

## DETERMINING OPTIMAL TRIAL SIZE USING SEQUENTIAL ANALYSIS

Paul Taylor<sup>1</sup>, Kwee-Yum Lee<sup>1</sup>, Raul Landeo<sup>1</sup>, Damien O'Meara<sup>1,2</sup> and Emma Millett<sup>1,2</sup>

<sup>1</sup>Australian Catholic University, Sydney, Australia

<sup>2</sup>New South Wales Institute of Sport, Sydney, Australia

This study examined the response of sequential analysis (SA) results to conditions of different trial size to ascertain the optimal number of trials for mean stability in kinematic data. Thirty overarm throws were performed by twenty participants. Discrete and time series kinematic data were submitted to SA in three main trial size conditions (first 10, 20, 30 trials) and three comparative conditions (mid and last 10, mid 20 trials) to derive the number of trials (SA score) to a stable mean for each condition. Results showed that the SA score is dependent on the trial size of the condition but not on the position where the subsample was drawn. As the first 10 condition responded differently to the other conditions, we recommend the use of the SA score of the 20 trial conditions to ascertain stable means for parsimonious data collection.

**KEY WORDS:** sequential analysis, mean stability, trial size, overarm throw, kinematic data.

**INTRODUCTION:** Sequential analysis (SA) can determine the minimum number of samples from an individual or group which provide an acceptable estimate of stability in the mean (Wald, 1947). As such, the SA technique has been used to determine the number of trials to stability in numerous biomechanical measures including ground reaction forces during running (Bates, Osternig, Sawhill, & James, 1983), walking (Hamill & McNiven, 1990), landing (James, Herman, Dufek, & Bates, 2007), jumping (Racic, Pavic, & Brownjohn, 2009) and cricket bowling (Stuelcken & Sinclair, 2009). The SA technique quantifies trials to stability using a moving point mean coupled with a stability criterion. This criterion is a bandwidth calculated from the mean and standard deviation (SD) of total trials (commonly mean  $\pm$  0.25 SD). Due to the arbitrary selection of the trial sizes used for SA, it is important to understand the effect they have on the estimation of the number of trials required to supply the stable trial mean. The aim of this study was to investigate the effect of applying different trial sizes on the results of the SA using kinematic data from overarm throwing.

**METHODS:** Ten males and ten females provided informed consent and participated in this study which was approved by the Human Research Ethics Committee of The Australian Catholic University. Three dimensional (3-D) motion capture was performed using 10 Vicon cameras (6 MX and 4 T-Series, Oxford Metrics, Oxford, UK) sampling at 400 Hz using Vicon Nexus software (Oxford Metrics, Oxford, UK). Two dimensional (2-D) data of the ball trajectory in the sagittal plane were captured using a Basler A602fc camera (Basler AG, Germany) sampling at 100Hz. Participants performed 30 overarm throws towards a circular target (70 cm dia.) projected on a cloth screen at a distance of 7 m. Participants were seated on an adjustable stool at 90° knee flexion with their hands on their knees and directed to throw a regulation tennis ball as accurately as possible toward the centre of the target using their preferred hand. No other instructions were given. A self-determined period of familiarisation was allowed before data collection. Time between throws was self-paced.

Several kinematic variables commonly analysed in overarm throwing were selected. To represent 3-D displacement values in three axes for proximal, distal, bony and fleshy location, four anatomical markers were chosen: 10<sup>th</sup> thoracic vertebra, lateral upper arm (over the muscle belly of triceps), radial epicondyle and distal end of the 3<sup>rd</sup> metacarpal of the throwing arm. Six joint angles - shoulder internal/external rotation and flexion/extension at the elbow and wrist - from the kinematic model (Unilateral Vicon Upper Limb Model) were chosen for their critical role in producing ball velocity (Van Den Tillaar & Ettema, 2004). Discrete values

of the final determinants of ball trajectory (ball release angle, height and velocity) were also included from 2-D data. Following analysis of the frequency content and residuals of the power spectra (Winter, 2005) a cut off frequency of 12 Hz was employed in a low pass filter (Butterworth dual-pass, 4<sup>th</sup> order) on the time series (TS) data. The start of the movement was defined as the beginning of elbow flexion during wind up, and the end point was ball release. Filtered data were trimmed to these instants and time-normalised to 101 data points.

