

## Measurement and management of errors in quantitative gait data

Michael H. Schwartz<sup>a,b,c,\*</sup>, Joyce P. Trost<sup>a</sup>, Roy A. Wervey<sup>a</sup>

<sup>a</sup> Gillette Children's Specialty Healthcare, 200 East University Avenue, St. Paul, MN 55101, USA

<sup>b</sup> University of Minnesota, Department of Orthopaedic Surgery, St. Paul, MN 55101, USA

<sup>c</sup> University of Minnesota, Department of Biomedical Engineering, St. Paul, MN 55101, USA

Accepted 8 September 2003

### Abstract

Gait analysis is a valuable tool in the evaluation of children and adults with movement disorders. The data produced from gait analysis, however, is not necessarily free of errors. The purpose of this study was two-fold: (i) to estimate the errors associated with quantitative gait data; and (ii) to propose a method for incorporating the knowledge of these errors into the clinical interpretation process. An experimental protocol was designed that allowed within-subject, within-observer and between-observer errors to be computed at each point in the gait cycle. The estimates were then used in a practical scheme for detecting significant deviations in joint angles. The results of this study provide a means for managing error, while simultaneously improving the rigor and objectivity of clinical interpretations.

© 2003 Elsevier B.V. All rights reserved.

*Keywords:* Quantitative gait data; Movement disorders; Clinical gait analysis

### 1. Introduction

Clinical gait analysis is the standard of practice for evaluating patients with gait abnormalities. The ability of clinicians to discern findings that are meaningful from those that are insignificant or artifactual is therefore essential. In most clinical laboratories, patient gait data is compared to the average response of able-bodied control subjects (control data). This data is often referred to as normal data. The control data provides a reference for the study of pathological gait patterns. Comparison of patient data to control data forms the basis for clinical decisions, as well as the standard by which the success of surgical intervention is measured. Consequently, the question frequently asked by clinicians is: "How many degrees does—have to deviate from the control before it is significant?" If experimental errors conceal important gait deviations, meaningful information will be lost. On the other hand, if the limitations of the method are not understood, small deviations may be considered meaningful, thereby leading to "over-interpretation".

Natural variability exists in the gait of able-bodied persons, and can be attributed to many factors including age,

height, and walking speed [1–4]. Natural variability, however, should not be confused with experimental error. Many specific sources of experimental error are widely known and well characterized. For instance, the effect of thigh coordinate system alignment on hip rotation, knee varus/valgus, and knee flexion/extension angles has been documented [5]. Overall experimental error has also been examined in recent studies. The work of Gorton et al. quantified errors arising within and between observers, laboratory sites, and commercial gait analysis systems [6]. The Gorton study exposed many unexpectedly large sources of experimental error. However, the study was not designed in a manner that allowed precise quantification of intra-subject, -observer and inter-observer errors. The errors in gait data inevitably lead to confusion over the proper diagnosis and course of treatment. Skaggs et al. studied this problem and found only slight to moderate agreement between physicians' interpretation of gait data [7]. It is likely that the differences in interpretation are due, at least in part, to the interpreters' inability to consistently separate those deviations that are statistically significant from those that are possible artifacts.

Two error estimates are germane to gait analysis: an estimate of the standard error of the control data and an estimate of the standard error of an individual patient's data. In clinical gait analysis, it is common to see the standard deviation

\* Corresponding author. Tel.: +1-651-229-3929;

fax: +1-651-229-3867.

E-mail address: [schwa021@tc.umn.edu](mailto:schwa021@tc.umn.edu) (M.H. Schwartz).

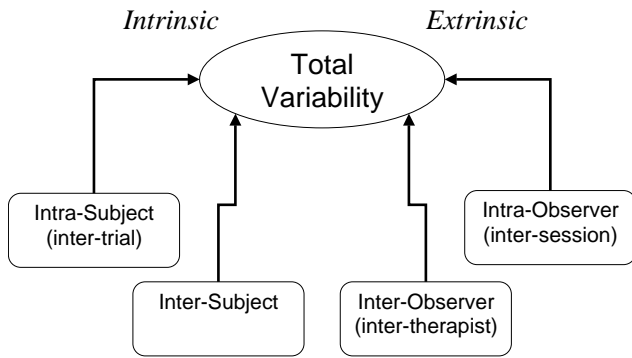


Fig. 1. The different sources of error in gait data. Intrinsic errors are those that arise naturally, either through trial-to-trial or subject-to-subject variability. These errors cannot be reduced, but need to be measured as a baseline for comparison.

of the control data used as the sole measure of experimental error. There are at least three reasons why this approach is wrong. First, the standard deviation of the control data conflates experimental errors with the natural variations that exist in the gait of able-bodied subjects. Second, the gait of a patient is typically compared to the able-bodied average, not to the gait of a randomly selected individual. Thus, the appropriate uncertainty estimate for the control data is the standard error of the mean ( $sem = \sigma / \sqrt{N_{\text{subjects}}}$ ) not the standard deviation ( $\sigma$ ). Third, by ascribing all of the uncertainty to the control data, the interpreter makes the errant assumption that the patient data is untainted by experimental error. Most laboratories have an estimate for the standard error of their control data; the standard error for an individual, however, has not been previously reported.

