Quantifying sources of variability in gait analysis

Kohleth Chia [a, b, *], Morgan Sangeux [a, b, c]

[a] Hugh Williamson Gait Analysis Laboratory, Royal Children's Hospital, Melbourne, Australia
[b] Murdoch Childrens Research Institute, Melbourne, Australia
[c] Department of Mechanical Engineering, University of Melbourne, Australia
[*] Corresponding author. Email: kohleth.chia@mcri.edu.au, kohleth@gmail.com. Address: MCRI, 50 Flemington Road, Parkville VIC 3052, Australia.

Abstract

Measurements from gait analysis are affected by many sources of variability. Schwartz *et al.* [1] illustrated an experimental design and methods to estimate these variance components. However, the derivation contains errors which could severely bias the estimation of some components. Therefore, in this paper, we presented correction to this method using ANOVA and Likelihood methods. Furthermore, we demonstrated how commonly used reliability indices like CMC and ICC may be derived from the variance components. We advocate the use of the variance components, in preference to reliability indices, because the variance components are easier to interpret, with understandable units.

Acknowledgement

We thank the two anonymous reviewers for their helpful feedback, which enriched the content of this paper.

DOI: 10.1016/j.gaitpost.2017.04.040 12 May. 17

Introduction

Measurements from gait analysis are variable. The sources of the variability may be intrinsic or extrinsic. Intrinsic variability corresponds to the variability of the subject under investigation, for example the variability between strides of the same individual, or between individuals [2]. Extrinsic variability corresponds to the variability of the gait analysis measurement process, for example marker replacement between sessions, and between assessors, or different marker placement protocols and processing workflows. Having the ability to differentiate and quantify these different sources of variability, or as we call it, variance components, is important for estimating the reliability and repeatability of gait analysis, comparing different methods and protocols, training assessors, and sharing data between laboratories. To this purpose, three statistics are commonly used in the literature, namely, the Coefficient of Multiple Correlation (CMC) as defined in Kadaba *et al.* [3], the Intra-Class Correlation coefficient (ICC) [4], and the explicit quantification of variances following the method proposed by Schwartz *et al.* [1].

CMC has been a popular choice because it is designed to handle curves rather than point data. However, several authors have highlighted issues with CMC, such as its strong dependence on sample size, or range of motion (ROM) [2, 5]. As we will elaborate below, we concur with Røislien *et al.* [5] that CMC in its current form should not be used in these studies. ICC is a similar index but works on the individual time point. However, as we will show below, both these indices may be derived from the variance component estimates themselves. Therefore, in our view, the most appropriate and fundamental framework is that of Schwartz *et al.* [1], which estimate the variance components directly.

The method of Schwartz *et al.* [1] has been adopted in several studies [6–12]. However, their proposed variance component estimators are biased, and as we will demonstrate, this bias may be severe. Our primary objective is to present a corrected set of variance component estimators. In addition, we will present methods to derive the ICC and CMC from the estimated variance components. Finally, we will discuss methods to move beyond point-based calculation for curve data.

Materials & Methods

1. Data

The data we used to illustrate our methods was presented in Schwartz *et al.* [1]. While we did not have access to the original data, figure 4 in Schwartz *et al.* [1] presents the inter-Trial, inter-Session, and inter-Therapist standard deviation for 11 joint angles. We extracted the information from these curves using Engauge Digitizer [13], which exported the coordinates of an irregular set of points on each curve. These points were subsequently fitted with a natural B-spline [14]. The fitted splines were plotted and visually inspected for significant discrepancy with the published curves. Finally, we resampled 101 points from these splines at time points t=0, …, 100 to form our dataset.

2. Experimental Design

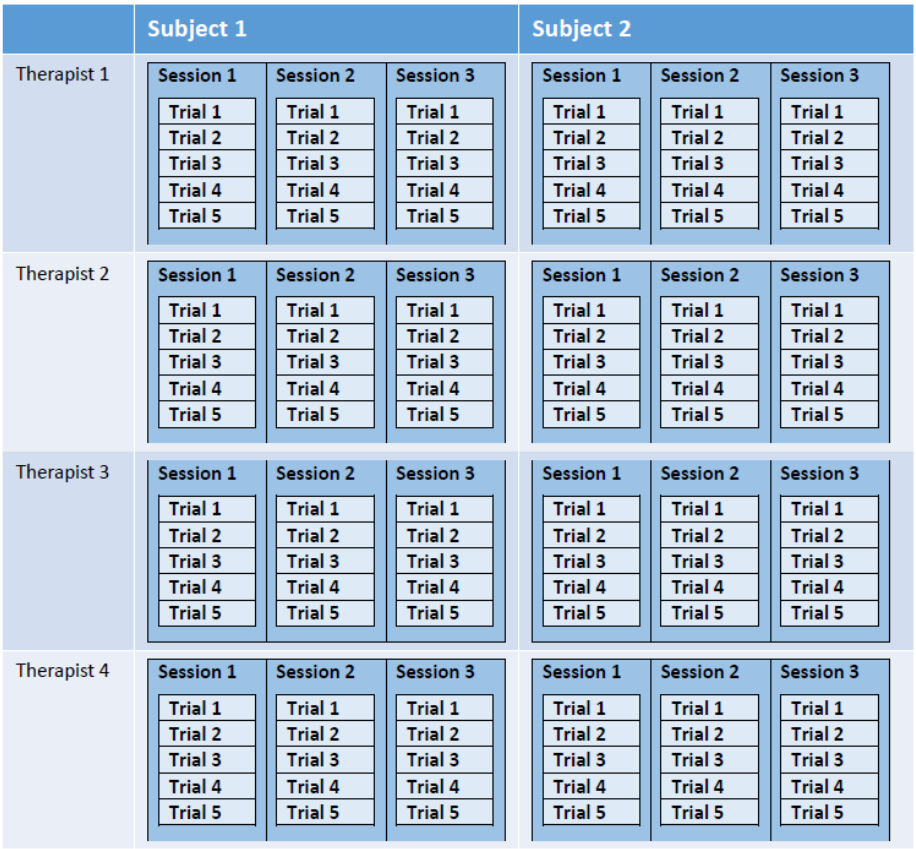
In general, to estimate variance components, one needs to conduct experiments that have multiple realizations of each factor (therapist, session, trial etc.). These factors may have two kinds of relationship with each other, crossed or nested, and it is important to differentiate between them.

