducibility

基金项目: 十一五国家科技支撑计划项目, 课题编号: 2007BAI07A14

第一作者简介: 男(1978-), 硕士研究生, 研究方向: 骨科临床

电话: (0991)4362829 E-mail: cht19781029@163.com (现调入新疆医科大学第六附属医院骨一科 830054 乌鲁木齐)

通讯作者: 金格勒 E-mail: jingl81@yahoo.com.cn

【Author's address】 Department of Orthopaedics, the First Teaching Hospital of Xinjiang Medical University, Urumchi, 830054, China

脊柱结核是常见的骨结核之一, 脊柱结核分型目的是为了统一规范对脊柱结核的诊治及预防, 国内外学者陆续提出过多种脊柱结核分型, 如 Jain 等<sup>[1]</sup>提出的 CT 分型、Mehta 等<sup>[2]</sup>的胸椎结核分型、Kush 等<sup>[3]</sup>提出的脊柱附件结核分型等, 但是这些分型尚存在某些不足, 不能全面、系统地评估脊柱结核, 实用性较差, 不能有效地指导临床治疗。2007 年土耳其学者 Oguz 等<sup>[4]</sup>在研究 76 例脊柱结核的基础上提出了 GATA 脊柱结核分型, 认为与其他分型比较, 此分型更全面, 床实用性明显提高, 其指导临床疗效明显。但是 GATA 脊柱结核分型提出时间尚短, 其临床实用性还有待进一步考证, 本研究主要拟针对 GATA 分型的一致性进行分析, 对其可信度和可重复性进行评价, 探讨影响其分型一致性的原因, 为今后的临床应用提供一定的参考和帮助。

## 1 资料和方法

### 1.1 一般资料

收集 2006 年 1 月~2008 年 1 月在我院住院治疗的脊柱结核患者 100 例, 其中男 55 例、女 45 例, 年龄 2~83 岁, 平均 43 岁。影像资料包括脊柱 X 线、MRI、CT, 其中颈椎结核 3 例、胸椎结核 22 例、腰椎结核 38 例、胸腰段结核 31 例、跳跃性结核 6 例。所有影像学资料均不带与分型有关的任何标记及患者信息。

### 1.2 GATA 脊柱结核分型标准

GATA 脊柱结核分型<sup>[4]</sup>主要依据临床表现、影像学检查(X 线、CT、MRI)两个方面提出分型的七项标准(脓肿形成、椎间盘退变破坏、椎体破坏塌陷、脊柱稳定性、后凸畸形、矢状面指数、神经系统的损害), 把脊柱结核分为四型:

I A 型, 病变局限于椎体, 合并一个椎间盘退变, 无椎体塌陷性破坏, 无脓肿形成, 无神经系统损害, 无后凸畸形, 椎体稳定性好。

I B 型, 该型病变侵犯椎体并有脓肿形成, 一至两个椎间盘受累, 无椎体塌陷性破坏, 无神经系统损害, 无后凸畸形, 椎体稳定。

II 型, 病变椎体塌陷且呈病理性骨折, 一至两个椎间盘破坏、脓肿形成、后凸畸形(角度不大前路矫形可纠正), 脊柱稳定、脊柱矢状面指数 $<20^{\circ}$ , 伴有或不伴有神经损害。

III 型, 病变椎体呈严重破坏, 塌陷严重, 一至两个椎间盘破坏, 脓肿形成, 严重后凸畸形, 椎体不稳定, 矢状面指数 $\geq 20^{\circ}$ (图 1), 伴有或不伴有神经损害。

注: GATA 脊柱结核分型体系依据七项指标评价, 其中有关神经系统的损害, 本研究已由专人整理, 不影响本研究 4 名医师之间对脊柱稳定一致性的判断, 对 GATA 最后的分型也没有实质性的影响。