The SA technique (see Hamill & McNiven, 1990 for more details) was employed on both discrete and TS kinematic data (Table 1). To perform SA on 3-D marker displacement and joint angle TS data, each of the 101 points were treated as a discrete point, providing trial to stability for each point along the entire TS. To determine the effect of different trial sizes on SA score, three main conditions (first 10, 20, 30 trials) were assessed with the  $n^{\text{th}}$  trial mean and 0.25 SD bandwidth calculated from each condition. In addition, the first, mid and last 10 and mid 20 conditions were compared to determine if results were dependant on where in the sequence of throws a sample was extracted.

Table 1: Discrete and time series (*italics*) variables included in sequential analysis

| <b>Marker Variables<br/>(X, Y, Z)</b> | <b>Three Joint Angle Variables<br/>(shoulder external rotation, elbow flexion, wrist extension)</b> | <b>Ball Release Variables</b> |
|---------------------------------------|---|-------------------------------|
| Maximum value                         | Peak angle value  | Release height                |
| Minimum value                         | Time of peak angle value  | Release velocity              |
| Value at release                      | Value at release  | Release angle                 |
| <i>Normalised time series</i>         | <i>Normalised time series</i>   |                               |

*Note.* Peak angle value represents relevant maximum or minimum displacement, occurring near the end of the wind up phase prior to release.

In order to compare discrete and TS SA results, repeated measures analysis of variance (ANOVA) was performed. The moving point average for all discrete variables across the 30 trials were plotted to qualitatively assess patterns relative to the 0.25 SD bandwidth (see Figure 1). Following initial analysis of these results, discrete SA scores were converted to a relative percentage by dividing the score by the trial size for that condition and compared using repeated measures ANOVA. These comparisons allowed identification of any differences in the behaviour of the technique relative to trial size. In addition to individual variables, group mean relative SA scores for each marker and joint angle variable were combined and compared across conditions using repeated measures ANOVA. Means and SDs of the discrete values at release were compared across conditions using repeated measures ANOVA to examine the effect of bandwidth criteria on the technique. Fisher's least significant difference (LSD) and Tukey post hoc tests were performed on discrete and TS variables respectively to further explore significant effects (alpha level = 0.05). All statistical analyses were performed using IBM SPSS Statistics, version 19 (SPSS Inc, Chicago, Illinois) for discrete variables and Statistica 7 (StatSoft Inc, Tulsa, Oklahoma) for time series.

**RESULTS:** Figure 1 shows a sample plot of the SA (on the minimum value of Finger marker in X axis) from one participant, illustrating the most common pattern (54%; 523/960) observed amongst the discrete variables in the 30 trial condition. It can be seen that the moving point mean (solid black line) undergoes a transition phase, commonly during the first 10 trials, moving up or down toward the criterion bandwidth (dash grey line). Following the 10<sup>th</sup> trial, fluctuations in the moving point mean become less severe. After the point of stability (trial 15 in this example), and even slightly before, the mean is robust to fluctuations in raw data (dash-dot grey line), illustrating the concept of SA score and mean stability.

SA scores were significantly different across the three main conditions for all discrete variables,  $F(1, 19) \geq 191.95$ ,  $p < 0.05$ . Scores from the first 10 condition were different from

the first 20 and 30 condition in all 90 pairwise comparisons, and between the first 20 and first 30 in 44 of 45 pairwise comparisons ( $p < 0.05$ ). Of 45 pairwise comparisons between the first and mid 10 conditions 2 were significantly different ( $p < 0.05$ ). No differences existed between first and last 10 conditions ( $p \geq 0.06$ ). First and mid 20 differed significantly once in 45 pairwise comparisons ( $p < 0.05$ ). SA scores were significantly different across all TS variables,  $F(3, 57) \geq 48.51$ ,  $p < 0.05$ . Pairwise comparisons displayed significant differences between all conditions of different sizes ( $p < 0.05$ ) while same sized conditions formed homogenous groups. Means and SDs of discrete variables at release were not significantly different across all conditions ( $p \geq 0.05$  for 64/72 pairwise comparisons)

