

Original Article

臨床経験年数と典型的な症例の教示が観察による歩行分析の評価者間信頼性に及ぼす影響

谷川広樹,¹ 大塚 圭,¹ 山田純也,² 向野雅彦,³ 松田文浩,¹
加賀谷齊,³ 才藤栄一,³ 金田嘉清,¹ 橋本修二⁴

¹藤田医科大学医療科学部リハビリテーション学科

²藤田医科大学病院リハビリテーション部

³藤田医科大学医学部リハビリテーション医学Ⅰ講座

⁴藤田医科大学医学部衛生学講座

要旨

Tanikawa H, Ohtsuka K, Yamada J, Mukaino M, Matsuda F, Kagaya H, Saitoh E, Kanada Y, Hashimoto S. Influence of clinical experience and instruction on typical cases on the inter-rater reliability of observational gait analysis. *Jpn J Compr Rehabil Sci* 2019; 10: 14-20.

【目的】観察による歩行分析において、臨床経験年数と典型的な症例を観察させることが信頼性に及ぼす影響を明らかにすること。

【方法】30名の理学療法士を教示なし群/あり群の2群に分け、さらに臨床経験年数で全4群に分けた。30名の片麻痺患者の歩行ビデオを観察させ、異常歩行パターンの重症度を5段階に判定させた際の一致率を算出した。教示あり群には判定前に各重症度の典型例を観察させた。

【結果】重症度判定の一致率は教示なし群は低く、あり群の方が高かった。臨床経験年数による違いはなかった。ほぼすべての評価者の組合せに有意な順位相関があった。有意な順位相関があり、かつWilcoxonの符号付順位検定で有意差がなかった組合せ数は教示あり群に多かった。

【考察】臨床経験年数に関わらず観察による歩行分析の信頼性は低く、評価者の主観的尺度の違いが主な原因と考えた。各重症度を示す標準的な症例の観察が信頼性向上に有効である。

キーワード：観察による歩行分析、片麻痺、信頼性、臨床経験、教示

はじめに

リハビリテーション医療において、歩行分析は歩行障害の重症度判定、治療方針の決定、治療の効果判定を行うために用いられる。機器を用いた歩行分析はおもに研究分野で行われ、臨床では低コスト性、簡便性の面から観察による歩行分析が広く用いられている [1-3]。しかし、観察による歩行分析は主観的であり、順序尺度であるという問題がある。先行研究では、観察による歩行分析の評価者間信頼性は低いという報告 [4-11] がある一方、高いという報告 [12-15] もある。これらの結果の違いは、観察の対象者、評価者の数、評価者の職種、評価の段階数、観察部位、観察方法（ライブか、ビデオか）などによると考えられる。さらに、これらの評価者間信頼性を検討した先行研究では、観察対象者の関節角度の異常を判定した際の一致率を算出している。

脳卒中患者の歩行分析において、歩容の特徴を明らかにすることは、異常歩行の病態を理解し、治療への示唆を得ることにつながる [16]。見落としを防ぎ、観察による歩行分析の信頼性を高めるために評価スケールが開発されている [4, 12, 17-19] が、これらの評価スケールはおもに歩行の相ごとに関節角度を評価するものである。なかには同時に異常歩行パターンを評価している項目もあるが、異常歩行パターンの重症度を記録するものではない。脳卒中患者に認める異常歩行パターンの重症度を観察により判定し、その信頼性を検討した研究は見当たらなかった。また、評価者の臨床経験年数が観察による歩行分析の信頼性に及ぼす影響は明らかになっていない [3, 13, 15]。さらにわれわれはこれまで、明確な判断基準を設定することが評価者の主観的な尺度を近似させ、その結果として観察による歩行分析の評価者間信頼性が高まる可能性について言及した [20]。しかし、観察による歩行分析の信頼性を高める具体的な方法論を検証した報告はなく、類似した研究として、トレーニングプログラムによって脳卒中患者の上肢の動きの滑らかさを観察した際の正確性が向上するかを検証した報告 [21] があるのみである。

