

肘関節屈筋群における modified Ashworth scale の 検者 5 名による信頼性の検討

中山 恭秀¹ 安保 雅博²

¹東京慈恵会医科大学附属第三病院リハビリテーション科

²東京慈恵会医科大学リハビリテーション医学講座

(受付 平成 19 年 10 月 5 日)

A REVIEW OF THE INTERRATER RELIABILITY OF FIVE EXAMINERS USING THE MODIFIED ASHWORTH SCALE FOR ELBOW FLEXORS

Yasuhide NAKAYAMA¹ and Masahiro ABO²

¹Department of Rehabilitation Medicine, Daisan Hospital, The Jikei University School of Medicine

²Department of Rehabilitation Medicine, The Jikei University School of Medicine

Opinions vary regarding the interrater reliability of the modified Ashworth scale. We thought three factors to difference on a lot of reports about the interrater reliability. Three factors was that to compare by 2 examiners, to use a rank correlation for statistical analysis, and to contain the data of “zero” by 2 examiners. The purpose of this study was to calculate the interrater reliability of the modified Ashworth scale by considering three factors. Twenty-five spastic elbow flexors of 23 patients were evaluated by 5 examiners. Analysis was performed with Spearman rank correlation coefficients and weighted kappa. We obtained high reliability with a weighted kappa is 0.53 to 0.78 and a Spearman rank correlation coefficient of 0.69 to 0.88. This study suggests that the modified Ashworth scale is a high probability examination that is satisfied with the difference of 1 grade error.

(Tokyo Jikeikai Medical Journal 2008; 123: 1-5)

Key words: spasticity, modified Ashworth scale, evaluation, inter-rater reliability, rehabilitation

I. 緒 言

modified Ashworth scale (以下 mAs とする)¹⁾ は、1964 年に多発性硬化症患者の痙縮評価法として Ashworth らにより考案された Ashworth scale²⁾ を、1987 年に Bohannon らが改良したものである。Bohannon らは、片麻痺患者の痙縮評価において Ashworth scale を使用した際に、グレード 1 が離散的となり、信頼を低くしていると指摘し、新たに 1+ というグレードを加えて改定した (Table 1)。現在のところ、mAs は痙縮評価として世界的に最も通用する尺度であり、国内で

も ASIA (American Spinal Injury Association) impairment scale, Brunnstrom recovery stage などと並んで利用頻度の高い機能障害評価法である³⁾。

mAs の表面的妥当性は、Bohannon らにおいて説明されている¹⁾、基準関連妥当性は、Katz ら⁴⁾ により Pendulum Test との相関 ($r = -0.67$) が高いことが示されており、Allison ら⁵⁾ によって、H 反射と中等度の相関関係があることが報告されている ($r = 0.39-0.49$)。しかし、Bakheit ら⁶⁾ は、mAs におけるグレード 1 とグレード 2 の両群で H/M 最大値比の差を比較し、有意差が無かった

Table 1. Classification of the modified Ashworth scale¹³⁾

0	no increase in muscle tone
1	slight increase in muscle tone, manifested by a catch and release or by minimal resistance at the end of the range of motion when the affected part(s) is moved in flexion or extension
1+	slight increase in muscle tone, manifested by a catch, followed by minimal resistance throughout the remainder (less than half) of the ROM
2	more marked increased in muscle tone through most of the ROM, but affected part(s) easily moved
3	considerable increase in muscle tone, passive movement difficult
4	Affected part(s) rigid in flexion or extension

This scale added the grade “1+” to Ashworth scale, and modified the definitions by Bohannon.

