©Journal of Sports Science and Medicine (2007) **6**, 126-134
http://www.jssm.org

Research article

Number of trials necessary to achieve performance stability of selected ground reaction force variables during landing

C. Roger James [1]✉, Joseph A. Herman [2], Janet S. Dufek [3] and Barry T. Bates [4]
[1] Center for Rehabilitation Research, Texas Tech University Health Sciences Center, Lubbock, Texas, USA
[2] Department of Health, Exercise, and Sport Sciences, Texas Tech University, Lubbock, Texas, USA
[3] Department of Kinesiology, University of Nevada, Las Vegas, Las Vegas, Nevada, USA
[4] Human Performance & Wellness, Inc., Henderson, Nevada, USA

Abstract

The objectives were to determine the number of trials necessary to achieve performance stability of selected ground reaction force (GRF) variables during landing and to compare two methods of determining stability. Ten subjects divided into two groups each completed a minimum of 20 drop or step-off landings from 0.60 or 0.61 m onto a force platform (1000 Hz). Five vertical GRF variables (first and second peaks, average loading rates to these peaks, and impulse) were quantified during the initial 100 ms post-contact period. Test-retest reliability (stability) was determined using two methods: (1) intra-class correlation coefficient (ICC) analysis, and (2) sequential averaging analysis. Results of the ICC analysis indicated that an average of four trials (mean 3.8 ± 2.7 Group 1; 3.6 ± 1.7 Group 2) were necessary to achieve maximum ICC values. Maximum ICC values ranged from 0.55 to 0.99 and all were significantly ($p \leq 0.05$) different from zero. Results of the sequential averaging analysis revealed that an average of 12 trials (mean 11.7 ± 3.1 Group 1; 11.5 ± 4.5 Group 2) were necessary to achieve performance stability using criteria previously reported in the literature. Using 10 reference trials, the sequential averaging technique required standard deviation criterion values of 0.60 and 0.49 for Groups 1 and 2, respectively, in order to approximate the ICC results. The results of the study suggest that the ICC might be a less conservative, but more objective method for determining stability, especially when compared to previous applications of the sequential averaging technique. Moreover, criteria for implementing the sequential averaging technique can be adjusted so that results closely approximate the results from ICC. In conclusion, subjects in landing experiments should perform a minimum of four and possibly as many as eight trials to achieve performance stability of selected GRF variables. Researchers should use this information to plan future studies and to report the stability of GRF data in landing experiments.

Key words: reliability, variability, sequential averaging, intra-class correlation coefficient.

Introduction

Stability of a performance variable refers to the repeatability of that variable across repeated trials (observed performances) over time and can be evaluated using test-retest reliability methods (Portney and Watkins, 2000). The stability of a variable across trials influences the stability of the mean value of the group of trials. When the mean value is not stable, both the reliability of the mean and its ability to represent a more generalized performance (validity) are limited. The number of trials obtained from an individual in an experiment is thought to influence stability (Bates et al., 1983; Salo et al., 1997) and thus is an important methodological consideration in the design of landing experiments.

Except for unique circumstances (e.g., a single trial is the subject of interest) several trials are thought to provide a more stable and representative mean value (Bates et al., 1983). Because variability is present in all human movement, using too few trials may not represent the individual's long-term performance. A single trial protocol has been suggested to be both invalid and unreliable (Bates et al., 1992) because of the potential inability of the single trial to represent the generalized performance. By chance the single trial could represent an average performance but also might be atypical. Greater movement variability results in less stable data and a greater likelihood of sampling an atypical performance from the population of all possible performances. Stability may be particularly important when trials are obtained in non-continuous activities (e.g., a discrete movement such as a jump or landing) or in a nonconsecutive manner in continuous activities (e.g., nonconsecutive strides in running). While, increasing the number of trials is thought to increase performance stability (Bates et al., 1983; Salo et al., 1997), how many trials are necessary to provide stable data? Although a few studies have examined this issue for nonconsecutive trials during the activities of running (Bates et al., 1983; 1992), walking (Hamill and McNiven, 1990), hurdling (Salo et al., 1997) , and vertical jumping (Rodano and Squadrone, 2002), little information is available about the number of trials necessary to achieve performance stability for nonconsecutive trials during landing. Moreover, different studies have used either different arbitrary criteria or different methods for determining stability, making comparisons among studies difficult.