よって本研究の目的は、理学療法士が脳卒中患者に

著者連絡先：谷川広樹
藤田医科大学医療科学部リハビリテーション学科
〒470-1192 愛知県豊明市杣掛町田楽ヶ窪 1-98
E-mail: tanikawa@fujita-hu.ac.jp
2018年9月7日受付；2018年12月10日受理

利益相反：本研究において、一切の利益相反や研究資金の提供はありません。

おける異常歩行パターンの重症度を観察により判定した際の評価者間信頼性を明らかにすることと、評価者の臨床経験年数と重症度判定前に教示として典型的な症例を観察させることが評価者間信頼性に及ぼす影響を明らかにすることとした。

方法

1. 教示としての典型的な症例の選出

50名の片麻痺患者を対象とした。選定基準は、片麻痺を呈すること、歩行中に分回し歩行か急激な膝関節の伸展が観察されること、介助なしで歩行できることとした。分回し歩行は、「麻痺側下肢が、遊脚初期～遊脚中期に股関節外転・外旋、遊脚中期～遊脚終期に股関節内転・内旋が起こり、半円の軌跡を描くもの」

[22-24]と定義した。急激な膝関節の伸展は、「麻痺側荷重応答期～立脚終期に麻痺側膝関節が急激に後方に向かい伸展するもの」[24]と定義し、つねに膝関節が過伸展位である症例は除外した。50名の片麻痺患者のうち、27名が右麻痺、23名が左麻痺であり、年齢は 57 ± 15 歳、身長 163.4 ± 9.5 cm、体重 59.2 ± 10.5 kg、発症後日数は $1,452 \pm 3,136$ 日(平均値 \pm 標準偏差)であった。下肢のBrunnstrom recovery stageの中央値はIV (I～VI)であった。対象の身体の動きを観察しやすくするため、対象に衣服のだぶつきを抑えるレッグウェアを着用させた。さらに直径30mmのカラマーカを両側の肩峰、腸骨稜、大転子、大腿骨外側上顆、外果、第5中足骨骨頭に貼付し、トレッドミル(ADAL 3D, Tecmachine社製)上を快適平地歩行速度と同速度で歩行させた。手すりや装具の使用は許可した。対象のトレッドミル歩行を、麻痺側と後方からビデオカメラ(HDR-FX1, SONY)で撮影した。

この50名の片麻痺患者のトレッドミル歩行のビデオ画像の中から、4名の臨床経験年数が高い理学療法士(平均臨床経験年数15.8年、8～34年)が二つの異常歩行パターン(分回し歩行と急激な膝関節の伸展)について、5段階の重症度ごとに1例ずつ典型例を選出した。5段階の重症度は、臨床経験に基づいて、異常なし、わずかな異常、軽度異常、中等度異常、重度異常とした。4名の理学療法士には、選出前に各異常歩行の定義を提示した。

2. 観察による歩行分析の評価者間信頼性および臨床経験年数と教示が信頼性に及ぼす影響

2.1 対象者

三つの異なる施設から集められた30名の理学療法士(男性21名、女性9名)を評価者とした。平均臨床経験年数は5.8年(0.4～14.4年)であった。評価者を、臨床経験年数が偏らないように2群に分け、一方を教示あり群、もう一方を教示なし群とした。教示あり群は、観察により異常歩行パターンの重症度判定をする前に各重症度の典型的な症例を観察する群とし、教示なし群は対照群として判定前に典型的な症例は観察しない群とした。さらに、両群内において臨床経験年数5年以上の群と5年未満の群に分けた(図1)。

2.2 実験方法

手順を図2に示す。まず、両群に異常歩行パターンの定義を示した。次に教示あり群には、4名の臨床経験年数が高い理学療法士が選出した症例を、異常歩行パターンの重症度を理解するまで繰り返し観察させた。その後、両群ともに片麻痺患者30例のトレッドミル歩行のビデオ画像を観察させた。30例の片麻痺患者は、前述した50例の片麻痺患者から、4名の理学療法士が選出した5症例を除き、さらに異常歩行パターンが明らかに観察されないかと筆者が判断した