ことを報告している ($t = -0.32$, $df22$, $p = 0.7$)。基準関連妥当性で意見がわかる要因として、検者の抵抗強度の加減が影響するとしており、痙縮ではなく緊張の高さを測定する可能性が指摘されている。

一方、信頼性に関しては、検者内信頼性において Gregson ら⁷⁾ が $\kappa = 0.84$ 、Blakburn ら⁸⁾ が $W = 0.62$ と、いずれも高いことを報告している。これに対して、検者間信頼性の報告は意見が分かれている。原典 ($W = 0.85$)、Gregson ら⁷⁾ ($\kappa = 0.83$)、Allison ら⁹⁾ ($rs = 0.73$) など、信頼性が高いと報告しているのに対して、Smith ら¹⁰⁾ ($\kappa = 0.14 - 0.35$)、Blackburn ら⁸⁾ ($W = 0.57$)、Haas ら¹¹⁾ ($\kappa = 0.34$) は必ずしも高くはないとする研究結果を報告している。Sloan ら¹²⁾ は、上肢 ($W = 0.91 - 0.56$) と比べて下肢 ($W = 0.62 - 0.26$) では劣るとしており、部位による精度に関する指摘もみられる。

検者間信頼性の意見がわかれている要因として以下の3つの原因が考えられる。第1に、検者2名で信頼性を検討しているものが多く、検者の個人差が直接反映されている可能性があることである。2名で判断することで、結果が極端に算出される可能性を含んでいる。第2に、順位相関をもって信頼性を検討しているものが多いため、必ずしも完全に一致していなくても高い相関として報告されやすいことである。そして第3に、検者がともにグレード0と判断した症例、つまり痙縮の無い症例も含んで検討されているため、高い信頼性

が得られやすいということである。複数名の検者で一致率をもって検討し、検者がともに0と判断した症例を除いて検者間信頼性を検討することで、より正確な精度を示すことが可能であると考えた。本研究においては、同一の症例を測定した際に、5名全員の検者が0と判断したものを除いている。ただし、4名が0と判断しても1名が1以上と判断した症例などの場合は、痙縮がないとは言えないため対象に含めた。

以上の理由から本研究の目的は、同一被験者に対して5名の検者すべてがグレード0と判断したものを除いた症例を対象として、実験計画法に則って測定を行ない、測定誤差に重みを加味させた一致率による分析で評価法の検者間信頼性を検討することである。なお、本研究は東京慈恵会医科大学倫理委員会の審査を受けて承認を得た後、内容を説明し協力が得られた患者を対象としている。

II. 対象と方法

東京慈恵会医科大学附属第三病院リハビリテーション科病棟に入院および外来通院している患者で、発症から1カ月以上経過しており、検者5名がグレード0(筋緊張増加無し)と判断しなかった23名25肢、平均年齢67.4歳(42~83)を対象とした。測定は肘関節屈筋群とした。検者は、5名の理学療法士(経験年数3~13年、平均7.3年、男性4名、女性1名、すべて右利き)であった。測定前にBohannonらが定める方法の和訳¹³⁾をすべての検者で確認し、判断基準の統一を図った。対象の肢位は背臥位とし、上肢を体側にそらせ前腕回内外中間位をとらせる。検者は一方の手で前腕遠位端を把持し、もう一方の手で上腕を軽く固定する。速度を変えて数回のデモンストレーションを行なった後に、1秒かけて他動による肘関節伸展運動を5回行なうこととした。測定はランダムとし、検者内の測定結果を反映しないように他の測定結果を伏せ、検者間の情報交換を行なわないように配慮した。また、対象には、測定前に十分な休息を取らせた。

分析は、5名の検者間(計10通り)による一致率を、Cohen's weighted kappa(以下kwと略す)を用いて検討した。重み付け係数は、完全一致を

0として、1グレードの誤差が生じた場合を1とした。2グレード以上の誤差は、明らかなる判断の相違ととらえ、2グレードの誤差には3、3グレードの誤差には5、4グレードの誤差には7、5グレードの誤差には9と、2ずつ重み付け係数を付加した (Table 2)。また、多くの論文で用いられている順位相関を Spearman rank correlation coefficient (以下 rs 値とする) にて算出し、kw 値との比較に用いた。これに、完全一致率 (%) と2グレード以上の誤差を生じる確率 (%) を求め、臨床的有用性を検討した。

III. 結 果

全評価値を Table 3 に、また 10 通りの kw 値を Table 4 に示す。10 通りの kw 値は、0.78~0.53、平均 0.68 であった。また、rs 値は Table 5 に示す。0.88~0.69、平均 0.82 となった。rs 値は、kw 値と比べてすべての検者間において高い値を示した ($p < 0.01$)。完全一致率 (Table 6) は、72~32% であり、平均 46% となった。また、2グレード以上の評価結果の違いが生じた確率は、250 のデータ中 10 (発生率 4%) であった。

Table 2. Quadratic weights to kappa statistics

tester 2 tester 1	0	1	1+	2	3	4
0	0	1	3	5	7	9
1	1	0	1	3	5	7
1+	3	1	0	1	3	5
2	5	3	1	0	1	3
3	7	5	3	1	0	1
4	9	7	5	3	1	0

The weight was added to each data from 0 to 9 in accordance with this table.