Variations in measured gait patterns arise from different sources (Fig. 1). Some variations arise from experimental errors, and are candidates for quality improvement measures (extrinsic). Other variations occur naturally, and can only be measured and managed (intrinsic). This study focused on inter-trial variation along with inter-session and -therapist errors. Inter-trial variation measures the intrinsic repeatability of gait patterns, thereby serving as an important reference level to which the extrinsic sources of error can be compared. Inter-session and -therapist errors are extrinsic,

arising from various methodological sources including palpation, anthropometric measurements, alignment processes, regression based joint centers and spatial resolution of the motion capture system to name a few.

## 2. Materials and methods

This study consisted of two parts: (i) estimating the experimental errors inherent in gait analysis; and (ii) incorporating these errors into the clinical interpretation of patient data.

### 2.1. Part I. Estimating experimental errors

The three-dimensional gait data of two healthy adult subjects (subj) was acquired. Subject 1 was a 40-year-old female, 175 cm in height and 75 kg in mass. Subject 2 was a 36-year-old male, 183 cm in height and 91 kg in mass. Each of the subjects underwent 12 gait analysis sessions (sess); three sessions conducted by each of four different staff physical therapists (ther) (Fig. 2). The subjects donned a standard clinical marker set (modified Helen Hayes marker set). Marker trajectories were acquired using a 12-camera optoelectronic system (Vicon 512, etc.). Subjects were asked to walk at a self-selected speed. Five walking trials (trial) were acquired during each session. A validated biomechanical model was then used to compute lower extremity joint kinematics (Vicon Clinical Manager). The inter-trial, -session and -therapist errors were estimated using standard statistical means. For clarity, the statistical procedures are briefly outlined. Let  $\Phi$  denote a gait variable (e.g.  $\Phi = \text{knee flex/extension}$ ) and let the indices  $k, l$  and  $m$  denote therapist, session and trial. Then  $\Phi_{klm}^{\text{subj}}(t)$  is a gait variable for one subject (subj) associated with a single trial ( $m$ ), a single session ( $l$ ), and a single therapist ( $k$ ). The variable  $\Phi_{klm}^{\text{subj}}(t)$  has an explicit dependence on a time-like parameter ( $t$ ). In this study,  $t$  was percent gait cycle, assuming 51 evenly spaced values from 0% to 100%. Except where necessary to eliminate ambiguity, the  $t$  is dropped from the notation and all equations apply at every point in the gait cycle.

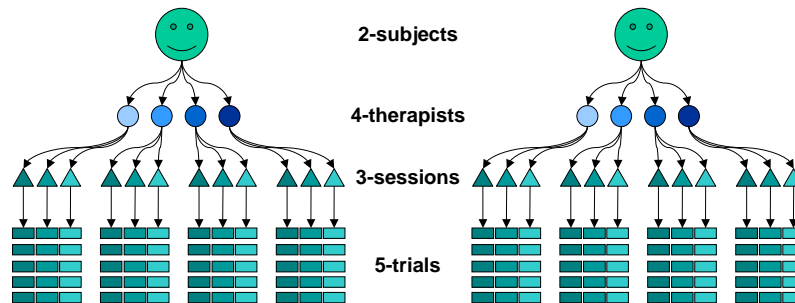


Fig. 2. The repeated-measures experimental design. Two healthy adult subjects (subj) were tested. Each subject underwent three gait analysis sessions (sess) conducted by each of four physical therapists (ther). Five walking trials were obtained during each session.

Session, therapist and subject averages for  $\Phi(t)$  were defined by pooling data accordingly:

$$\begin{aligned} \text{session : } \quad \bar{\Phi}_{kl}^{\text{subj}} &= \frac{1}{N_{\text{trials}}} \sum_{m=1}^{N_{\text{trials}}} \Phi_{klm}^{\text{subj}} \\ \text{therapist : } \quad \bar{\Phi}_k^{\text{subj}} &= \frac{1}{N_{\text{sess}} N_{\text{trials}}} \sum_{l=1}^{N_{\text{sess}}} \sum_{m=1}^{N_{\text{trials}}} \Phi_{klm}^{\text{subj}} \quad . \quad (2.1) \\ \text{subject : } \quad \bar{\Phi}^{\text{subj}} &= \frac{1}{N_{\text{ther}} N_{\text{sess}} N_{\text{trials}}} \sum_{k=1}^{N_{\text{ther}}} \sum_{l=1}^{N_{\text{sess}}} \sum_{m=1}^{N_{\text{trials}}} \Phi_{klm}^{\text{subj}} \end{aligned}$$

