Gillette Gait Index as a Gait Analysis Summary Measure

Comparison With Qualitative Visual Assessments of Overall Gait

Tishya A.L. Wren, PhD,† Khanh Patrick Do, BS,* Reiko Hara, MA, ATC,* Frederick J. Dorey, PhD,‡§ Robert M. Kay, MD,*† and Norman Y. Otsuka, MD||¶*

Background: The Gillette Gait Index (GGI) is a summary measure incorporating 16 clinically important kinematic and temporal parameters. The purpose of this study was to compare GGI scores from computerized gait analysis versus qualitative visual assessments of overall gait to assess the validity of the GGI as a summary score for gait analysis.

Methods: The GGI was calculated for 25 children with cerebral palsy who underwent computerized gait analysis before and 1 year after lower extremity surgery to correct gait problems. Twelve observers reviewed video recordings from the gait analysis to assess the severity of each patient's gait impairment preoperatively and postoperatively and the amount of preoperative to postoperative change. Variability of the video ratings was assessed, and GGI scores were compared with the mean video ratings.

Results: The individual ratings showed some variability, with moderate intrarater agreement (weighted $\kappa = 0.49–0.56$) and slight to fair interrater agreement ($\kappa = 0.11–0.25$). However, the mean scores from all raters were much more consistent, as demonstrated by a highly significant relationship in preoperative to postoperative change viewing the videos separately versus together ($r^2 = 0.62$; $P = 0.0001$). GGI scores were correlated with these mean scores preoperatively ($r^2 = 0.34$; $P = 0.003$), postoperatively ($r^2 = 0.30$; $P = 0.005$), and in preoperative to postoperative change ($r^2 = 0.30$, $P = 0.006$ for absolute change; $r^2 = 0.22$, $P = 0.02$ for percentage change).

Conclusions: These results support the validity of GGI as a gait analysis summary score and suggest that GGI may be a useful outcome measure in patients undergoing gait analysis.

Clinical Relevance: Clinicians and researchers should consider using the GGI as a quantitative outcome measure for assessing overall gait.

Key Words: gait analysis, outcome assessment, outcome measure

(*J Pediatr Orthop* 2007;27:765–768)

From the *Childrens Orthopaedic Center, Childrens Hospital Los Angeles; †Department of Orthopaedics, Keck School of Medicine, University of Southern California; ‡Department of Pediatrics, Childrens Hospital Los Angeles; §University of Southern California; ||Shriners Hospital for Children; and ¶Geffen School of Medicine, University of California, Los Angeles, CA.

This study was supported by the Clinical Outcomes Advisory Board of the Shriners Hospitals for Children and by the Agency for Healthcare Research and Quality (grant no. 5 R01 HS014169).

Reprints: Tishya A.L. Wren, PhD, Childrens Orthopaedic Center, Childrens Hospital Los Angeles, 4650 Sunset Blvd, M.S. 69, Los Angeles, CA 90027. E-mail: twren@chla.usc.edu.

Modern computerized gait analysis produces a large volume of data. Although these data are highly informative for evaluating specific gait parameters such as velocity or peak dorsiflexion in stance, a quantitative measure describing a patient's overall gait is not usually reported. The Gillette Gait Index (GGI), formerly called the Normalcy Index, has been proposed as a summary measure that takes into account 16 clinically important kinematic and temporal parameters.[1] GGI values have been reported to be less than 46 in children without disability and to range from 32 to 1827 in children with cerebral palsy (CP).[1,2] GGI has been shown to be more reliable than any of its constituent variables, with an error of less than 7% compared with 10% to 15% for the most reliable kinematic variables, and an intraclass correlation coefficient of 0.95 in children with CP.[3,4] Previous studies have shown that the GGI can detect differences in groups of subjects who either have different diagnoses or have undergone surgery.[2,5] However, to fully validate this measure for use in individual patients, studies are needed to evaluate the validity of the GGI on an individual level.

In individual patients, the GGI has been shown to correlate with the Edinburgh Gait Score (EGS),[6] the Gillette Functional Assessment Questionnaire walking score,[6] the Gross Motor Function Measure,[7] and the Global Function score of the Pediatric Outcomes Data Collection Instrument.[8] These measures all use specific criteria for assigning scores, requiring them to focus on particular features of gait or function. The purpose of this study was to provide an additional comparison of GGI scores versus qualitative visual assessments of overall gait in individual patients. This will establish whether the GGI reflects observers' overall impressions of gait and will help to validate the GGI for use as a global measure of gait in individual patients.