### 1.3 研究方法

从我院脊柱外科抽取 4 名脊柱外科专业医师(包括主任医师 1 名、副主任医师 1 名、主治医师

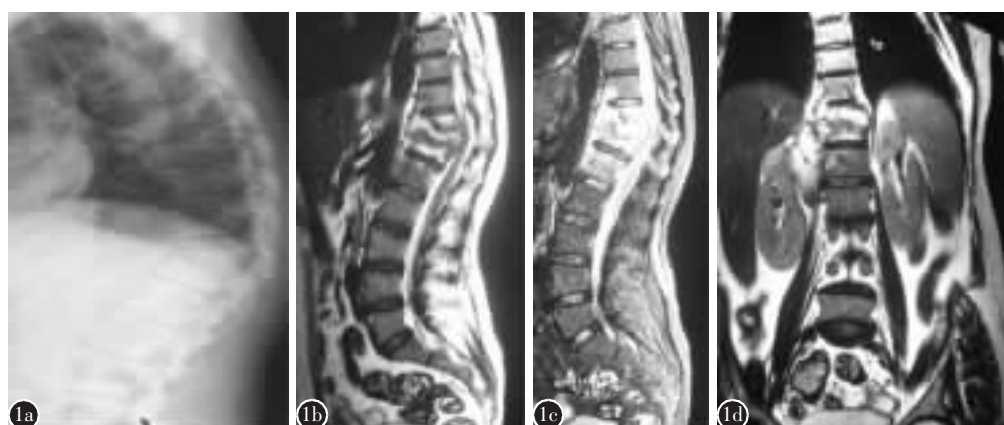


图 1 a~d 术前 X 线片及 MRI 示腰胸段椎体严重破坏塌陷并严重后凸畸形、椎旁脓肿形成、椎间隙破坏、稳定性丧失、病变椎体矢状面指数 $>20^{\circ}$ , 符合 GATA 分型 III 型标准

1 名、住院医师 1 名)进行 GATA 分型培训,向参与分型的医师发送 GATA 分型学习资料,并进行讨论,对分型标准的定义及注意事项进行详细讲解,培训完成后,选取 10 例患者的影像资料由 4 名医师单独分型,考核对分型的掌握程度,考核合格后再由 4 名医师对 100 例患者的影像资料进行单独分型,分型时容许参考原始文献,第一次分型完成后,分型影像资料及分型结果由专人(不参与分型的医师)收集,1 个月后再对相同的 100 例影像资料随机打乱顺序后向参与分型医师发放再次分型,收集分型结果作可信度和可重复性的分析,4 名医师在一次分型中只要有 1 名医师判断不同,即认定为不一致。可信度通过对观察者之间前后两次分型的数据进行两两比较,可重复性通过对观察者个人前后两次分型的数据进行两两比较,均使用 SPSS 17.0 软件计算 Kappa 值和 Spearman 相关系数来判定,探讨影响分型一致性的原因。

1.4 统计方法

数据分析采用一致性检验。使用 SPSS 17.0 软件计算 Kappa 值和 Spearman 相关系数,取均值作为统计量,Kappa 值在-1~+1 之间,大于 0 时有意义,越高表示一致性越好,Kappa 值<0.50 时为一致性较差,0.5~0.75 为有中、高度一致性,Kappa 值>0.75 时,表示有极好的一致性。Spearman 相关系数用来检验等级资料观察者判断的一致性。Spearman 相关系数值在-1~+1 之间,Spearman 相关系数值<0.40 时表示没有正性相关,0.40~0.70 为有中度正性相关,>0.70 为有高度正性相关。

2 结果

2.1 GATA 分型各型的数量与百分比

4 名医师在两次分型中进行了 800 例次(100 例×4 人×2 次)分型,其中 I A 型 50 例(6%)、I B 型 159 例(20%)、II 型 313 例(39%)、III 型 278 例(35%)。

2.2 判断不一致的数量及百分比

通过对前后两次共 200 例病例资料研究后,统计出对 GATA 最后分型及各指标判断不一致例数及百分比,具体见表 1,由表 1 可见 4 名医师最后分型判断不一致占了 75.5%,说明 4 名医师对 GATA 最后分型的判断存在较大的差别,其中影响一致性的主要因素是对脊柱稳定性、后凸畸形、椎体破坏塌陷的判断。