Deviations ( $\Delta$ ) of  $\Phi_{klm}^{\text{subj}}$  were then computed for each trial

$$\begin{aligned} \text{inter-trial : } \quad \Delta \Phi_{klm}^{\text{subj,trial}} &= \Phi_{klm}^{\text{subj}} - \bar{\Phi}_{kl}^{\text{subj}} \\ \text{inter-session : } \quad \Delta \Phi_{klm}^{\text{subj,sess}} &= \Phi_{klm}^{\text{subj}} - \bar{\Phi}_k^{\text{subj}} \quad . \quad (2.2) \\ \text{inter-therapist : } \quad \Delta \Phi_{klm}^{\text{subj,ther}} &= \Phi_{klm}^{\text{subj}} - \bar{\Phi}^{\text{subj}} \end{aligned}$$

Inter-trial deviations measure the stride-to-stride (intrinsic) variability of the subject's gait pattern. Inter-session deviations measure the errors introduced when a single therapist (observer) repeats the gait evaluation. Finally, inter-therapist deviations measure the errors introduced when multiple therapists (observers) measure the gait of a subject. The deviations in Eq. (2.2) are relative to subject means, hence inter-subject variability is eliminated and the deviations for the two subjects can be pooled as follows:

$$\begin{aligned} \Delta \Phi_p^{\text{trial}} &= \{ \Delta \Phi_{klm}^{\text{subj}=1,\text{trial}}, \Delta \Phi_{klm}^{\text{subj}=2,\text{trial}} \} \\ \Delta \Phi_p^{\text{sess}} &= \{ \Delta \Phi_{klm}^{\text{subj}=1,\text{sess}}, \Delta \Phi_{klm}^{\text{subj}=2,\text{sess}} \} \quad . \quad (2.3) \\ \Delta \Phi_p^{\text{ther}} &= \{ \Delta \Phi_{klm}^{\text{subj}=1,\text{ther}}, \Delta \Phi_{klm}^{\text{subj}=2,\text{ther}} \} \end{aligned}$$

In Eq. (2.3)  $p$  is a trial counting index ranging from one to the total number of trials ( $N_{\text{total}} = N_{\text{subj}} N_{\text{ther}} N_{\text{sess}} N_{\text{trial}} = 120$ ). The estimated standard error of  $\Phi$  is the standard deviation of  $\Delta \Phi^{\text{source}}$ ,

$$\sigma_{\Phi(t)}^{\text{source}} = \sqrt{\frac{1}{N_{\text{total}} - 1} \sum_{p=1}^N (\Delta \Phi^{\text{source}})^2}, \quad (2.4)$$

where source assumes the values trial, sess or ther, thus measuring the intra-session, -observer and inter-observer sources of error noted above. Note that  $\sigma_{\Phi(t)}^{\text{source}}$  is an explicit function of  $t$ , meaning that it takes on different values for each point. Eqs. (2.1)–(2.4) can be represented graphically for a single subject (Fig. 3).

## 2.2. Part II. Incorporating error estimates into data interpretation

Once  $\sigma_{\Phi(t)}^{\text{source}}$  has been calculated, it can be used to aid in the evaluation of kinematic deviations between patients or conditions. Four obvious comparisons exist:

1. Patient pre-treatment versus average control.
2. Patient post-treatment versus average control.
3. Patient pre-treatment versus patient post-treatment.
4. One patient versus another.

At issue in these comparisons is whether or not an observed deviation reaches a level of statistical significance. Failure to reach statistical significance suggests that the deviation is too small to be reliably measured. Point-by-point assessment of gait deviations can be made using standard statistical techniques as follows:

Let  $\Phi^1(t)$  and  $\Phi^2(t)$  represent the kinematics for two conditions or subjects. In scenario (1) above,  $\Phi^1(t)$  is a patient's pre-treatment data and  $\Phi^2(t)$  is the average laboratory control value for the same data. The point-by-point difference between patient and average control is

$$\delta(t) = \Phi^1(t) - \Phi^2(t) \quad (2.5)$$

The null hypothesis is that the expected value for  $\delta(t)$  is zero. In other words, there is an underlying assumption that the patient and laboratory control data are equal on a point-by-point basis. The test statistic is:

$$z(t) = \frac{\Phi^1(t) - \Phi^2(t)}{\sqrt{\text{S.E.}_{\Phi^1(t)}^2 + \text{S.E.}_{\Phi^2(t)}^2}}, \quad (2.6)$$