METHODS

This study examined 25 children (17 boys, 8 girls) with diplegic and quadriplegic CP aged 9.0 ± 3.1 years (range, 4–17 years) who underwent multilevel lower extremity orthopaedic surgery to correct gait problems (Table 1). All research was approved by the appropriate institutional review board. The subjects underwent computerized gait analysis preoperatively and approximately 1 year postoperatively. Both the preoperative and postoperative gait analyses were performed to assess outcomes only; the gait data were not used in treatment planning. For the gait analysis, retroreflective markers were attached over specific bony landmarks of the pelvis and lower extremities.[9,10] Subjects made several passes down a 15-m path, whereas motion of the

TABLE 1. Surgeries Performed

Surgical Procedure	No. Subjects (%)
Hamstring lengthening	18 (72)
Psoas lengthening	10 (40)
Hip adductor lengthening	8 (32)
Tendo-Achilles lengthening	6 (24)
Other	4 (16)

markers was tracked by a 3–dimensional motion analysis system (Vicon Motion Systems, Oxford, UK). The resulting kinematic and temporal data from at least 3 passes were averaged, and the GGI was calculated. The GGI, formerly called the Normalcy Index,[11,12] is a summary measure that takes into account 16 clinically important gait parameters[1] (Table 2). Higher GGI values indicate greater deviation from the gait of able-bodied individuals.

The gait analysis testing also included videotaping of the children walking. Twelve raters reviewed the videos from the same trials used to calculate GGI. The raters included 3 gait laboratory clinicians (physical therapists), 4 other gait laboratory personnel, 3 biomedical engineering students, and 2 persons not affiliated with the gait laboratory. This range of raters was included to represent not only observers highly experienced in gait analysis but also observers not familiar with gait analysis who can represent patients' families or community observers. First, the raters watched the 25 preoperative and 25 postoperative videos mixed together in random order along with 6 repeats (56 videos). The repeats were randomly distributed with an average of 30 other videos between repeat viewings (range, 15–48). Each video was rated in 3 categories: overall walking ability, walking efficiency, and stability and balance. One of the following scores was selected for each category: 1, minimal impairment; 2, moderate impairment; 3, substantial impairment; and 4, severe impairment. By design, the raters were not given specific instructions or guidelines for scoring; this was done to capture the observers' overall impressions without bias toward any particular aspects of gait. The videos for all subjects were viewed in one sitting.

Next, the raters watched each patient's preoperative video followed immediately by the patient's postoperative video. The preoperative to postoperative change was rated as 3, significantly improved; 2, moderately improved; 1, slightly improved; 0, not much better or worse; −1, slightly worse; −2, moderately worse; and −3, much worse. Again, no specific scoring instructions were given to avoid biasing the observers' global assessments.

Before comparing the GGI and observational gait analysis scores, we first examined intrarater and interrater reliability of the raters' scores. Weighted κ statistics were derived from the scores for the 6 repeated videos to evaluate intrarater reliability. κ statistics were also derived from the scores given by the different raters to evaluate interrater reliability of the observational gait ratings. Based on the κ values, agreement was classified as poor (<0.00), slight (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), or almost perfect (0.81–1.00).[13] The

TABLE 2. Parameters Used to Calculate GGI

Time of toe off (% gait cycle)
Walking speed/leg length
Cadence (steps/second)
Mean pelvic tilt (degrees)
Range of pelvic tilt (degrees)
Mean pelvic rotation (degrees)
Minimum hip flexion (degrees)
Range of hip flexion (degrees)
Peak abduction in swing (degrees)
Mean hip rotation in stance (degrees)
Knee flexion at initial contact (degrees)
Time of peak knee flexion in swing (% gait cycle)
Range of knee flexion (degrees)
Peak dorsiflexion in stance (degrees)
Peak dorsiflexion in swing (degrees)
Mean foot progression angle in stance (degrees)

Kruskal-Wallis test was used to compare the raters' assessments of change when viewing the preoperative and postoperative videos separately versus viewing them together. Because only a small number of scores for preoperative to postoperative change were negative, scores of −1 to −3 were considered equivalent in the κ and Kruskal-Wallis analyses. Linear regression was used to compare GGI scores and the mean scores from the 12 raters. Regressions were also performed using the mean scores from only the 3 gait laboratory physical therapists.