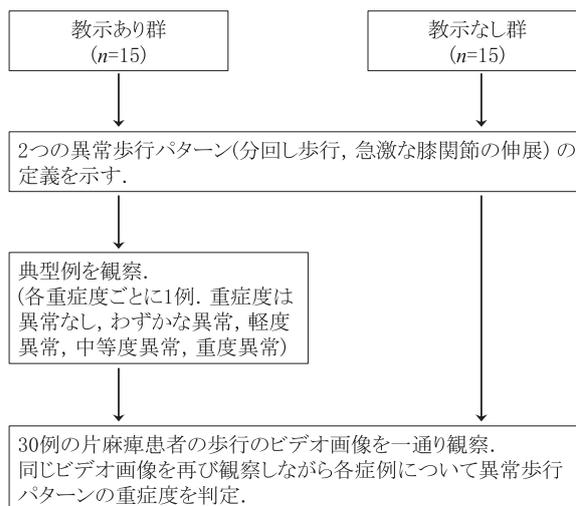


図2. 本研究における実験手順

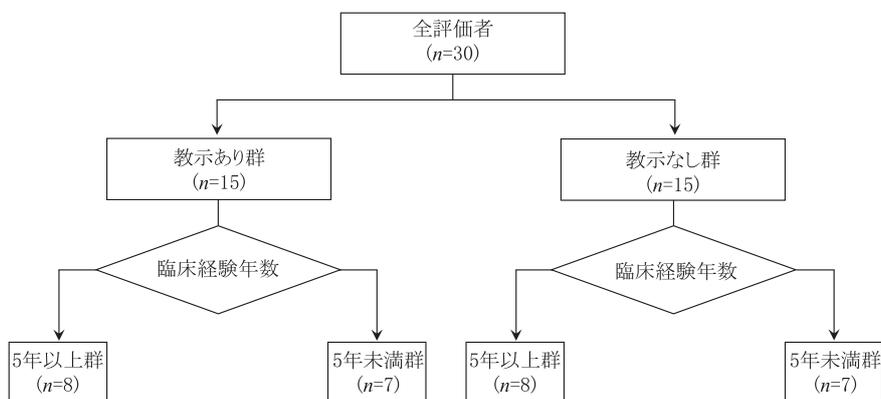


図1. 評価者の群分け

15症例を除いた30症例とした。ビデオ画像はスクリーン（高さ182 cm，幅244 cm）に映写した。ビデオ画像は1症例につき15秒とし、一時停止、スローモーション、他の評価者との議論は許可しなかった。再度、同じビデオ画像を観察させ、同時に重症度を5段階（異常なし、わずかな異常、軽度異常、中等度異常、重度異常）で判定させた。教示あり群と教示なし群の実験日は別日とし、上記手順を異常歩行パターンごとに繰り返した。

2.3 データ分析

判定された重症度は1点（重度異常）から5点（異常なし）に数値化し、各症例における重症度判定結果の中央値を算出した。重症度に関わらず異常のあり/なしの判定の一致率を算出するため、教示あり群と教示なし群それぞれにおいて、各評価者の組み合わせのうち両者とも5点（異常なし）もしくは1~4点（わずかな異常、軽度異常、中等度異常、重度異常のうちいずれか）と判定した組み合わせ数の割合を算出した。また、重症度判定の評価者間信頼性と、臨床経験年数および典型例として選出された症例を事前に観察することが観察による歩行分析の評価者間信頼性に及ぼす影響を検討するため、教示あり群と教示なし群それぞれの臨床経験年数5年以上群と5年未満群内の各評価者の組み合わせにおける判定結果の一致率をCohenの κ 係数、重み付き κ 係数で算出し、さらにSpearmanの順位相関検定、Wilcoxonの符号付順位検定を行った。重み付き κ 係数は、重症度の1段階のズレを半分一致しているとして重み（0.50）をつけた。統計処理にはJMP 7（SAS Institute Inc., Cary, NC, USA）を用い、危険率5%未満を統計学的有意とした。

本研究は藤田医科大学医学研究倫理審査委員会の承認（13-069）を受け、実験参加者には本研究の趣旨と内容を口頭および文章で十分に説明し、書面で同意を得て実施した。