2.3 可信度和可重复性的结果

可信度分析结果见表 2。4 名医师两次分型对最后分型判断相同的百分比平均为 52.33%,可信度的 Kappa 值为 0.343、Spearman 相关系数为 0.576,表明 GATA 分型可信度较低。其中主要影响因素为对椎体塌陷、椎体稳定性、后凸畸形的判断不一致。可重复性研究中椎体塌陷、脊柱稳定性、脊柱后凸畸形、病变椎矢状面指数、椎间盘破坏退变的 Kappa 值和 Spearman 相关系数均达到较高的一致性(表 3)。

表 1 对各指标及最后分型判断不一致的例数及百分比

|        | 判断不一致例数 | 百分比   |
|--------|---------|-------|
| 椎间盘破坏  | 67      | 33.5% |
| 脓肿形成   | 30      | 15.0% |
| SI     | 80      | 40.0% |
| 脊柱稳定性  | 125     | 62.5% |
| 后凸畸形   | 137     | 68.5% |
| 椎体破坏塌陷 | 132     | 66.0% |
| 最后分型   | 151     | 75.5% |

表 2 GATA 分型的可信度分析 (n=100)

| 评价指标      | 分型相同的百分比(%)  | Kappa值 | Spearman相关系数 |
|-----------|--------------|--------|--------------|
| 椎体塌陷      | 61.92(44~73) | 0.430  | 0.392        |
| 椎旁脓肿      | 91.83(87~96) | 0.519  | 0.526        |
| 脊柱稳定性     | 63.33(43~85) | 0.317  | 0.392        |
| 脊柱后凸畸形    | 60.08(50~79) | 0.380  | 0.503        |
| 矢状面指数(SI) | 78.25(66~98) | 0.536  | 0.569        |
| 椎间盘破坏     | 82.00(74~86) | 0.531  | 0.582        |
| 最后分型      | 52.33(34~72) | 0.343  | 0.576        |

表 3 GATA 分型的可重复性分析 (n=100)

| 评价指标      | 分型相同的百分比(%)  | Kappa值 | Spearman相关系数 |
|-----------|--------------|--------|--------------|
| 椎体塌陷      | 72.75(67~85) | 0.574  | 0.748        |
| 椎旁脓肿      | 91.00(89~93) | 0.453  | 0.473        |
| 脊柱稳定性     | 85.25(79~91) | 0.627  | 0.632        |
| 脊柱后凸畸形    | 70.75(63~75) | 0.523  | 0.660        |
| 矢状面指数(SI) | 88.75(86~93) | 0.724  | 0.730        |
| 椎间盘破坏     | 81.75(77~85) | 0.584  | 0.613        |
| 最后分型      | 73.00(68~79) | 0.591  | 0.762        |

3 讨论

3.1 对椎体破坏塌陷判断不一致的分析

本研究对椎体破坏塌陷主要通过脊椎 X 线、CT 来评价,少数病例由于没有 CT 影像资料,通

过结合 X 线及 MRI 来分析,本研究参照多数文献及临床应用将椎体破坏 $<1/2$  定为轻度、椎体破坏 $\geq 1/2$  定为重度破坏。4 名研究者前后两次对椎体破坏塌陷严重程度的判断可信度较差,特别是椎体塌陷介于  $1/3\sim 1/2$  时,判断不一致的较多。对资料分析发现少数胸椎结核的病例,由于缺乏 CT 资料,采用 X 线评价时,受肺、纵隔软组织、肋骨和双肩阴影影响,不易辨别清楚椎体形态,这对上胸椎椎体压缩破坏塌陷程度的判断受到一定影响;另外存在个别研究者对部分病例椎体塌陷严重程度评判时只是主观经验评价而没有对椎体前后缘高度进行测量,结果出现不一致;最后影像资料的清晰度也是影响一致性判断的因素之一。