where the standard errors (S.E.) are computed from the standard deviations based on the nature of the comparison (i.e. single trials, session averages, etc.). If there is a 95% chance that the null hypothesis is wrong, and the patient's data differs from the control, then the joint angle at time  $t$  is characterized as pathological (probability of Type I error <5%). It must be emphasized that finding a joint angle to be pathological at a single point, or a collection of points, does not imply that the entire curve deviates from the control. Further, it is important to recognize that the probability of errantly characterizing pathological data as typical (Type II error) is neither controlled for nor reported in this study. Instead of reporting Type I errors, it would be equally valid to report Type II errors. These errors could be estimated using the experimental/analytical methodology described here. The choice to report Type II errors would reflect an underlying assumption that a patient's gait data is assumed to differ from the control until proven otherwise. Both views have merit, and neither is without shortcomings. Proper interpretation of statistical error probabilities necessarily calls for an understanding of statistical assumptions and limitations.

## 3. Results

### 3.1. Part I. Estimation of experimental errors

The inter-trial, -session and -therapist errors in lower extremity kinematics were calculated for every point in the gait cycle (Fig. 4). The reliability of a joint angle is measured

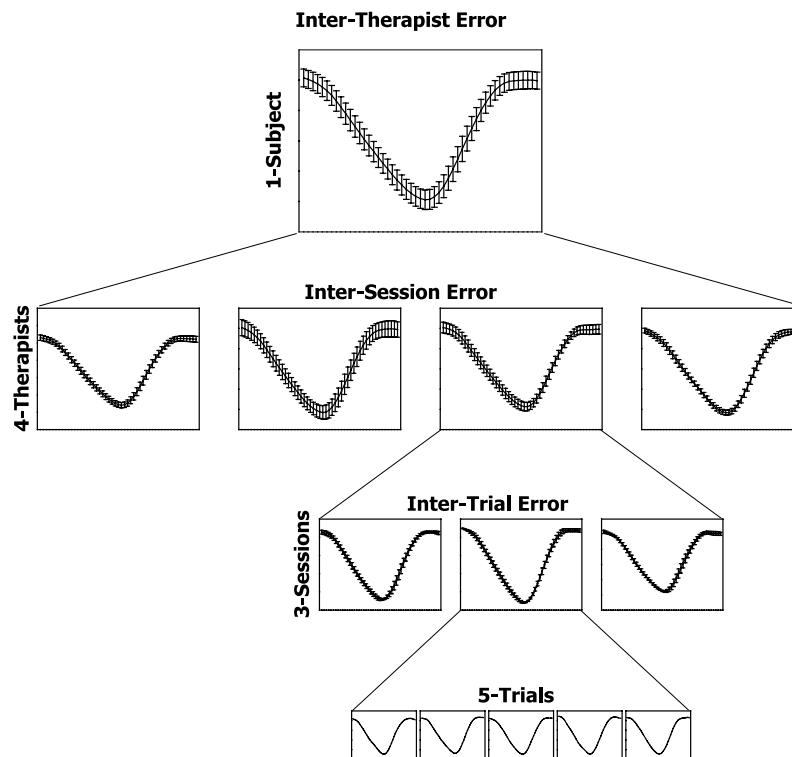


Fig. 3. A schematic depiction of the statistical analysis. Inter-trial differences arise due to the natural (intrinsic) variability that occurs from stride-to-stride. Inter-session errors reflect the ability of a single therapist (observer) to repeatedly acquire gait data. Inter-therapist errors reflect the overall reliability of gait data for the laboratory.

by its standard error. Valuable information is contained in both the magnitude of the inter-therapist error and its ratio to inter-trial error ( $r = \sigma^{\text{ther}}/\sigma^{\text{trial}}$ ). The inter-trial error is free of methodological errors, and thereby serves as an appropriate baseline for comparisons.

Of the 11 joint angles, pelvic obliquity and pelvic rotation were the most reliable (mean  $\sigma^{\text{ther}} = 1.5^\circ$ , mean  $\sigma^{\text{ther}}/\sigma^{\text{trial}} = 1.4$ ). Pelvic rotation depends primarily on the transverse plane orientation of the line connecting the anterior superior iliac spines (ASIS). This direction is essentially free of palpation error since the frontal plane of the pelvis acts as an impenetrable alignment reference. Pelvic obliquity contained smaller inter-therapist errors than pelvic rotation (mean  $\sigma^{\text{ther}} = 1.2^\circ$ ) but a larger ratio of inter-therapist to inter-trial error ( $\sigma^{\text{ther}}/\sigma^{\text{trial}} = 2.6$ ). Pelvic obliquity depends primarily on the vertical alignment of the ASIS markers, which requires manual palpation and alignment, accounting for the increased influence of methodological error. Pelvic tilt contained significantly larger errors than the other pelvic angles (mean  $\sigma^{\text{ther}} = 2.8^\circ$ ,  $\sigma^{\text{ther}}/\sigma^{\text{trial}} = 4.0$ ). This is largely due to the introduction of a third palpated marker at the posterior superior iliac spine.