結果

分回し歩行と急激な膝関節の伸展の両パターンにおいて、各症例に対する評価者の重症度判定結果の中央値は、異常なしから重度異常までほぼ同等に分布していた（図3）。教示あり群には重症度判定前に典型的な症例の歩行を繰り返し観察させたが、5回以上繰り返す必要はなかった。

異常のあり/なしの判定の一致率は、分回し歩行は教示なし群が81.2%、教示あり群が83.2%であった。急激な膝関節の伸展は教示なし群が77.6%、教示あり群が84.9%であった。教示なし群と教示あり群のCohenの κ 係数と重み付き κ 係数を表1に示す。両異常歩行パターンにおいて、 κ 係数、重み付き κ 係数ともに教示なし群よりも教示あり群が高かった。

各群内における各評価者の組合せのうち、有意な順位相関があった組合せ数を表2に示す。ほぼすべての組合せで有意な順位相関を認めた。

各群内の各評価者の組合せで実施したSpearmanの順位相関検定とWilcoxonの符号付順位検定の結果を表3に示す。分回し歩行において、有意な順位相関があり、かつWilcoxonの符号付順位検定で有意差がなかった組合せ数は、教示なし群の臨床経験年数5年未満群と教示あり群の臨床経験年数5年未満群で同数であった。一方、教示あり群の臨床経験年数5

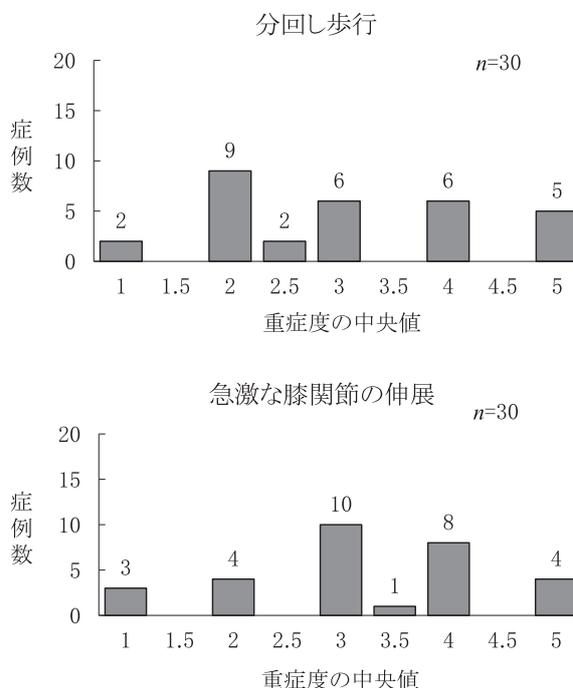


図3. 各症例に対する評価者の重症度判定結果の中央値30名の理学療法士が各症例について判定した重症度の中央値に対するヒストグラムを示す。横軸は重症度の中央値、縦軸は症例数を示す。

表1. 観察による重症度判定結果の一致率

	分回し歩行		急激な膝関節の伸展	
	Kappa (range)	Weighted kappa (range)	Kappa (range)	Weighted kappa (range)
教示あり群				
臨床経験年数5年以上群	0.24 (0.00-0.49)	0.40 (0.00-0.64)	0.23 (0.00-0.46)	0.35 (0.13-0.51)
5年未満群	0.26 (0.07-0.68)	0.43 (0.15-0.78)	0.20 (0.03-0.38)	0.31 (0.06-0.43)
教示なし群				
臨床経験年数5年以上群	0.21 (0.00-0.61)	0.35 (0.04-0.67)	0.18 (0.00-0.49)	0.32 (0.02-0.63)
5年未満群	0.16 (0.00-0.61)	0.26 (0.00-0.71)	0.19 (0.00-0.59)	0.29 (0.04-0.64)

表2. 重症度判定結果に有意な順位相関を認めた評価者の組合せ数

	分回し歩行	急激な膝関節の伸展
教示あり群		
臨床経験年数 5 年以上群	28 (100%)	27 (96%)
5 年未満群	21 (100%)	21 (100%)
教示なし群		
臨床経験年数 5 年以上群	28 (100%)	28 (100%)
5 年未満群	21 (100%)	19 (90%)