RESULTS

Intrarater Reliability

There was moderate agreement between the raters' first and second assessments of the 6 repeated videos (overall κ = 0.52; efficiency κ = 0.49; stability κ = 0.56). For overall

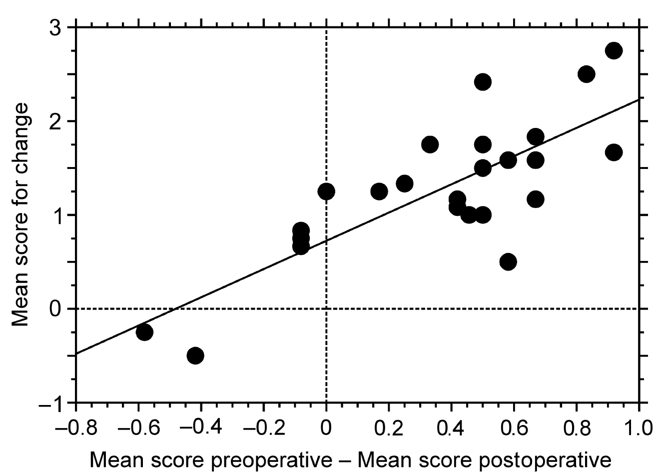

FIGURE 1. There was a highly significant relationship in postoperative change viewing the preoperative and postoperative videos together vs viewing them separately. Results are shown for overall walking ability. Similar results were obtained for walking efficiency and for stability and balance.

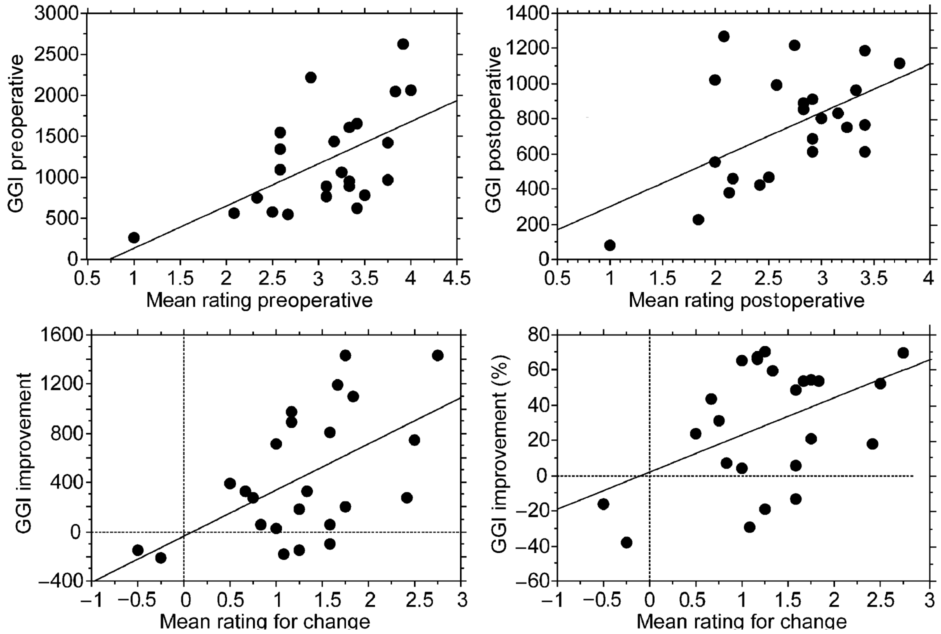

FIGURE 2. The GGI scores reflected the mean visual scores preoperatively, postoperatively, and in preoperative to postoperative change (one outliner excluded).

walking ability, 42 (58%) ratings were exactly the same, 29 (40%) differed by ±1, and 1 (1%) differed by ±2. For walking efficiency, 43 (60%) ratings were exactly the same, 26 (36%) differed by ±1, and 3 (4%) differed by ±2. For stability and balance, 46 (64%) ratings were exactly the same, 21 (29%) differed by ±1, and 5 (7%) differed by ±2. Of the 9 cases in which scores differed by ±2, only 1 involved ratings by a gait laboratory clinician. Thus, differences of more than 1 grade occurred in 4% (9/216) of all assessments, 2% (1/54) for the physical therapists, and 5% (8/162) for the other raters.