At the hip, the largest errors were in transverse plane rotation (mean  $\sigma^{\text{ther}} = 4.5^\circ$ ,  $\sigma^{\text{ther}}/\sigma^{\text{trial}} = 4.2$ ). Problems with orienting the knee alignment device (KAD) contribute significantly to the errors in this angle. Hip flex/extension was more reliable than hip rotation (mean  $\sigma^{\text{ther}} = 3.5^\circ$ ,  $\sigma^{\text{ther}}/\sigma^{\text{trial}} = 3.0$ ). Hip ad/abduction was the most reliable

of the hip angles overall (mean  $\sigma^{\text{ther}} = 2.2^\circ$ ). Closer inspection, however, revealed that hip ad/abduction had high ratio of extrinsic-to-intrinsic error ( $\sigma^{\text{ther}}/\sigma^{\text{trial}} = 3.8$ ), suggesting a significant proportion of error arose from methodological sources such as the medial-lateral position of the hip center, as determined by the regression equations.

At this point, it is appropriate to point out that the biomechanical model used in this study is hierarchical. In a hierarchical model, errors propagate “downstream”, from proximal to distal. Errors in pelvic angles, therefore, generate errors in hip, knee, and ankle/foot angles. In addition to hierarchical effects, hip–knee errors are also influenced by the location of the hip and knee joint centers. Marker placement, measured anatomical dimensions and KAD placement determine these joint locations.

Errors in knee angles mainly arise from three sources: upstream effects, joint center locations, and KAD alignment. Misalignment of the KAD generates errors through “cross talk”, the mathematical interaction between planes of finite rotation. The cross talk effect is most prominent in the swing phase knee var/valgus data, when knee flexion is large. During stance phase, when the small knee flexion magnitude minimizes the cross talk effect, the knee var/valgus errors are similar to the hip ad/abduction errors. It is worth noting that knee var/valgus contains the smallest inter-trial errors and the largest inter-therapist to inter-trial ratio (mean  $\sigma^{\text{ther}} = 0.5^\circ$ , mean  $\sigma^{\text{ther}}/\sigma^{\text{trial}} = 5.0$ ). The logical conclusion to be drawn from this is that deviations in knee