数字は組合せ数を示す。カッコ内の数字は評価者の全組合せ数（臨床経験年数 5 年以上群は 28 組，5 年未満群は 21 組）に対する割合を示している。

表3. 各評価者の組合せで実施した Spearman の順位相関検定と Wilcoxon の符号付順位検定結果

Spearman の順位相関検定	Wilcoxon の符号付順位検定	分回し歩行				急激な膝関節の伸展			
		教示なし群		教示あり群		教示なし群		教示あり群	
		臨床経験年数 5 年未満群	臨床経験年数 5 年以上群						
有意な相関あり	有意差なし	7 (33%)	6 (21%)	7 (33%)	15 (54%)	5 (24%)	12 (43%)	9 (43%)	20 (71%)
	有意差あり	14 (67%)	22 (79%)	14 (67%)	14 (46%)	14 (66%)	16 (57%)	12 (57%)	8 (25%)
有意な相関なし	有意差なし	0	0	0	0	1 (5%)	0	0	1 (4%)
	有意差あり	0	0	0	0	1 (5%)	0	0	0

数字は組合せ数を示す。カッコ内の数字は評価者の全組合せ数（臨床経験年数 5 年以上群は 28 組，5 年未満群は 21 組）に対する割合を示している。

年以上群におけるその組合せ数は，教示なし群の臨床経験年数 5 年以上群よりも多かった。急激な膝関節の伸展において，有意な順位相関があり，かつ Wilcoxon の符号付順位検定で有意差がなかった組合せ数は，教示あり群と教示なし群ともに臨床経験年数 5 年以上群が 5 年未満群よりも多かった。さらに，両異常歩行パターンにおいて，教示あり群の臨床経験年数 5 年以上群と教示なし群の 5 年以上群のその組合せ数の差は，教示あり群の 5 年未満群と教示なし群の 5 年未満群の差よりも大きかった。

考察

本研究では，理学療法士が脳卒中患者に特徴的な異常歩行パターンの重症度を 5 段階で判定した際の評価者間信頼性を検討した。脳卒中患者の歩行障害に対するリハビリテーションにおいて，異常歩行パターンの重症度を評価することは治療方針の決定や治療効果判定をするうえで重要である [16, 25]。本研究において，評価者の異常のあり/なしの判定の一致率は高かった。しかし，異常の重症度判定の一致率は低かった。一般的に κ 係数は，0.00~0.20 が slight, 0.21~0.40 が fair, 0.41~0.60 が moderate, 0.61~0.80 が substantial, 0.81~1.00 が almost perfect の一致と解釈できる [26]。本研究において，教示なし群の重症度判定の一致率は臨床経験年数に関わらず低かった (slight~fair)。歩行以外の動作においても，観察に

よる分析の評価者間信頼性が報告されており，Tohara らは繰り返し動作を観察させたにも関わらず評価者間信頼性は低く，その理由は主観的で評価者間のバイアスが生じやすいためと述べている [27]。しかし，本研究では，ほぼすべての評価者の組合せで重症度判定結果に有意な順位相関を認めたことから，重症度の順位は一致していたと考えられた。つまり，評価者が判定した重症度は一致していなかったが，症例を重症度順に並べた際の順序は一致していたと言えた。このことから，観察による重症度判定の評価者間信頼性が低い原因として，評価者の主観的尺度の基点が異なる，主観的尺度における各重症度間の幅が評価者間で同一ではない，各重症度間の幅が評価者内で均一ではないことが考えられた。

本研究ではさらに，異常歩行パターンの重症度判定の前に教示として各重症度を示す典型例を観察させることと，臨床経験年数が観察による歩行分析の評価者間信頼性に及ぼす影響を検討した。臨床判断を行ううえで信頼性の高い歩行分析が重要であるが，臨床現場では観察による歩行分析が主流である [5]。そのため，観察による歩行分析の信頼性を高めるトレーニングプログラムが必要である [13, 20, 21]。よって，重症度判定前の教示と臨床経験年数が評価者間信頼性に及ぼす影響を明らかにすることは，観察による歩行分析の信頼性を向上させる方法論に示唆を与えると考える。