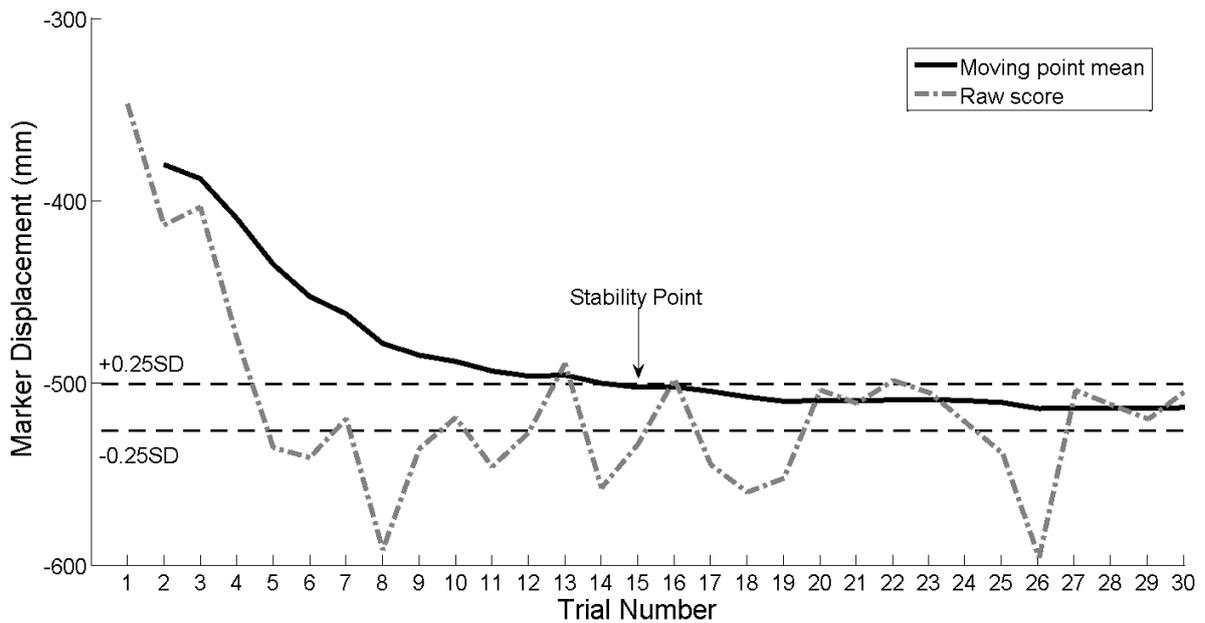


Figure 1: Example of sequential analysis (SA) of a finger marker minimum value in 30 trial condition showing the relationship between raw kinematic data (dash-dot grey), moving point mean (solid black) and 0.25 standard deviation bandwidth (dash grey). Stability point (SA score) is indicated by an arrow.

Results for discrete relative group mean marker variables,  $F(1, 11) \geq 3304.52$ ,  $p < 0.05$ , group mean joint angle variables,  $F(1, 8) = 2831.88$ ,  $p < 0.05$ , and all individual discrete variables,  $F(1, 19) \geq 420.31$ ,  $p < 0.05$ , displayed significance. Main condition pairwise post hoc results are summarised in Table 2, and results for subsample comparisons are listed in Box 1. Of 15 relative TS data, 12 displayed significance,  $F(3, 57) \geq 16.70$ ,  $p < 0.05$ . The first 10 condition was significantly greater than the other main conditions in 8 of 12 significant TS variables (excluding T10 Z, Upper arm Z and Elbow Y),  $F(3, 57) \geq 4.78$ ,  $p < 0.05$ . Same sized conditions formed homogenous groups based on sample size.