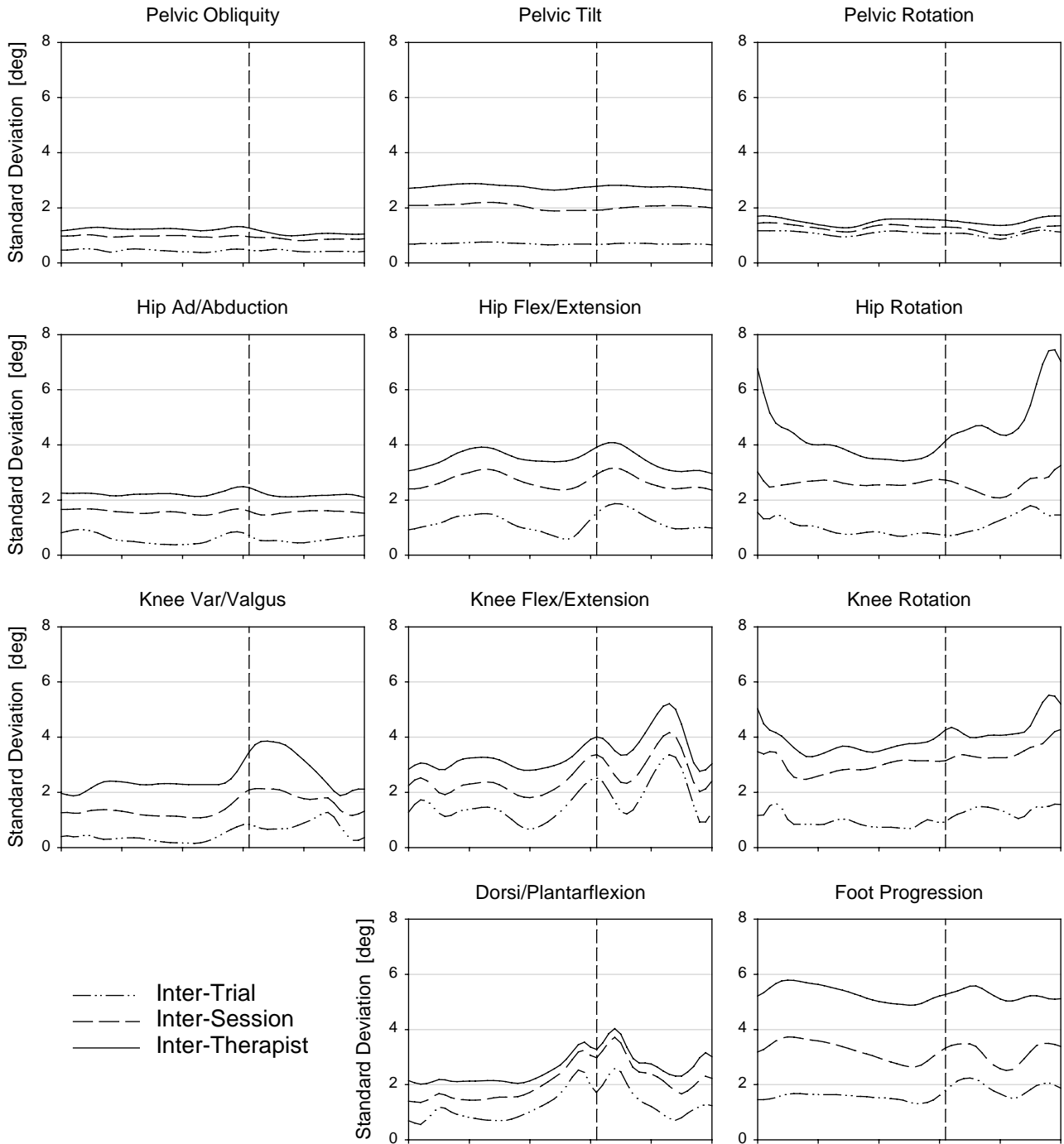


Fig. 4. Errors in lower extremity joint kinematics for the two able-bodied subjects. The toe-off line is the average for the two subjects. The overall reliability of a joint angle is measured by the inter-therapist error. The ratio of inter-therapist to inter-trial error reveals the influence of experimental (extrinsic) errors.

var/valgus data almost certainly arise from experimental errors. This, in turn, supports the increasingly common use of the knee var/valgus data as an ad hoc quality assurance tool. Errors in knee flex/extension data were similar to those seen in hip flex/extension (mean  $\sigma^{ther} = 3.4^\circ$ ). Inter-trial errors in knee flex/extension were relatively large (mean  $\sigma^{trial} = 1.6^\circ$ ), implying that trial-to-trial variability detracts from the reliability of this data. Knee rotation errors es-

entially mirror hip rotation errors owing to their shared genesis.

Foot progression exhibited the greatest inter-therapist error (mean  $\sigma^{ther} = 5.3^\circ$ ) and a large ratio to inter-trial error (mean  $\sigma^{ther}/\sigma^{trial} = 3.2$ ). This finding was particularly disconcerting for two reasons. First, mean foot progression is central to the planning of tibial derotational osteotomies. Second, at the laboratory where the study was conducted

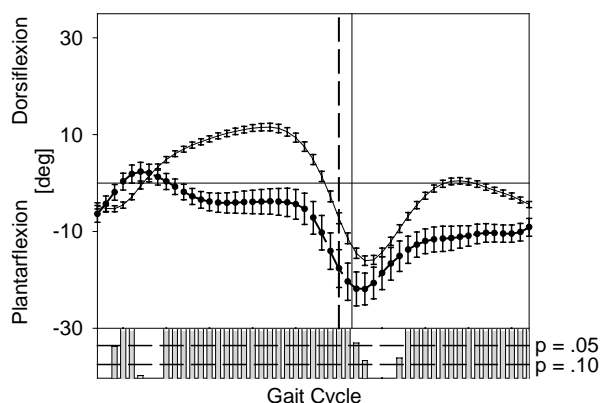


Fig. 5. Point-by-point assessment of significance. A typical dorsi/plantarflexion curve for a patient with a short gastrocnemius is shown. The error bars on the pre-operative patient data (●) are the inter-therapist errors for dorsi/plantarflexion. The error bars on the control data (no symbol) are the standard error of the mean laboratory control. The statistical significance ( $P$ -value) of the deviation is reported on a point-by-point basis along the bottom of the graph. For this example, significant deviations exist for a majority of the gait cycle ( $\approx 80\%$ ).

there was a prevailing impression that foot progression was reliable. Errors in foot progression arise from the placement and alignment of foot markers. Post-hoc examination of the laboratory marker-placement protocols revealed large inter-therapist discrepancies in both the appreciation of proper foot marker placement and the methods used to achieve that placement. Dorsi/plantarflexion exhibited small inter-therapist errors ( $\approx 2^\circ$ ) throughout most of the gait cycle. The error increased sharply near toe-off, where there is a doubling of inter-trial variations, implying that foot position at toe-off is inherently variable.