Factors A and B are crossed, if all possible combination between them occurs in the experiment, and their effect remains the same regardless of the other variable. For example, in the design of Schwartz *et al.* [1], the factors subject and therapist are crossed, because all subject-therapist combination existed in the experiment, andthe effect of subject i is always the same regardless of which therapist was doing the assessment. In addition, when two factors are crossed, the possibility for interaction between factors arises. In contrast, when factors A and B are nested, there is a natural hierarchical relationship between them, and the effect of the inner factor will depend on the level of the outer factor. For example, the factor trial is nested within the factor session,because the effect of trial l in session k will not be the same as that in session k', where k≠k'.

We need to differentiate between crossed and nested factors in order to account for all possible sources of variability, including that from the interaction effect. Therefore, the accurate description of the experimental design of Schwartz *et al.* [1] would be that 2 subjects are crossed with 4 therapists, and within each subject-therapist combination, there are 3 sessions, and 5 trials nested within each session.

Similarly, when representing the experimental design by figures, for example like in [1,6,7], it may be informative to differentiate graphically between crossed and nested factors. The experimental design of Schwartz *et al.* [1] (figure 2 in the original paper) may be redrawn to highlight the structure of the factors in figure 1.

Figure 1:Graphical representation of the experimental design.



In gait analysis experiments, the degrees of freedom for the innermost factor (i.e. trial) is high, while that for the outermost factors (subject or therapist) is low. For example, in our dataset, there are 4 therapists, thus 3 degrees of freedom for the estimation of the inter-Therapist variability. This is akin

to estimating a variance from 4 numbers, which is inadequate. Therefore, it is often better to increase the number of subjects and therapists (or just therapists if inter-Subject variability is not of interest), while decreasing the number of trials to offset the increase in the resources demanded. Of course, in practice, it is less costly to increase the number of trials than the number of therapists, so a trade-off between cost and precision needs to be made.

3. Variance Estimation

Below, we present the original biased estimators of variance by Schwartz *et al.* [1], the corrected version, as well as the ANOVA method of estimating variance components.

3.1. Schwartz et al. estimators of variance

Let $\Phi_{ijkl}(t)$ denotes the gait measurement (scalar) at time point t, for subject $i = 1,\dots,I$, assessed by therapist $j = 1,\dots,J$, at session $k = 1,\dots,K$ in trial $l = 1,\dots,L$. For notational simplicity we will drop the (t) notation, except in cases where it might cause confusion.

The various means are defined as followed:

- Session mean: $\bar{\Phi}_{ijk} = \frac{1}{L}\sum_l \Phi_{ijkl}$
- Subject-Therapist (Interaction) mean: $\bar{\Phi}_{ij} = \frac{1}{KL}\sum_{kl} \Phi_{ijkl}$
- Therapist mean: $\bar{\Phi}_j = \frac{1}{IKL}\sum_{ikl} \Phi_{ijkl}$
- Subject mean: $\bar{\Phi}_i = \frac{1}{JKL}\sum_{jkl} \Phi_{ijkl}$
- Overall mean: $\bar{\Phi} = \frac{1}{IJKL}\sum_{ijkl} \Phi_{ijkl}$

Schwartz *et al.* [1] defined the estimator for the variance components as followed:

- Inter-trial: $s^2_{trial} = \frac{1}{IJKL-1}\sum_{ijkl}[\Phi_{ijkl} - \bar{\Phi}_{ijk}]^2$
- Inter-session:$s^2_{sess} = \frac{1}{IJKL-1}\sum_{ijkl}[\Phi_{ijkl} - \bar{\Phi}_{ij}]^2$
- Inter-therapist: $s^2_{ther} = \frac{1}{IJKL-1}\sum_{ijkl}[\Phi_{ijkl} - \bar{\Phi}_{i}]^2$

However, these are biased for the target quantity σ^2_{source}, where $source$ can be $trial$, $sess$, or $ther$. Table 1 shows the expected value of these estimators, assuming the linear model as described in section 2.3.3. The quantity $\sigma^2_{interaction}$ denotes the variance components associated with the interaction effect between a subject and a therapist. Unless we have strong reason to believe there is no such effect, it is prudent to include it in the model.