Table 2: Number of significant post hoc pairwise comparisons ( $p < 0.05$ ) for group mean (out of 4) and individual (out of 48) discrete variables for three main conditions

|          | Group Mean Pairwise (out of 4) |          |          | Individual Variable Pairwise (out of 48) |          |          |
|----------|--------------------------------|----------|----------|--|----------|----------|
|          | First 10                       | First 20 | First 30 | First 10                                 | First 20 | First 30 |
| First 10 |                                | 4        | 4        |  | 11       | 12       |
| First 20 | 4                              |          | 1        | 11                                       |          | 4        |
| First 30 | 4                              | 1        |          | 12                                       | 4        |          |

Box 1: Number of significant pairwise comparisons ( $p < 0.05$ ) between subsamples

|  |  |
|--|--|
| <b>Group Means (out of 4):</b>           | First 10 & Mid 10 (1)<br>First 20 & Mid 20 (1)                         |
| <b>Individual Variables (out of 48):</b> | First 10 & Mid 10 (2)<br>Mid 10 & Last 10 (1)<br>First 20 & Mid 20 (1) |

**Discussion:** SA results for both discrete and TS variables showed that the outcome of this technique is dependent on the trial size from which criterion mean and SD values are drawn. Yet, results were not dependant on the position in the total sample where the subsample is drawn. Qualitative assessment of the SA plots suggests that the results from the first 10 condition may be affected by the “transition” phase of the moving point mean. Results of relative scores support this, showing that the first 10 condition often produces a relative score (65.6%) higher than the first 20 and 30 conditions (59.0% and 56.9% respectively). These results are sufficient to exclude the first 10 condition as a supply of valid SA results to determine the number of trials to stable means.

With the first 10 condition excluded, it must also be considered whether to accept SA values from either the first 20 or 30 trials. Determinants of SA score might lie outside the bandwidth criteria of the sample, since mean and the variance at release were not significantly different across all conditions. As the raw mean does not vary statistically, nor are relative scores consistently different from the 20<sup>th</sup> to 30<sup>th</sup> trial, collecting 20 trials would suffice to estimate stable means. Time and/or budget constraints, learning and fatigue are some factors that may provide further justification to perform SA on a sample of 20 trials. Qualitatively, some evidence of fatigue was noticed within the population in the final ten throws. However, it is possible that the 30 trials condition may be more appropriate for estimating stable mean values in other tasks and populations.

**Conclusion:** Based on the results, performing SA on a sample of 20 trials to ascertain an acceptable estimate of the mean of the kinematic data in the overarm throwing task is recommended. Furthermore, the use of similar methods presented here to determine the required sample size for SA in other populations and tasks are suggested. For researchers and practitioners this method may be implemented on pilot samples of the target population to guide data collection and trial size decisions in studies with larger samples. As with any statistical technique, however, it should also be carried out with a good understanding of the expected outcome and the factors (e.g., learning and fatigue) that may influence the analysis.

#### REFERENCES:

- Bates, B. T., Osternig, L. R., Sawhill, J. A., & James, S. L. (1983). An assessment of subject variability, subject-shoe interaction, and the evaluation of running shoes using ground reaction force data. *Journal of Biomechanics*, 16(3), 181-191. doi: 10.1016/0021-9290(83)90125-2
- Hamill, J., & McNiven, S. L. (1990). Reliability of selected ground reaction force parameters during walking. *Human Movement Science*, 9(2), 117-131. doi: 10.1016/0167-9457(90)90023-7
- James, C. R., Herman, J. A., Dufek, J. S., & Bates, B. T. (2007). Number of trials necessary to achieve performance stability of selected ground reaction force variables during landing. *Journal of Sports Science and Medicine*, 6(1), 126-134. Retrieved from <http://www.jssm.org/>
- Racic, V., Pavic, A., & Brownjohn, J. M. W. (2009). Number of successive cycles necessary to achieve stability of selected ground reaction force variables during continuous jumping. *Journal of Sports Science and Medicine*, 8, 639-647. Retrieved from <http://www.jssm.org/>
- Stuelcken, M. C., & Sinclair, P. J. (2009). A pilot study of the front foot ground reaction forces in elite female fast bowlers. *Journal of Science and Medicine in Sport*, 12(2), 258-261. doi: 10.1016/j.jsams.2007.10.016

Van Den Tillaar, R., & Ettema, G. (2004). A force-velocity relationship and coordination patterns in overarm throwing. *Journal of Sports Science and Medicine*, 3(4), 211-219. Retrieved from <http://www.jssm.org/>

Wald, A. (1947). *Sequential analysis*. New York: Wiley.

Winter, D. A. (2005). *Biomechanics and motor control of human movement* (3rd ed.). Hoboken, NJ: Wiley.