重症度判定結果に有意な順位相関があり，かつ Wilcoxon の符号付順位検定で有意差がなかった評価

者の組合せ数は、教示あり群が教示なし群よりも多かった。このことは、教示が、観察による歩行分析における評価者の主観的尺度を校正したと言えた。また、有意な順位相関があり、かつ Wilcoxon の符号付順位検定で有意差がなかった組合せ数は、経験年数5年以上群の方が5年未満群よりも教示あり群と教示なし群の差が大きかった。このことから、臨床経験年数が長くても観察による歩行分析の評価者間信頼性は低いものの、教示が与えられるという条件下では、臨床経験年数が長いほうが評価者間信頼性が向上しやすいと考えられた。観察による関節角度の異常判定において、臨床経験年数はその判定能力に寄与しないが、異常歩行の原因予測についてはプラスの影響があるという報告 [28] がある。本研究においても、評価者の臨床経験年数が異常歩行パターンの重症度判定に及ぼす影響はわずかであったが、教示を理解し、主観的尺度を校正する点においては良い影響があったと言えた。

本研究において、重症度判定前に典型的な症例を教示として観察させることで臨床経験年数に関わらず評価者間信頼性が向上することがわかった。観察による歩行分析において、ビデオの使用、判定基準の設定、主観的な尺度と客観的な尺度に一貫性を持たせること、トレーニングプログラム、観察後の議論は、評価者間信頼性の改善につながると考えられる [21, 27, 29, 30]。本研究により、観察による歩行分析を臨床現場で応用する前に、基準となる典型例を観察し理解するというトレーニングプログラムが評価者間信頼性の向上に有効だと考えられた。さらに、評価者の主観的尺度を標準化させ、観察による歩行分析の評価者間信頼性を向上させるためには各重症度を表す標準的な症例を設定する必要があると言えた。今回、教示としての各重症度を示す症例の選定は主観的な方法で行われたが、今後は三次元動作分析により健康者との乖離の程度を客観的に表現する方法を用いて選定すべきであろう。

異常歩行パターンの重症度判定に用いられる標準的なスケールは存在しないため、本研究では異常歩行パターンの重症度を5段階に設定した。また、各段階の定義は先行研究 [7, 19, 20] を参考にし、重複がなく異常の程度の順序が明確になるよう留意した。5段階よりも少ない段階数に設定すると一致率は向上すると考えられるが、異常の検出精度が低下し、病態を捉えることができなため臨床応用には不十分である。しかし、7段階以上に設定すると、各段階の意味合いが不明瞭になってしまう [31]。よって、5段階の設定は妥当であったと考えた。さらに、評価者は症例のトレッドミル歩行のビデオ画像を観察した。歩行分析にビデオ画像を用いることの利点は、患者に疲労感を与えることなく評価者が繰り返し観察できる点である [14]。一方、欠点として得られる情報が二次元平面になってしまうこと、解像度が低いと関節角度の判定が難しくなる [6] ことが挙げられる。これらの問題点を解決し、解剖学的骨指標を可能な限り観察しやすくするため、本研究では症例の体表にカラーマーカを貼付した。歩行分析にトレッドミルを用いることの利点は、歩行速度を制御できること、多数歩の分析が容易であること、対象は同一空間内を歩行することが挙げられる [22]。本研究では、トレッドミル歩行

の画像を観察に用いることで症例の歩行速度に関わらず対象をつねに画面の中心に写し、また観察時間を統一することができた。平地歩行とトレッドミル歩行には関節角度変化と時間因子にわずかな違いがある [32, 33] もの、評価者は同一の歩容を観察して重症度判定したため、この点が一致率が低かった主たる原因ではないと考えた。