Table 1 Schwartz et al. estimators and their expected value.

Source	Schwartz's estimators	Expected value of Schwartz's estimators
Therapist (target $= \sigma^2_{ther}$)	$\frac{1}{IJKL-1}\sum_{ijkl}[\Phi_{ijkl}-\bar{\Phi}_{ijk}]^2$	$\frac{IJKL-I}{IJKL-1}\sigma^2_{trial}+\frac{IJKL-IL}{IJKL-1}\sigma^2_{sess}+\frac{I(J-1)KL}{IJKL-1}\sigma^2_{cell}+\frac{I(J-1)KL}{IJKL-1}\sigma^2_{ther}$ $=\frac{118}{119}\sigma^2_{trial}+\frac{110}{119}\sigma^2_{sess}+\frac{90}{119}\sigma^2_{interaction}+\frac{90}{119}\sigma^2_{ther}$
Session (target $= \sigma^2_{sess}$)	$\frac{1}{IJKL-1}\sum_{ijkl}[\Phi_{ijkl}-\bar{\Phi}_{ij}]^2$	$\frac{IJK(L-I)}{IJKL-1}\sigma^2_{trial}+\frac{IJ(K-1)L}{IJKL-1}\sigma^2_{sess}$ $=\frac{112}{119}\sigma^2_{trial}+\frac{80}{119}\sigma^2_{sess}$
Trial (target $= \sigma^2_{trial}$)	$\frac{1}{IJKL-1}\sum_{ijkl}[\Phi_{ijkl}-\bar{\Phi}_{i}]^2$	$\frac{IJK(L-1)}{IJKL-1}\sigma^2_{trial}$ $=\frac{96}{119}\sigma^2_{trial}$

There were two sources of bias in the estimators. Firstly, it over-estimated the degrees of freedom associated with the target quantity. The degrees of freedom was always set to $N-1$ $(i.e. IJKL-1)$, whereas the true degrees of freedom was less for all factors. This reduction of degrees of freedom reflects the fact that we do not know the various true means used in the estimators, but have to estimate them from the data. For example, in the estimation of the inter-Trial variability (bottom row in table 1), while we have used $IJKL$ many data, Φ_{ijkl}, in the sums of squares, we also estimated IJK many $\bar{\Phi}_i$, therefore the correct degrees of freedom was $IJKL-IJK=IJK(L-1)$.

Secondly, the expectation of the estimators was a linear combination of the variance components rather than the target variance component itself. For example, the expected value of the Schwartz et al. estimator for the inter-Session variability is not σ^2_{sess}, but a linear combination of that and σ^2_{trial}. The reason for the appearance of the latter variance component is that the mean involved in the estimators, $\bar{\Phi}_{ij}$, is itself estimated by averaging the data at the innermost level, Φ_{ijkl}, and therefore is associated with variability. In summary, these estimators will slightly under-estimate the innermost variance components (inter-Trials), but over-estimate the rest, with increasing severity.

Table 1 provides a means to correct these estimators retrospectively for published literature. Assuming there was no interaction effect (i.e. it was not estimable with the given data), Table 1 provides a system of 3 linear equations to calculate the unbiased estimates of the true variance components: σ^2_{ther}, σ^2_{sess}, σ^2_{trial}.

3.2. ANOVA

The relevant ANOVA method we employed is the random effect ANOVA, where we assume the factors are a random quantity associated with the variance parameter σ^2_{source}. This method is in spirit similar to that of Schwartz et al., except that it uses a different set of equations. A full ANOVA table associated with the experimental design presented in figure 1 is shown in table 2.

Table 2 ANOVA table for the experiment

Source	Sum of Squares (SS)	Degree of freedom (Df)	Expected Mean Square ($EMS = E[SS/Df]$)
Subject	$JKL\sum_i[\bar{\Phi}_i - \bar{\Phi}]^2$	$I-1$	$\sigma^2_{trial} + L\sigma^2_{sess} + KL\sigma^2_{interaction} + JKL\sigma^2_{subj}$
Therapist	$IKL\sum_j[\bar{\Phi}_i - \bar{\Phi}]^2$	$J-1$	$\sigma^2_{trial} + L\sigma^2_{sess} + KL\sigma^2_{interaction} + IKL\sigma^2_{ther}$
Interaction	$KL\sum_{ij}[\bar{\Phi}_{ij} - \bar{\Phi}_i - \bar{\Phi}_j + \bar{\Phi}]^2$	$(I-1)(J-1)$	$\sigma^2_{trial} + L\sigma^2_{sess} + KL\sigma^2_{interaction}$
Session	$L\sum_{ijk}[\bar{\Phi}_{ijk} - \bar{\Phi}_{ij}]^2$	$IJ(K-1)$	$\sigma^2_{trial} + L\sigma^2_{sess}$
Trial	$\sum_{ijkl}[\Phi_{ijkl} - \bar{\Phi}_{ijk}]^2$	$IJK(L-1)$	σ^2_{trial}
Total	$\sum_{ijkl}[\Phi_{ijkl} - \bar{\Phi}]^2$	$IJKL-1$	

If we believe there was no interaction effect, the Interaction and Session strata may be combined, and the combined values are:

$$SS = \sum_{ijk}[\bar{\Phi}_{ijk} - \bar{\Phi}_i - \bar{\Phi}_j + \bar{\Phi}]^2,\ Df = IJK - I - J + 1, \text{ and } EMS = \sigma^2_{trial} + L\sigma^2_{sess}.$$

By equating the mean squares ($MS = SS/Df$) with their expectation (EMS), we obtain a system of linear equations, which solves to the required variance components.