### 3.2 对脊柱稳定性判断不一致的分析

本研究统一采用 White-panjabi<sup>[5]</sup>量化标准来评价脊柱的稳定性,通过评价前后柱的破坏、椎体间相对矢状面成角移位的大小、椎体间相对矢状面移位、神经损伤等综合评分来判断脊柱稳定性。分析发现出现不一致的原因有对脊柱不稳诊断标准掌握不熟练,如对少数胸腰段结核稳定性的测量,个别医师套用了腰椎稳定性评价标准,胸腰段稳定性评价标准中椎体间的移位为 $>2.5\text{mm}$ ,相对矢状面成角 $>5^\circ$ ,而腰椎的稳定性评价标准中椎体间的移位为 $>4.5\text{mm}$ ,相对矢状面成角 $>22^\circ$ ,导致错误地套用标准值,测出的数值偏差大;还有个别医师测量方法不正确,对少数病变椎的测量只测量了两个相对不稳椎体间的成角,而没有按标准测量包括不稳椎之间、不稳椎与邻近上下正常椎之间的夹角,其结果测出的成角数值大小必然存在偏差;最后影像资料清晰度、测量者的主观经验也是影响脊柱稳定判断的因素。

### 3.3 对脊柱后凸畸形判断不一致的分析

通常脊柱后凸畸形影像学评价指标有后凸 Cobb 角,骶骨倾斜角,重力线前移值<sup>[6,7]</sup>,本研究采用后凸角 Cobb 角来评价。脊柱结核所致后凸畸形,参考 Kongstam 和 Blesovsky<sup>[8]</sup>规定,测定后凸角来表示后凸畸形的程度,测量病变椎体上一椎体上终板与病变椎下一椎体下终板的夹角为 Cobb 角。分析发现出现判断不一致的原因有对病变椎所属节段定位错误,个别医师对胸腰段结核如 T12~L1 椎体结核选择了腰椎标准,结果测出的数据存在误差,另外个别医师对个别病椎的上下端椎定位错误也是影响脊柱后凸畸形一致性判

断的原因;影像资料的清晰度,研究人员的测量技术,凭主观经验判断而没有实际测出 Cobb 角,这些也是影响判断不一致的原因。

### 3.4 对椎间盘破坏、椎旁脓肿、病变椎矢状面指数(SI)判断不一致的分析

对椎间盘退变破坏、椎体脓肿、SI 的判断虽然不是影响最后分型一致性的主要原因,但仍存在对部分病例的判断不一致,研究发现影响对椎间盘破坏及椎旁脓肿的判断主要还是影像资料特别是脊柱 MRI 清晰程度,尤其对于脓肿形成不明显、椎间盘破坏轻时,易出现判断不一致,其中主要影响了 I A、I B 型的判断。对病变椎体矢状面指数(SI)的判断标准本研究主要参照 Farcy 评分方法<sup>[9]</sup>,即测量病变椎上位椎体下终板与病变椎下终板之间的交角,减去基数值为 SI。SI 的基数在胸椎为  $5^\circ$ ,在胸腰段为  $0^\circ$ ,在腰椎为  $-10^\circ$ 。对于影响病变椎体矢状面指数的判断主要还是对病变椎体定量不一致,例如个别研究员对少数椎体破坏轻的脊柱结核判断时,定量病变椎时忽略了破坏较轻的相邻椎体,结果导致测出 SI 数值有差异,另外一个影响 SI 指数的判断是对病变椎相对应的脊柱节段判断错误,如 T12~L1 椎体结核定位为腰椎,直接导致 SI 数值偏大  $10^\circ$ 。

### 3.5 GATA 脊柱结核分型系统自身的缺陷对分型一致性的影响

本研究中发现按 GATA 分型的七项指标评判,有部分病例最后测出的数据按分型标准找不到相适应的分型,前后共 800 人次分型中有 17 例无法确定最后的分型,特别是对 II 型和 III 型的判断。例如存在少数病例椎体破坏严重但无或轻度后凸畸形、矢状面指数 $<20^\circ$ ,此时按 GATA 标准无套用分型,最后无法确定是 II 型和 III 型(图 2),这也是 GATA 分型不足之处。本研究认为对 GATA 脊柱结核分型系统如果对各指标依据严重程度进行评分,依据总分决定最后的分型及治疗,这可能更合理。