本研究の限界は、異常歩行パターンの重症度判定における評価者内信頼性を検討していない点である。しかし、観察による歩行分析の評価者間信頼性は低い一方で、評価者内信頼性は高かったという報告がある [3, 4, 8]。よって、本研究において評価者間信頼性が低かった原因は、評価者内の判定のばらつきではなく、評価者間の主観的尺度の相違の影響が大きいと考えた。

結論

観察による歩行分析は簡便であり、臨床での歩行評価に広く用いられている。脳卒中片麻痺患者の歩行分析において、異常歩行パターンを同定し、その重症度を判定することは重要である。本研究では、評価者とした理学療法士は異常歩行パターンの同定および重症度の順位付けは可能であった。しかし、重症度判定の一致率は低く、その原因は評価者の主観的尺度の相違であった。評価者間信頼性は臨床経験年数が長くても低く、臨床経験年数の長さが観察による歩行分析の信頼性に及ぼす影響はわずかであった。しかし、重症度判定前に各重症度の典型例を観察させることで臨床経験年数に関わらず評価者間信頼性は向上した。理学療法士が臨床で観察による歩行分析を行う場合、事前にこのようなトレーニングプログラムを実施すべきであろう。また、三次元動作分析のような客観的な手法による評価と整合性の取れた、異常歩行パターンの重症度を明確に示す標準的な判断基準を設定する必要がある。

文献

1. Robinson JL, Smidt GL. Quantitative gait evaluation in the clinic. *Phys Ther* 1981; 61: 351-3.
2. Stuberger WA, Colerick VL, Blanke DJ, Bruce W. Comparison of a clinical gait analysis method using videography and temporal-distance measures with 16-mm cinematography. *Phys Ther* 1988; 68: 1221-5.
3. Brunnekreef JJ, van Uden CJ, van Moorsel S, Kooloos JG. Reliability of videotaped observational gait analysis in patients with orthopedic impairments. *BMC Musculoskelet Disord* 2005; 17: 1-9.
4. Goodkin R, Diller L. Reliability among physical therapists in diagnosis and treatment of gait deviations in hemiplegics. *Percept Mot Skills* 1973; 37: 727-34.
5. Eastlack ME, Arvidson J, Snyder-Mackler L, Danoff JV, McGarvey CL. Interrater reliability of videotaped observational gait-analysis assessments. *Phys Ther* 1991; 71: 465-72.
6. Keenan AM, Bach TM. Video assessment of rearfoot movements during walking: a reliability study. *Arch Phys Med Rehabil* 1996; 77: 651-5.
7. Maathuis KG, van der Schans CP, van Iperen A, Rietman