3.3. Maximum Likelihood

The ANOVA estimators are method-of-moments estimators. One drawback of this method is that it requires the dataset to be balanced. Fortunately, both theoretical and computational advances have since led to an alternative way of estimating variance components which does not require a balanced dataset. This method is maximum likelihood estimation.

Implicitly assumed in the ANOVA method is the following linear mixed model:

$$\Phi_{ijkl} = \mu + Sub_i + Ther_j + Interaction_{ij} + Sess_{ijk} + Trial_{ijkl} \qquad (1)$$

where μ denotes the overall mean, and the remaining terms denote the respective random effects, each following an independent normal distribution with zero mean and variance σ^2_{source} for their respective $source$. Instead of computing the sums of squares, maximum likelihood method estimates all unknown parameters directly by maximizing the likelihood function implied by the linear mixed model. Due to this different approach of estimation, the estimated values may be slightly different to

that from the ANOVA method. The actual calculation involved is beyond the scope of this paper, but we point interested readers to Brown & Prescott [15]. The associated standard errors are also available through the matrix of second derivatives, but it should be warned that they provide limited value due to the asymmetry of the estimated variance components. Better alternatives for estimating uncertainties exist but they require more involved computation. We refer interested readers to [15]. In practice, a variant of the maximum likelihood method, called residual maximum likelihood (REML) [16] is preferred due to its ability to correct for the bias in the traditional maximum likelihood method.

3.4. Clinical usability of variance components

The estimated variance components can be used quite flexibly. For example, as we will show in section 2.4, they can be used to derive reliability indices like CMC and ICC. Alternatively, following Schwartz *et al.*[1], we can compute the ratio $\sigma^2_{source}/\sigma^2_{trial}$. This ratio is useful because σ^2_{trial} represents the intrinsic variability that we cannot control, and thus serves as a good benchmark figure.

For comparison with the ratios computed in section 3 of Schwartz *et al.* [1], we have computed the ratio of the extrinsic to intrinsic variation defined as:

$$r = \sqrt{\frac{1}{T}\sum_t \left(\frac{\sigma^2_{therapist}(t) + \sigma^2_{session}(t)}{\sigma^2_{trial}(t)}\right)}$$

The principal advantage of the variance components is that they can be used to derive quantities for situations which have a completely different setup to the ones they were derived from. We give a scenario to illustrate the point below.

Suppose a subject had two gait analysis visits, one use of the variance components is to estimate the variability of the change between the two sessions. Using the estimated variance components, we can calculate that, if the therapist was the same, then the variance of change would be:

$$var\left(\frac{\Phi_{ijkl} - \Phi_{ijk'l'}}{2}\right) = \frac{\sigma^2_{session} + \frac{\sigma^2_{trial}}{L}}{2}$$

whereas if the therapists were different it becomes:

$$var\left(\frac{\Phi_{ijkl} - \Phi_{ij'k'l'}}{2}\right) = \frac{\sigma^2_{therapist} + \sigma^2_{session} + \frac{\sigma^2_{trial}}{L}}{2}$$

So using different therapists increases variability and thus reduces our clinical precision.

Another use is to do sample size calculation. That is, calculate the required L in order to achieve a required threshold of variability, S, by solving the following equation:

$$var\left(\frac{\Phi_{ijkl} - \Phi_{ijk'l'}}{2}\right) = \frac{\sigma^2_{session} + \frac{\sigma^2_{trial}}{L}}{2} < S$$

which leads to

$$L > \frac{2S - \sigma^2_{session}}{\sigma^2_{trial}}$$

Yet another use is to simply inspect which of $\sigma^2_{session}, \sigma^2_{therapist}$, and σ^2_{trial}/L is the largest, and then target our managerial effort to reduce that source of variability.

From this scenario, it is worth noticing that, the setup of repeated visits, with or without the same therapist, is different to the one from which the variance components were estimated. So once we have estimated the variance components, they can be used in many different contexts.

4. Reliability indices

The variance components are the most fundamental quantities and below we illustrate how ICC and CMC can be derived from them.

4.1. Intra-Class Correlation (ICC)

Recall that implicit in our estimation is the linear mixed model (1). The ICC is defined as the correlation between two measurements that have exactly the same components, except for the 'class' variable. For example, the ICC for therapist, is the correlation between two measurements from the same subject, session, and trial, but different therapist. This quantity turns out to be a simple proportion of the relevant variance components (proof in appendix):

$$ICC(Therapist) = Corr\left(\Phi_{ijkl}, \Phi_{ij'kl}\right) = \frac{\sigma^2_{subj} + \sigma^2_{sess} + \sigma^2_{trial}}{\sigma^2_{subj} + \sigma^2_{ther} + \sigma^2_{sess} + \sigma^2_{trial}}$$

Using our dataset, we have computed $ICC(Therapist)$, but since σ^2_{subj} is unknown, we removed this term from both the numerator and denominator of the formula. Therefore, the $ICC(Therapist)$ we computed is $[\sigma^2_{sess} + \sigma^2_{trial}]/[\sigma^2_{ther} + \sigma^2_{sess} + \sigma^2_{trial}]$.