### 3.2. Part II. Incorporating error estimates into data interpretation

The point-by-point assessment of deviations was demonstrated using a single curve (dorsi/plantarflexion) from a diplegic patient. The point of this example is to demonstrate the utility of the method, not to evaluate the actual patient or the outcome of specific interventions. In the clinical setting, other factors would be important. These factors are beyond the scope of this article.

The following comparisons were made.

- Pre-intervention versus average control (Fig. 5).
- Post-intervention versus average control (Fig. 6).
- Pre-intervention versus post-operative (Fig. 7).

The comparison curve is that of the average laboratory control, not the two subjects used in the determination of the error estimates in Part I. The error estimates depend on the experimental methods used in the laboratory. It was, therefore, deemed proper to ascribe these error estimates to the patient's data.

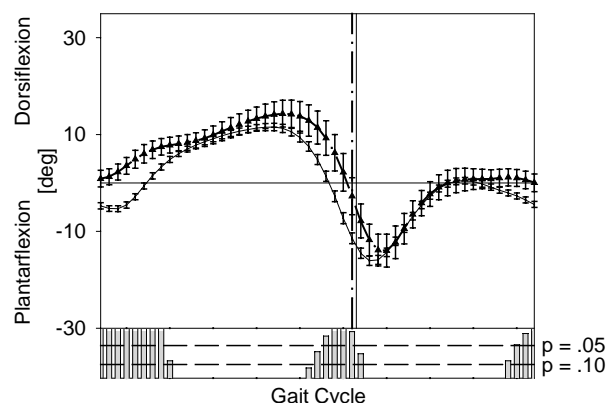


Fig. 6. The dorsi/plantarflexion data following surgical gastrocnemius lengthening (▲) is displayed along with the average control data. Significant deviations from the control data exist near initial contact and toe-off. The percent of the gait cycle in which significant deviations exist has been reduced from 80% to 30%.

Clinical interpretation of gait data involves identifying problems, proposing possible causes, and selecting solution strategies. A point-by-point assessment objectively highlights the portions of the gait cycle in which deviations are more prevalent or significant, thereby highlighting problems and/or the efficacy of treatments directed at those problems. The sample pre-operative ankle dorsi/plantarflexion data showed significant deviations into plantarflexion throughout most (80%) the gait cycle (Fig. 5). While there are many possible causes for this sort of deviation, it is commonly attributed to shortness of the gastrocnemius in the spastic-diplegic population. The post-operative data demonstrated the effect of a surgically lengthened gastrocnemius. Some dorsiflexion deviations existed near initial/terminal contact and at toe-off. However, these deviations comprised only 30% of the gait cycle (Fig. 6). Comparison of the patient to her/himself showed significant pre–post changes in terminal-stance and swing (Fig. 7).

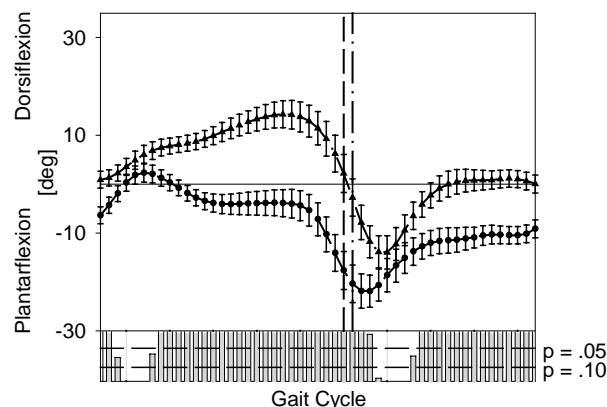


Fig. 7. Comparison of pre-operative (●) and post-operative (▲) data shows that significant changes occurred in terminal stance and swing suggesting an effective intervention.

#### 4. Discussion

The errors in quantitative gait data were evaluated using a repeated measures experimental design and standard statistical analyses. The values presented here are only valid for the laboratory that was studied. A laboratory with a different number of persons involved in the clinical testing, a different set of testing protocols, or a different biomechanical model, would be likely to find somewhat different results. Nevertheless, the errors measured here show trends, and have general magnitudes, that are likely to be consistent across laboratories and commercial gait analysis systems. Furthermore, the methodology presented is widely applicable.

Using gait data to enhance clinical decision-making requires three distinct tasks. The first is to identify significant gait deviations. The second is to decide whether the deviations reflect primary pathology, secondary symptoms, or compensatory mechanisms. The third is to decide if and how the deviations can be treated. Among clinicians, there is considerable disagreement in the surgical decisions made using gait data [7]. This is due in large part to differences in treatment philosophy. However, the confounding effect of methodological errors and the lack of objectivity in assessing the significance of deviations certainly contributes to this inconsistency. Some of the most frequently used variables are also those that exhibit the greatest errors (e.g. hip rotation and foot progression for the planning of femoral and tibial derotational osteotomies). It is clear that the reduction of methodological errors must be a high priority in gait analysis. To reduce errors, however, it is first necessary to identify their source. Measuring inter-trial, -session, and -therapist errors aids in identifying the angles and portions of the gait cycle most susceptible to methodological shortcomings.