跳跃性脊柱结核是指结核病灶侵及不连续的多个椎体,在 GATA 脊柱结核分型系统中作者并没有具体描述,本研究病例中存在 6 例跳跃性脊柱结核,分型中发现有 3 名研究员对跳跃性结核只选择了病变严重椎体分析作为最后分型,这也是影响最后准确分型的原因。印度学者 Sumant Samuel<sup>[10]</sup>认为 GATA 脊柱结核分型针对的是椎体



图 2 a~c 通过测量提示 L1 椎体严重破坏 $>1/2$ 、椎体不稳、病变椎体矢状面指数 $<20^\circ$ 、椎旁脓肿形成、椎间隙破坏、但是无后凸畸形, GATA 分型标准中 II 型没有椎体严重破坏, III 型的判断标准应有  $SI \geq 20^\circ$  且存在重度后凸畸形, 所以此病例最后无法确定是 II 型还是 III 型?

结核的分型治疗, 对脊柱附件结核并没涉及, 脊柱附件结核的外科治疗和椎体结核有很大不同, 这也是 GATA 脊柱结核分型需完善之处。

综上所述 GATA 脊柱结核分型的可信度较差, 可重复性较高, 在临床制定手术策略时应充分考虑到以上这些影响因素, 进一步提高 GATA 分型一致性的判断, 做到: (1) 对分型指标判断标准熟练掌握, 如准确定位病变椎, 准确定量病变椎, 熟练掌握测量方法; (2) 避免主观凭经验判断; (3) 提高测量技术进一步减少误差; (4) 提高影像资料的清晰度; (5) 进一步完善 GATA 分型。对 GATA 脊柱结核分型系统可信度及可重复性的评价及在临床应用的实际效果还需多中心大样本分析结果验证。

#### 4 参考文献

1. Jain R, Sawhney S, Berry M. Computed tomography of vertebral tuberculosis: patterns of bone destruction[J]. Clinical Radiology, 1993, 47(3): 196-199.
2. Mehta JS, Bhojraj SY. Tuberculosis of the thoracic spine: a classification based on the selection of surgical strategies[J]. J Bone Joint Surg Br, 2001, 83(6): 859-863.

3. Kush Kumar. A clinical study and classification of posterior spinal tuberculosis [J]. International Orthopaedics (sicot), 1985 (9): 147-152.
4. Oguz E, Schirlioglu A, Altinmakas M, et al. A new classification and guide for surgical treatment of spinal tuberculosis[J]. International Orthopaedics(sicot), 2008, 32(1): 127-133.
5. White AA, Panjabi MM. Clinical Biomechanics of the Spine[M]. Philadelphia: JB Lippincott, 1990. 342-351.
6. Gelb DE, Lenke LG, Bridwell KH, et al. An analysis of sagittal spinal alignment in 100 asymptomatic middle and older aged volunteers[J]. Spine, 1995, 20(12): 1351-1358.
7. An HS, Balderston RA. Juvenile kyphosis[J]. Spine, 1992, 18(10): 493.
8. Konstam PG, Blesovsky A. The ambulant treatment of spinal tuberculosis[J]. Br J Surg, 1962, 50(219): 26-38.
9. Farcy JP, Weidenbaum M, Glassman SD. Sagittal index in management of thoracolumbar burst fractures[J]. Spine, 1990, 15(9): 958-965.
10. Samuel S. Comment on Oguz, et al: a new classification and guide for surgical treatment of spinal tuberculosis [J]. Int Orthop, 2008, 32(1): 127-133.

(收稿日期: 2010-07-02 修回日期: 2010-12-7)

(英文编审 蒋欣/刘思麒)

(本文编辑 彭向峰)