The above quantity is useful for, for example, the training of a team of therapists, where the objective is to have an ICC as close to 1 as possible. In practice, measurements from different therapists also implies they are obtained from different sessions and trials. So, the ICC for inter-Therapist-Session-Trial might also be of interest to, say, data quality control, and is defined as

$$ICC(Therapist, Session, Trial) = Corr\left(\Phi_{ijkl}, \Phi_{ij'k'l'}\right) = \frac{\sigma^2_{subj}}{\sigma^2_{subj} + \sigma^2_{ther} + \sigma^2_{sess} + \sigma^2_{trial}}$$

4.2. Coefficient of Multiple Correlation (CMC)

Kadaba *et al.* [3] originally defined CMC in terms of various mean squares. In his paper, the experiments involve multiple days and runs, instead of sessions and trials. Therefore, we have used a different set of terms in this section. But statistically speaking, their 'days' is equivalent to our 'sessions' and their 'runs' is equivalent to our 'trials'. So, consider an experiment where a gait variable was measured on J runs nested within I days at time points $t_1, \ldots, t_T$. Kadaba et al. defined the intra-day CMC to be:

$$R^2_{intra-day} = 1 - \frac{\sum_{ijt}[\Phi_{ij}(t) - \bar{\Phi}_i(t)]/I(J-1)T}{\sum_{ijt}[\Phi_{ij}(t) - \bar{\Phi}_i]/I(JT-1)}$$

and the inter-day CMC to be:

$$R^2_{inter-day} = 1 - \frac{\sum_{ijt}[\Phi_{ij}(t) - \bar{\Phi}(t)]/(IJ-1)T}{\sum_{ijt}[\Phi_{ij}(t) - \bar{\Phi}]/(IJT-1)}$$

The idea behind such definitions is that the two numerators should capture the variance between the J runs and the I days respectively. An argument similar to section 2.3.1 should indicate that the

numerator for the inter-day CMC is a biased estimate of its target quantity. However, if we assume the following linear model

$$\Phi_{ij}(t) = \mu + T_t + Day_i + Run_j$$

where μ is the overall mean, and the rest are random effects for the time point, day, and run respectively. Then it is possible to derive a CMC based on the variance components:

$$\tilde{R}^2_{intra-day} = 1 - \frac{\sigma^2_{run}}{\sigma^2_{day}+\sigma^2_{run}+\sigma^2_t} = \frac{\sigma^2_{day}+\sigma^2_t}{\sigma^2_{day}+\sigma^2_{run}+\sigma^2_t}$$

$$\tilde{R}^2_{inter-day} = 1 - \frac{\sigma^2_{run} + \sigma^2_{day}}{\sigma^2_{day} + \sigma^2_{run} + \sigma^2_t} = \frac{\sigma^2_t}{\sigma^2_{day} + \sigma^2_{run} + \sigma^2_t}$$

The reason why CMC is dominated by the range of movement (ROM) [2,5] is because CMC considers the inter-time point difference as a variability, σ^2_t, which we believe is inappropriate. As a result, for joints with large ROM, σ^2_t will be a lot larger than all the other variance components. Nonetheless, we have shown how the CMC can be derived from the estimated variance components.

Results

Figure 2 plots the variance component curves as found in figure 4 of Schwartz *et al.* [1], together with the corrected estimates (same as the ANOVA estimates) using methods outlined above. In addition, the square root of the mean of the ANOVA curve, representing the average standard deviation, is also reported. Finally, the ratio of extrinsic to intrinsic variation (see 2.3.4) is reported in the left panel. This ratio will be slightly different to the ratio that would have been calculated using the individual average standard deviations. This is due to the difference in the order of the mathematical operations involved in the calculation (average then square-root vs square root then average). From the figure, we can see that the bias of Schwartz *et al.* method [1] increases from inter-Trial, inter-Session, to inter-Therapist. In particular, the original estimators usually over-estimate the inter-Therapist variability, and in some cases, quite severely so. In cases such as Knee Rotation and Ankle Dorsiflexion, the corrected (ANOVA) estimates show that the inter-Therapist variability is no longer the dominant source of variability.

DOI: 10.1016/j.gaitpost.2017.04.040 12 May. 17

Figure 2 Estimated variance component in standard deviation (degree) using both Schwartz et al. original estimators and the ANOVA estimators. Both time points estimate (curves) and average over the gait cycle (ANOVA only, top left hand corner) are provided. The ratio of extrinsic to intrinsic variability, r (cf. section 2.3.4), is provided for each angle.