Table 3. Result of five testers evaluation by modified Ashworth scale ($n=250$)

	0	1	1+	2	3	4
0	17	41	1	5		
1		35	28	4		
1+			16	39		
2				40	16	
3					8	
4						

Ten out of the 250 data caused the error more than two grades.

Table 4. Weighted kappa (kw)

	tester A	tester B	tester C	tester D	tester E
tester A		0.73	0.64	0.77	0.78
tester B			0.53	0.63	0.69
tester C				0.71	0.65
tester D					0.66
tester E					

Overall, reliability of the mAs for assessing elbow flexor spastic patients was found to be adequate to support its continued use.

Table 5. Spearman rank correlation coefficient (rs)

	tester A	tester B	tester C	tester D	tester E
tester A		0.89	0.77	0.88	0.87
tester B			0.69	0.81	0.80
tester C				0.80	0.83
tester D					0.85
tester E					

$p < 0.01$

On the all intertester comparison, spearman rank correlation coefficients were higher than weighted kappa coefficients.

Table 6. Complete agreement rate (%)

	tester A	tester B	tester C	tester D	tester E
tester A		52	40	52	72
tester B			32	36	52
tester C				52	36
tester D					40
tester E					

It was 32% in the combination with the lowest ratio and the highest ratio was 72%. The average is 46%, it is not easy to say it is high.

IV. 考 察

今回、5名の検者が肘関節屈筋群の痙縮を、mAsを用いて判断したrs値、および誤差幅を考慮したkw値はいずれも高かった。また、すべての検者間でkw値と比べてrs値は高く、順位相関では一致率よりも高い分析結果が示されることが確認された。しかしながら、いずれの分析結果をもってしても、肘関節の測定における信頼性は高いこととなる。

5名で測定したことによる興味深い知見として、完全一致率が、高い検者の組み合わせでは72%であったのに対して、低い組み合わせでは32%と明らかな差が見られた。5名の検者間(10通り)における平均が48%であり、5割を下回る結果となっている。これは、2名の検者間で測定されている文献において、検者間信頼性が高いとするものと、逆に低いとするものがある理由の1つであると考えられる。要因としては、グレードの表現が的確であるかという内的妥当性の問題と、検者が理解できていたかという実験計画および習熟の問題とがもたらしている誤差であると考えられる。誤差の要因として高岡¹⁴⁾は、グレードの表現にある1秒間という他動運動の時間でグレードを取り違えずに判断するには、軟部組織の弾性の影響を受けるとしており、可動時間を1秒で正確に行なうことや、感じ取る抵抗量を正確にはかる検者の要素に関する指摘をしている。しかしながら、今回得られた結果より、2グレード以上の格差を生じない高い確率を有しているという特性を加味することで、臨床的有用性を高めることに繋がるであろう。さらに、本研究では5名の検者ともにグレード0と判断した症例を含まずに検討している

ため信憑性を高めているものと考えられる。

痙縮の評価は、mAsをはじめとして、Pendulum Test, H反射, トルク値測定などを用いて行われているが、いずれの評価法でも他の上位運動神経障害や環境・肢位・感情などによる影響を受ける筋緊張との混同や区別がしにくいことが指摘されている⁵⁾⁶⁾。6段階からなるmAsのグレードは、筋緊張の程度を中心に表現されており、それに加えて速度依存性を可動域と、検者の抵抗感の強さで分類されている。臨床的に広く用いられている普及率や簡便性を備えているという利点を高く価値付け、測定精度の向上方法、手続きなどを検討することが臨床場面で重要な課題であろう。辻ら¹⁵⁾は、評価する上での注意点を文章化し、測定前に確認することで信頼性が向上すると報告している。名義尺度で作成されている主観的判断に基づく評価法は、検者間誤差が常につきまとうものであるため、施設における評価者の鍛錬も要求され続ける課題といえる。