Interrater Reliability

Agreement among the 12 raters was fair for assessment of the individual videos (overall $\kappa = 0.25$; efficiency $\kappa = 0.22$; stability $\kappa = 0.23$). Agreement was slightly higher among the gait laboratory clinicians ($\kappa = 0.31–0.33$) compared with the other raters ($\kappa = 0.19–0.22$). Agreement was also better for scores at the ends of the scale ($\kappa = 0.32–0.52$ for scores 1 and 4) than for intermediate scores ($\kappa = 0.09–0.23$ for scores 2 and 3). Despite the low κ values, at least 80% of scores were within 2 adjacent ratings (usually 2–3 or 3–4) for 22 of 25 subjects (88%).

Agreement among the raters was only slight for the assessment of preoperative to postoperative change ($\kappa = 0.11$). Agreement was better for scores at the ends of the scale ($\kappa = 0.22–0.29$ for scores −1 and 3) than for intermediate scores ($\kappa = 0.03–0.15$ for scores 0, 1, and 2). Agreement was highest for scores of −1 ($\kappa = 0.37$) and 3 ($\kappa = 0.53$) among the gait laboratory clinicians. However, because these clinicians had only slight agreement in the intermediate scores ($\kappa < 0.08$), overall agreement was not higher for the gait laboratory clinicians compared with the other raters.

Assessment of Preoperative to Postoperative Change

The raters were more likely to score patients as having improved when the preoperative and postoperative videos were viewed together compared with when the videos were viewed separately. When viewing the videos together, 270 (73%) of the scores given indicated improvement, but only 138 (37%) of the preoperative scores for overall walking ability improved postoperatively, with 192 (52%) remaining unchanged. Nevertheless, using the mean scores from all raters, there was a highly significant relationship between the change in scores when the videos were viewed separately and the score given when the videos were viewed together ($r^2 = 0.62$; $P = 0.0001$) (Fig. 1). The mean ratings were therefore used as the basis for comparison with the GGI scores.

Relationship between GGI and Observational Gait Ratings

The GGI scores reflected the mean video ratings preoperatively ($r^2 = 0.34$; $P = 0.003$), postoperatively ($r^2 = 0.30$; $P = 0.005$), and in preoperative to postoperative change ($r^2 = 0.30$, $P = 0.006$ for absolute change; $r^2 = 0.22$, $P = 0.02$ for percent change) (Fig. 2). Similar results were obtained when only ratings from the gait laboratory physical therapists were used ($r^2 = 0.35$, $P = 0.003$ preoperatively; $r^2 = 0.24$, $P = 0.02$ postoperatively; $r^2 = 0.19$, $P = 0.03$ for absolute change; $r^2 = 0.12$, $P = 0.10$ for percent change). One subject was a clear outlier and was excluded from the analyses. This subject demonstrated substantial improvements in dorsiflexion, which led to a mean score of 1.5 from the raters, but had a worsening in GGI (708–1670) due to deterioration of less visual variables such as range of pelvic tilt and time to peak knee flexion in swing.

DISCUSSION

The intrarater and interrater reliabilities observed in this study were similar to other studies of observational gait analysis using the Physician Rating Scale, Observational Gait Scale, and Edinburgh Visual Gait Scale.[14–17] There were no obvious differences in the ratings assigned by the gait laboratory physical therapists and the other raters. The raters differed in some individual ratings and tended to note more improvement when viewing the preoperative and postoperative videos together. However, using mean scores from the group as a whole, there was good correlation between viewing the videos separately versus together. The mean scores therefore provide a reasonable basis for comparing the GGI scores. The GGI scores were consistent with these mean scores in 24 of 25 patients.

The one subject whose GGI scores did not agree with the visual assessments demonstrates the difference between the visual and quantitative assessments of gait. Although the visual ratings were primarily influenced by the patient's obvious improvement in dorsiflexion, this improvement was outweighed by worsening of several less visual variables in calculation of the GGI. Both assessments are "true"; the difference lies in the relative weight given to different variables and the ability of the quantitative assessment to account for more variables. When the video for this subject was reviewed again after the GGI results were known, the raters were able to see the changes reflected in the GGI. These changes were especially evident when kinematic graphs from the gait analysis were examined. One advantage of the GGI and computerized gait analysis is the ability to account for many variables in an objective manner, which is difficult to do even with structured observational gait analysis. Because of the limited sample size of the current study, it is difficult to determine how frequently the GGI can be expected to differ from qualitative assessments. However, the GGI can be expected to agree with visual assessments of gait in most cases (96% in this study).