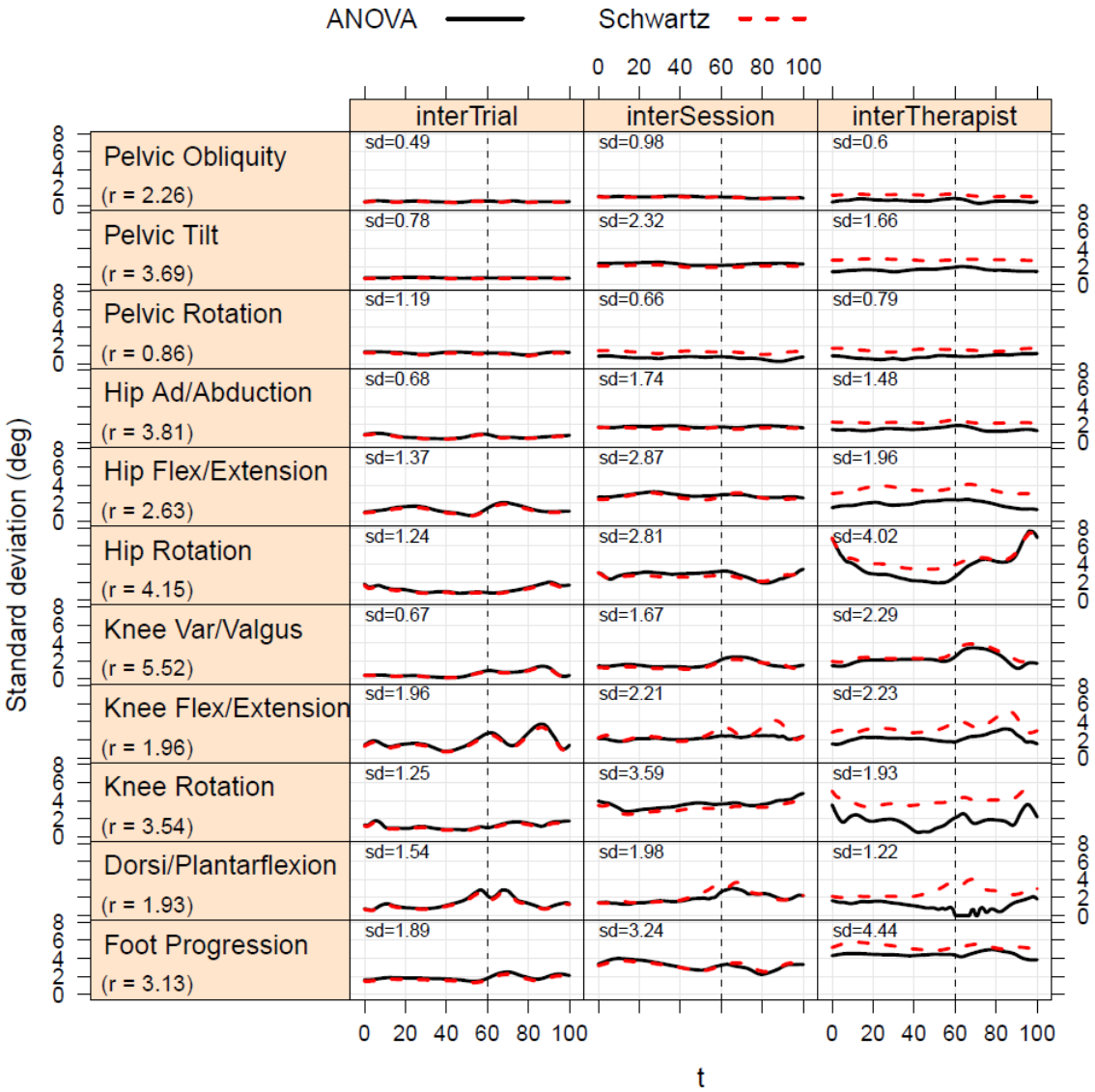
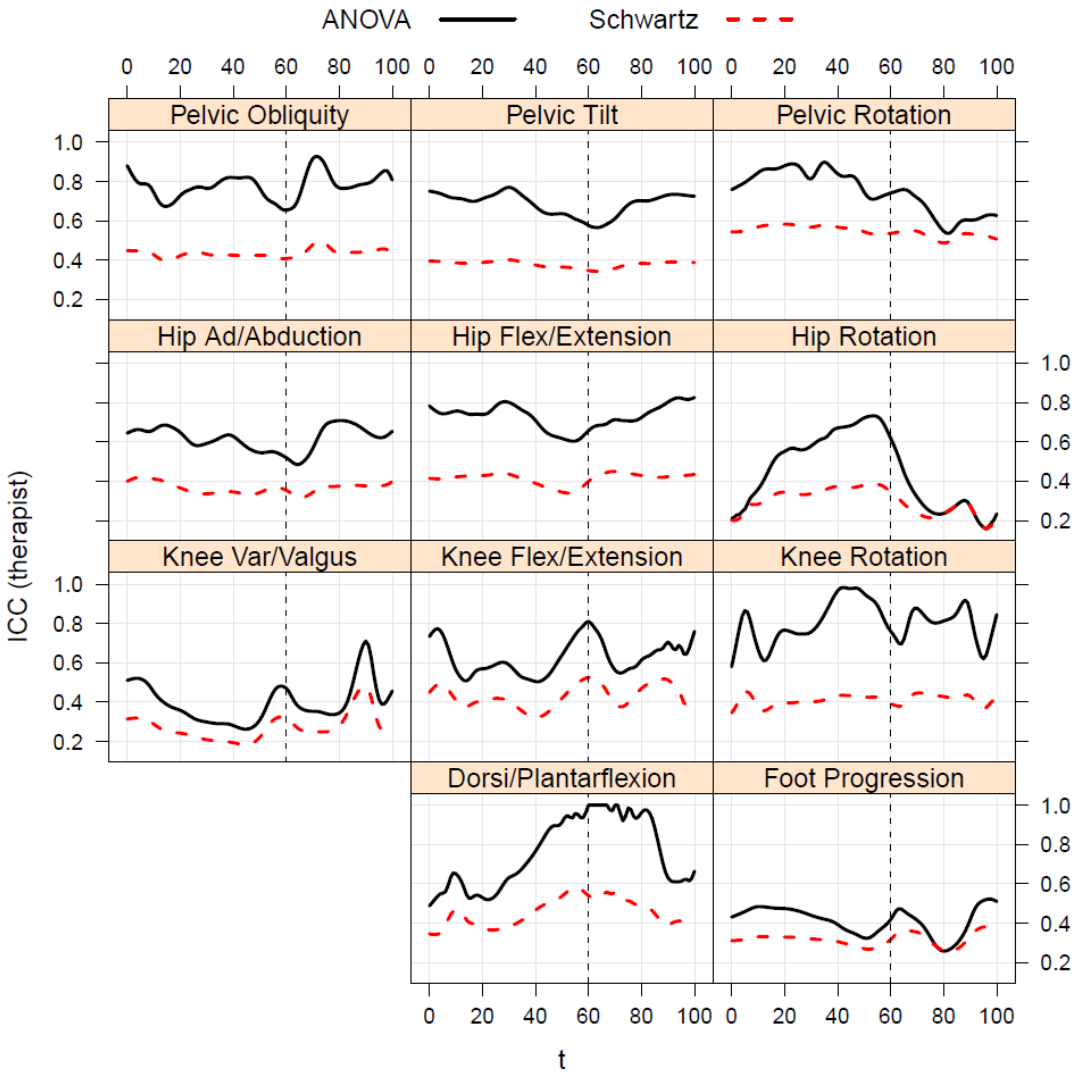