V. 結 語

医学的リハビリテーションにおいて痙縮を客観的にとらえることは、徒手理学療法での即時効果や長期効果などの治療効果判定、さらには装具の適合判定などにおいて重要である。今後のリハビリテーション医療のEBMを検証するためにも、簡便なmAsの臨床的有用性の検討は重要であり、検者間信頼性向上は課題であるといえる。Allisonら⁹⁾は、もともとmAsが肘関節屈筋群の測定方法として作成されていることを理由に、下肢筋群などの応用には検討が必要であると述べている。今後我々は、この指摘を踏まえて、下肢筋群におけるmAsの測定精度ならびに測定方法の改訂を念

頭に入れ、同様の手続きによる検者間信頼性の検証を行なう必要があると考える。また、誤差を発生させる要因について、検者の経験年数や被験者の関節可動域制限などによる影響を検討する必要があると考える。

文 献

- 1) Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys Ther* 1987; 67: 206-7.
- 2) Ashworth B. Preliminary trial of carisprodol in multiple sclerosis. *Practitioner* 1964; 192: 540-2.
- 3) 園田 茂, 大橋正洋, 小林一成, 近藤和泉, 豊倉 穰, 森本 茂 ほか. リハビリテーション関連雑誌における評価法使用動向調査. *リハ医学* 2001; 38: 796-8.
- 4) Katz RT, Rovai GP, Brait C, Rymer WZ. Objective quantification of spastic hypertonia, Correlation with clinical findings. *Arch Phys Med Rehabil* 1992; 73: 339-47.
- 5) Allison SC, Abraham LD. Correlation of quantitative measures with the modified Ashworth scale in the assessment of planter flexor spasticity in patients with traumatic brain injury. *J Neurol* 1995; 242: 699-706.
- 6) Bakheit AM, Maynard VA, Curnow J, Hudson N, Kodapala S. The relation between Ashworth scale scores and the excitability of the alpha motor neurons in patients with post-stroke muscle spasticity. *J Neurol Neurosurg Psychiatry* 2003; 74: 646-8.
- 7) Gregson JM, Leathley M, Moore AP, Sharma AK, Smith TL, Watkins CL. Reliability of the Tone Assessment scale and the modified Ashworth scale as clinical tools for assessing psotstroke spasticity. *Arch Phys Med Rehabil* 1999; 80: 1013-6.
- 8) Blackburn M, van Vliet P, Mockett SP. Reliability of measurements obtained with the modified Ashworth scale in the lower extremities of people with stroke. *Phys Ther* 2002; 82: 25-34.
- 9) Allison SC, Abraham LD, Petersen CL. Reliability of the Modified Ashworth Scale in the assessment of plantarflexor muscle spasticity in patients with traumatic brain injury. *Int J Rehabil Res* 1996; 19: 67-78.
- 10) Smith AW, Jamshidi M, Lo SK. Clinical measurement of muscle tone using a velocity-corrected modified Ashworth scale. *Am J Phys Med Rehabil* 2002; 81: 202-6.
- 11) Haas BM, Bergstrom E, Jamous A, Bennie A. The inter rater reliability of the original and of the modified Ashworth scale for the assessment of spasticity in patients with spinal cord injury. *Spinal cord* 1996; 34: 560-4.
- 12) Sloan RL, Sinclair E, Thompson J, Taylor S, Pentland B. Inter-rater reliability of the modified Ashworth scale for spasticity in hemiplegic patients. *Int J Rehabil Res* 1992; 15: 158-61.
- 13) 中山恭秀. modified Ashworth scale. 内山 靖編. *臨床評価指標入門*. 東京: 協同医書出版; 2003. p. 61-6.
- 14) 高岡 徹. 痙縮の評価. *MB Med Reha* 2004; 8-12.
- 15) 辻 哲也, 大田哲生, 木村彰男, 千野直一. 脳血管障害片麻痺患者における痙縮評価: modified Ashworth scale の評価者間信頼性の検討. *リハ医学* 2002; 39: 409-15.