- HS, Geertzen JH. Gait in children with cerebral palsy: observer reliability of physician rating scale and Edinburgh visual gait analysis interval testing scale. *J Pediatr Orthop* 2005; 25: 268–72.
8. Krebs DE, Edelstein JE, Fishman S. Reliability of observational kinematic gait analysis. *Phys Ther* 1985; 65: 1027–33.
9. Skaggs DL, Rethlefsen SA, Kay RM, Dennis SW, Reynolds RA, Tolo VT. Variability in gait analysis interpretation. *J Periatr Orthop* 2000; 20: 759–64.
10. Miyazaki S, Kubota T. Quantification of gait abnormalities on the basis of continuous foot-force measurement: correlation between quantitative indices and visual rating. *Med Biol Eng Comput* 1984; 22: 70–6.
11. Kawamura CM, de Moraes Filho MC, Barreto MM, de Paula Asa SK, Juliano Y, Novo NF. Comparison between visual and three-dimensional gait analysis in patients with spastic diplegic cerebral palsy. *Gait Posture* 2007; 25: 18–24.
12. Lord SE, Halligan PW, Wade DT. Visual gait analysis: the development of a clinical assessment and scale. *Clin Rehabil* 1988; 12: 107–19.
13. McGinley JL, Goldie PA, Greenwood KM, Olney SJ. Accuracy and reliability of observational gait analysis data: judgments of push-off in gait after stroke. *Phys Ther* 2003; 83: 146–60.
14. Mackey AH, Lobb GL, Walt SE, Stott NS. Reliability and validity of the observational gait scale in children with spastic diplegia. *Dev Med Child Neurol* 2003; 45: 4–11.
15. Bernhardt J, Bate PJ, Matyas TA. Accuracy of observational kinematic assessment of upper-limb movements. *Phys Ther* 1988; 78: 259–70.
16. Olney SL, Griffin MP, McBride ID. Temporal, kinematic, and kinetic variables related to gait speed in subjects with hemiplegia: a regression approach. *Phys Ther* 1994; 74: 872–85.
17. Zimbelman J, Daly JJ, Roenigk KL, Butler K, Burdsall R, Holcomb JP. Capability of 2 gait measures for detecting response to gait training in stroke survivors: gait assessment and intervention tool and the tinetti gait scale. *Arch Phys Med Rehabil* 2012; 93: 129–36.
18. Hughes KA, Bell F. Visual assessment of hemiplegic gait following stroke: pilot study. *Arch Phys Med Rehabil* 1994; 75: 1100–7.
19. Rodrigues AA, Black PO, Kile KA, Sherman J, Stellberg B, McCormik J, et al. Gait training efficacy using a home-based practice model in chronic hemiplegia. *Arch Phys Med Rehabil* 1996; 77: 801–5.
20. Tanikawa H, Ohtsuka K, Saitoh E, Itoh N, Yamada J, Muraoka Y, et al. Inter-rater reliability of visual inspection about abnormal gait. *Sogo Rehabil* 2010; 38: 1175–81. Japanese.
21. Bernhardt J, Bate PJ, Matyas TA. Training novice clinicians improves observation accuracy of the upper extremity after stroke. *Arch Phys Med Rehabil* 2001; 82: 1611–8.
22. Itoh N, Kagaya H, Saitoh E, Ohtsuka Y, Yamada J, Tanikawa H, et al. Quantitative assessment of circumduction, hip hiling, and forefoot contact gait using Lissajous figures. *Jpn J Compr Rehabil Sci* 2012; 3: 78–84.
23. Kerrigan DC, Frates EP, Rogan S, Riley PO. Hip hiking and circumduction: quantitative definitions. *Am J Phys Med Rehabil* 2000; 79: 247–52.
24. Perry J. *Gait Analysis: Normal and Pathological Function*. Thorofare: SLACK; 1992.
25. Tanikawa H, Ohtsuka K, Mukaino M, Inagaki K, Matsuda F, Teranishi T, et al. Quantitative assessment of retropulsion of the hip, excessive hip external rotation, and excessive lateral shift of the trunk over the unaffected side in hemiplegia using three-dimensional treadmill gait analysis. *Top Stroke Rehabil* 2016; 23: 311–7.
26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–74.
27. Tohara H, Chiba Y, Nakane A, Goto S, Ouchi Y, Teranaka S, et al. Inter- and intra-rater reliability in videofluoroscopic swallowing study. *Jpn J Dysphagia Rehabil* 2005; 9: 139–47. Japanese.
28. Bonkohara S, Yamamoto S. Reliability and accuracy of observational gait analysis. *Rigakuryoho Kagaku* 2008; 23: 747–52. Japanese.
29. McCullough GH, Wertz RT, Rosenbek JC, Mills RH, Webb WG, Ross KB. Inter- and intrajudge reliability for videofluoroscopic swallowing evaluation measures. *Dysphagia* 2001; 16: 110–8.
30. Stoeckli SJ, Huisman TA, Seifert B, Martin-Harris BJ. Interrater reliability of videofluoroscopic swallow evaluation. *Dysphagia* 2003; 18: 53–7.
31. Krosnick JA, Presser S. Number of points on rating scale. In: Marsden PV, Wright JD, editors. *Handbook of Survey Research*. 2nd ed. Bingley: Emerald Group Publishing; 268–75.
32. Alton F, Baldey L, Caplan S, Morrissey MC. A kinematic comparison of overground and treadmill walking. *Clin Biomech* 1998; 13: 434–40.
33. Riley PO, Paolini G, Della Croce U, Paylo KW, Kerrigan DC. A kinematic and kinetic comparison of overground and treadmill walking in healthy subjects. *Gait Posture* 2007; 26: 17–24.