Figure 3 Estimated Intra Class Correlation (ICC) for therapist using both Schwartz et al. estimator and the ANOVA estimator.



Discussion

We have presented ANOVA based, and maximum likelihood based, methods to estimate and correct the variance components as produced by Schwartz *et al.* [1]. In addition, we have presented methods to derive reliability indices such as ICC and CMC from the estimated variance components. We have also shown that the original uncorrected estimator severely over-estimated the variance components at the higher level, which led to a more severely under-estimated ICC.

Interestingly, the ANOVA style of estimating variance components was found in McDowell et al [17], and the maximum likelihood type method in McGinley et al [18,19]. However, these methods are not widely adopted by the gait analysis community. Instead, it is quite common to see CMC [3] being used as reliability measure. We believe the variance components should always be the fundamental quantities to be computed when investigating issues related to variability, reliability, or repeatability. This is because the variance components are actually variances which are well-understood statistical quantities in meaningful units (degree^2). In addition, reliability indices such as ICC and CMC can be derived from these variances. Furthermore, they are applicable to a wide range of scenarios.

So far, we have shown how to estimate the variance components at individual time points. However, the interest may be to obtain a variability measure for the entire curve as one number. Consider the curve as represented by the vector $\mathbf{\Phi} = (\Phi(\mathrm{t}_1), \Phi(\mathrm{t}_2), \ldots, \Phi(\mathrm{t}_\mathrm{T}))'$, the variance of $\mathbf{\Phi}$ is naturally a $T \times T$ covariance matrix $\Sigma_{\mathbf{\Phi}}$, where the (i, j) entry is the covariance $cov(\Phi(t_i), \Phi(t_j))$. Under the linear model (1), without interaction, Σ_Φ can be decomposed into $\Sigma_{Subj} + \Sigma_{Ther} + \Sigma_{Sess} + \Sigma_{Trial}$. We then seek a single number summary for each of the Σ_{source}. Borrowing from the literature on optimal experimental design [20], some of the common options are the trace and the determinant of the covariance matrix. The determinant requires the entire covariance matrix to be estimated, which will be difficult in practice. The trace, however, sums the variance at the individual time points (diagonal terms), and is therefore equivalent to the average variance [2]. Therefore, purely for practicality we recommend averaging the variance components at each time point to arrive at a curve based measure.