The results of this study do not imply that visual assessment of gait can accurately replace objective gait data such as the GGI. Only when the visual ratings were averaged among multiple observers were they considered consistent enough for comparison with the GGI. There was a relatively high level of variability in the individual visual assessment of the children's gait even when the assessments were performed by observers with greater than 10 years of gait laboratory experience. The variability in the visual assessments demonstrates the need for more objective measures such as the GGI.

A recent study by Hillman et al[6] performed detailed visual analysis of specific gait variables by persons experienced in gait analysis to calculate the EGS. This study found high correlations between EGS and GGI. Our study differs by assessing gait globally, rather than in specific gait variables. This may better reflect "real-world" opinions of patient outcome, especially by persons not involved in gait analysis. Along with additional studies relating GGI to functional measures,[7,8] our study and that of Hillman et al suggest that GGI is a valid summary measure for gait analysis in individual patients. Clinicians and researchers should consider using GGI as a global outcome measure in patients undergoing gait analysis.

REFERENCES

1. Schutte LM, Narayanan U, Stout JL, et al. An index for quantifying deviations from normal gait. *Gait Posture*. 2000;11:25–31.
2. Romei M, Galli M, Motta F, et al. Use of the normalcy index for the evaluation of gait pathology. *Gait Posture*. 2004;19:85–90.
3. Schwartz M. The reliability of the normalcy index for quantifying gait pathology. *Gait Posture*. 2003;18:S119.
4. Bothner KE, Fischer R, Alderink G. Assessment of reliability of the normalcy index for children with cerebral palsy. *Gait Posture*. 2003;18:S3–S4.
5. Schwartz MH, Viehweger E, Stout J, et al. Comprehensive treatment of ambulatory children with cerebral palsy: an outcome assessment. *J Pediatr Orthop*. 2004;24:45–53.
6. Hillman SJ, Hazlewood ME, Schwartz M, et al. Correlation of the Edinburgh Gait Score with the Gillette Gait Index, the Gillette Functional Assessment Questionnaire, and dimensionless speed. *J Pediatr Orthop*. 2007;27:7–11.
7. Romei M, Galli M, Schwartz M, et al. Correlation between gait analysis, magnetic resonance imaging and gross motor function measure in cerebral palsy. *Gait Posture*. 2003;18:S1–S30.
8. Tervo RC, Azuma S, Stout J, et al. Correlation between physical functioning and gait measures in children with cerebral palsy. *Dev Med Child Neurol*. 2002;44:185–190.
9. Davis RBI, Ounpuu S, Tyburski D, et al. A gait analysis data collection and reduction technique. *Hum Mov Sci*. 1991;10:575–587.
10. Kadaba MP, Ramakrishnan HK, Wootten ME. Measurement of lower extremity kinematics during level walking. *J Orthop Res*. 1990; 8:383–392.
11. Romei M, Galli M, Motta F, et al. Reply to "Letter to the Editor" regarding Romei et al. *Gait Posture*. 2005;22:377; author reply 378.
12. Theologis T, Thompson N, Harrington M. Letter to the editor regarding Romei et al. *Gait Posture*. 2005;22:377; author reply 378.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
14. Wren TA, Rethlefsen SA, Healy BS, et al. Reliability and validity of visual assessments of gait using a modified physician rating scale for crouch and foot contact. *J Pediatr Orthop*. 2005;25:646–650.
15. Maathuis KG, van der Schans CP, van Iperen A, et al. Gait in children with cerebral palsy: observer reliability of Physician Rating Scale and Edinburgh Visual Gait Analysis Interval Testing scale. *J Pediatr Orthop*. 2005;25:268–272.
16. Mackey AH, Lobb GL, Walt SE, et al. Reliability and validity of the Observational Gait Scale in children with spastic diplegia. *Dev Med Child Neurol*. 2003;45:4–11.
17. Brunnekreef JJ, van Uden CJ, van Moorsel S, et al. Reliability of videotaped observational gait analysis in patients with orthopedic impairments. *BMC Musculoskelet Disord*. 2005;6:17.