While improving accuracy is an important goal, accurate data is of little use if it is injudiciously interpreted. The point-by-point assessment approach gives an unbiased technique for identifying significant deviations during routine interpretation sessions. Clinical significance may then be subjectively judged, aided by the knowledge of statistical significance and inherent methodological errors. Clinical significance is often subjectively defined by a spectrum of factors including: functional changes in a patient's gait, changes in the desired direction of correction and the magnitude of the observed change. There is no objective way to decide on the clinical significance of a statistically significant finding. For example, modest improvements in mid-stance dorsiflexion may have statistical significance but not provide the patient with any tangible functional advantage. On the other hand, the same magnitude change during swing phase may allow the patient to attain foot clearance that was lacking prior to intervention. Evaluating statistical significance is therefore only the first important step necessary to make rational decisions based on quantitative gait data.

One important limitation of the proposed method is the use of point-by-point comparisons. The value of a joint angle

at one time is not independent of values of that angle at other times. Several variables, including walking speed, can influence the magnitude, pattern and timing of joint angle data. It is, therefore, possible that what appears to be a deviation in magnitude may actually constitute a shift, contraction or dilation of the gait cycle. Time normalization reduces, but does not eliminate, this effect. The bootstrap method attempts to account for these effects in a more robust manner [8]. The bootstrap method generally produces larger (more conservative) error estimates than the point-by-point method. Another means for dealing with the interdependence of the data is to use a Bonferonni adjustment; assuming that each point in the gait cycle is influenced by every other point, or by some number of neighboring points [9]. As with the bootstrap method, the Bonferonni adjustment produces a more conservative estimate of significance. These methods differ almost exclusively in the computed degree of significance, not in the regions of the gait cycle identified as most deviant. Therefore, the question of interdependence comes down to the degree of certainty required to designate a deviation as statistically significant. Degree of certainty also influences the choice to report Type I versus Type II errors.

This study is similar to many statistical analyses in that the answer obtained depends partly on the way the question is asked. Regardless of whether or not statistical significance is computed, whether Type I or Type II errors are reported, or whether a  $P < 0.05$  or  $P < 0.01$  level is chosen for significance, the error estimates computed with this methodology reflect the reliability of the calculated gait data. If only for this reason, the method has value for the clinical gait community.

Technical advances will continue to improve the reliability and validity of gait data. The results described here quantify some strengths and weaknesses of the standard clinical gait model. In doing so, they provide a valuable means for directing quality assurance and model improvement efforts. The use of error estimates in clinical gait analysis can improve the rigor and objectivity of clinical interpretation by shifting the focus to gait deviations that exceed the level of experimental uncertainty.

#### References

- [1] Oeffinger DJ, Augsburger S, Cupp T. Pediatric kinetics: age related changes in able-bodied populations. *Gait Posture* 1997;5:155–6.
- [2] Ounpuu S, Gage JR, Davis RB. Three-dimensional lower extremity joint kinetics in normal pediatric gait. *J Pediatr Orthop* 1991;11:341–9.
- [3] Sutherland DH, Olshen RA, Biden EN, Wyatt MP. *The Development of Mature Walking*. Philadelphia: J.B. Lippincott Co.; 1988.
- [4] Stansfield BW, Hillman SJ, Hazlewood E, Lawson AA, Mann AM, Loudon IR, et al. Sagittal joint kinematics, moments, and powers are predominantly characterized by speed of progression, not age, in normal children. *J Pediatr Orthop* 2001;21:403–11.
- [5] Baker R. A new approach to determining the hip rotation profile during clinical gait analysis. *Gait Posture* 1999;9:134.

- [6] Gorton G, Hebert D, Goode B. Assessment of the kinematic variability between twelve shriners motion analysis laboratories. *Gait Posture* 2001;13:247.
- [7] Skaggs DL, Rethlefsen SA, Kay RM, Dennis SW, Reynolds RAK, Tolo VT. Variability in gait analysis interpretation. *J Pediatr Orthop* 2000;20:759–64.
- [8] Lenhoff MW, Santner TJ, Otis JC, Peterson MG, Williams BJ, Backus SI. Bootstrap prediction and confidence bands: a superior statistical method for analysis of gait data. *Gait Posture* 1999;9: 10–7.
- [9] Devore JL. *Probability and Statistics for Engineering and the Sciences*. Pacific Grove (California): Brooks/Cole; 1991.