References

[1] M.H. Schwartz, J.P. Trost, R.A. Wervey, Measurement and management of errors in quantitative gait data, Gait Posture. 20 (2004) 196–203.
[2] M. Sangeux, E. Passmore, H.K. Graham, O. Tirosh, The gait standard deviation, a single measure of kinematic variability, Gait Posture. 46 (2016) 194–200.
[3] M.P. Kadaba, H.K. Ramakrishnan, M.E. Wooten, J. Gainey, G. Gorton, G.V.B. Cochran, Repeatability of Kinematic, Kinetic, and EMG Data in Normal Adult Gait.pdf, J. Orthop. Res. 7 (1989) 849–860.
[4] J.L. Fleiss, The Design and Analysis of Clinical Experiments, John Wiley & Sons, Inc., Hoboken, NJ, USA, 1999.
[5] J. Røislien, Skare, A. Opheim, L. Rennie, Evaluating the properties of the coefficient of multiple correlation (CMC) for kinematic gait data, J. Biomech. 45 (2012) 2014–2018.
[6] K. Deschamps, F. Staes, H. Bruyninckx, E. Busschots, E. Jaspers, A. Atre, K. Desloovere, Repeatability in the assessment of multi-segment foot kinematics, Gait Posture. 35 (2012) 255–260.
[7] P. Salvia, S.V.S. Jan, A. Crouan, L. Vanderkerken, F. Moiseev, V. Sholukha, C. Mahieu, O. Snoeck, M. Rooze, Precision of shoulder anatomical landmark calibration by two approaches: A CAST-like protocol and a new anatomical palpator method, Gait Posture. 29 (2009) 587–591.
[8] K. Deschamps, F. Staes, H. Bruyninckx, E. Busschots, G.A. Matricali, P. Spaepen, C. Meyer, K. Desloovere, Repeatability of a 3D multi-segment foot model protocol in presence of foot deformities, Gait Posture. 36 (2012) 635–638.
[9] K. Deschamps, P. Roosen, H. Bruyninckx, K. Desloovere, P.A. Deleu, G.A. Matricali, L. Peeraer, F. Staes, Pattern description and reliability parameters of six force-time related indices measured with plantar pressure measurements, Gait Posture. 38 (2013) 824–829.
[10] P. Caravaggi, M.G. Benedetti, L. Berti, A. Leardini, Repeatability of a multi-segment foot protocol in adult subjects, Gait Posture. 33 (2011) 133–135.
[11] M. Manca, A. Leardini, S. Cavazza, G. Ferraresi, P. Marchi, E. Zanaga, M.G. Benedetti, Repeatability of a new protocol for gait analysis in adult subjects, Gait Posture. 32 (2010) 282–284.
[12] K. Kaufman, E. Miller, T. Kingsbury, E. Russell Esposito, E. Wolf, J. Wilken, M. Wyatt, Reliability of 3D gait data across multiple laboratories, Gait Posture. 49 (2016) 375–381.
[13] M. Mitchell, Engauge Digitizer, (2014).
[14] C. De Boor, A practical guide to splines, Springer-Verlag New York, 1978.
[15] H. Brown, R. Prescott, Applied Mixed Models in Medicine, John Wiley & Sons, Ltd, Chichester, UK, 2014.
[16] H.D. Patterson, R. Thompson, Recovery of inter-block information when block sizes are unequal, Biometrika. 58 (1971) 545–554.
[17] B.C. McDowell, R. Baker, V. Hewitt, a. Nurse, T. Weston, T. Dusoir, The variability of goniometric measurements in children with spastic cerebral palsy, Gait Posture. 10 (1999) 64.
[18] J. McGinley, R. Baker, R. Wolfe, Quantification of kinematic measurement variability in gait analysis, Gait Posture. 24 (2006) 55–56.
[19] J. McGinley, R. Wolfe, M. Morris, M.G. Pandy, R. Baker, Variability of walking in able-bodied adults across different time intervals, J. Phys. Med. Rehabil. Sci. 17 (2014) 6–10.
[20] A. Atkinson, A. Donev, R. Tobias, Optimum Experimental Designs, With SAS, Oxford university press, Oxford, 2007.

DOI: 10.1016/j.gaitpost.2017.04.040 12 May. 17

Appendix

Proof that ICC is a function of the estimated variance components

We prove the following equation only, all other ICCs equalities can be proved in a similar manner.

$$\widetilde{ICC_{ther}} = Corr(\Phi_{ijkl}, \Phi_{ij'k'l'}) = \frac{\sigma^2_{subj}}{\sigma^2_{subj} + \sigma^2_{ther} + \sigma^2_{sess} + \sigma^2_{trial}}$$

First, recall the linear model without interaction,

$$\Phi_{ijkl} = \mu + Sub_i + Ther_j + Sess_{ijk} + Trial_{ijkl}$$

Then,

$$\Phi_{ijkl} - \Phi_{ij'k'l'} = Ther_j - Ther_{j'} + Sess_{ijk} - Sess_{ij'k'} + Trial_{ijkl} - Trial_{ij'k'l'}$$

So the variance is,

$$var(\Phi_{ijkl} - \Phi_{ij'k'l'}) = 2\sigma^2_{ther} + 2\sigma^2_{sess} + 2\sigma^2_{trial} \qquad (2)$$

But we also know that,

$var(\Phi_{ijkl} - \Phi_{ij'k'l'}) = var(\Phi_{ijkl}) + var(\Phi_{ij'k'l'}) - 2cov(\Phi_{ijkl}, \Phi_{ij'k'l'})$

Therefore,

$var(\Phi_{ijkl} - \Phi_{ij'k'l'}) = 2(\sigma^2_{subj} + \sigma^2_{ther} + \sigma^2_{sess} + \sigma^2_{trial}) - 2cov(\Phi_{ijkl}, \Phi_{ij'k'l'})$ (3)

Equating (3) and (2) we get

$$cov(\Phi_{ijkl}, \Phi_{ij'k'l'}) = \sigma^2_{subj}$$

And dividing both sides by the total variance $\sigma^2_{subj} + \sigma^2_{ther} + \sigma^2_{sess} + \sigma^2_{trial}$ concludes our proof.