Running (Bates et al., 1983), walking (Hamill and McNiven, 1990), and vertical jumping (Rodano and Squadrone, 2002) all have been examined for performance stability of nonconsecutive trials using a sequential averaging estimation technique (see Methods). For running, results of the sequential averaging technique (using 10 reference trials and a 0.25 standard deviation criterion value) demonstrated that eight nonconsecutive steps (trials) were necessary to obtain stable data in 43 ground reaction force variables (Bates et al., 1983). Similar results were found when increasing the number of reference trials from 10 to 20 (Bates et al., 1983). For walking, the sequential averaging technique (using 20 reference trials

Received: 23 August 2006 / Accepted: 26 January 2007 / Published (online): 01 March 2007

and a 0.25 standard deviation criterion value) was used to determine that 10 nonconsecutive trials were necessary to reach performance stability of selected ground reaction force variables (Hamill and McNiven, 1990). For vertical jumping, the sequential averaging technique (using 25 reference trials and a 0.30 standard deviation criterion value) was used to determine that 12 trials were needed to establish performance stability of selected joint kinetic variables (Rodano and Squadrone, 2002). A limitation of the sequential averaging technique is that the number of reference trials and the standard deviation criterion value both influence the results, yet the values selected are arbitrary.

Other investigators have used a variety of methods for examining the reliability, stability, and variability of gait variables both within and between days (Belli et al., 1995; Kadaba et al., 1989; Owings and Grabiner, 2003; Winter, 1984) and for consecutive (Belli et al., 1995; Owings and Grabiner, 2003) and nonconsecutive (Kadaba et al., 1989; Winter, 1984) trials. For example, Kadaba and colleagues calculated the coefficient of variation (CV) both within and between days to estimate the repeatability of spatiotemporal gait parameters, while the repeatability of kinematic, kinetic, and electromyographic wave forms were examined using an adjusted coefficient of multiple determination method (Kadaba et al., 1989). They suggested that data obtained from nonconsecutive trials from subjects walking at their preferred speeds were sufficiently repeatable (Kadaba et al., 1989). However, a limitation of their method was that the number of trials used to calculate repeatability was selected arbitrarily (three per session and nine per day). Conversely, Owings and Grabiner used running mean and standard deviation functions similar to the sequential averaging technique to examine the stability of selected gait variables over consecutive trials during treadmill walking for the purpose of calculating step variability (Owings and Grabiner, 2003). They suggested that at least 400 steps were required for accurate estimation of step kinematics (Owings and Grabiner, 2003). However, a limitation of their method was that many criteria used to establish stability across multiple steps of data also were selected arbitrarily. Belli and colleagues examined the absolute variability of total body vertical displacement and step time for consecutive trials during treadmill running at different velocities (Belli et al., 1995). They demonstrated that variability was relatively low at sub-maximal velocities, but increased at higher velocities (Belli et al., 1995). The absolute variability of each parameter was calculated as the standard deviation of each mean value, and was expressed as a percentage of the mean. They suggested that 32-64 steps were required to obtain better than 1% accuracy on the mean value (Belli et al., 1995). However, a limitation of their method was that the percentage value used to represent a desired accuracy was selected arbitrarily.

Using a more traditional statistical method for examining performance stability, Salo and colleagues utilized the intra-class correlation coefficient (ICC) to examine the stability of selected kinematic variables in nonconsecutive trials during sprint hurdling (Salo et al., 1997). They predicted that as few as one to as many as 78 trials were necessary to reach a reliability of 0.90, depending on the specific kinematic variable examined. However, a limitation of this study was that only eight trials were actually collected from subjects and evaluated for reliability. Moreover, the value eight (i.e., eight trials) was selected arbitrarily. Additionally, the number of trials predicted to reach pre-determined reliability values was determined using the Spearman-Brown Prophecy formula, which likely overestimated the reliability values for large numbers of trials.

While the number of trials necessary to achieve performance stability has been examined for a number of different locomotor tasks, the activity of landing has not been evaluated. Landing is an activity that has recently received much attention in the literature because of its implicit link to many lower extremity injuries, especially in female athletes (Griffin et al., 2000). Because the number of trials necessary to achieve performance stability during landing has not been established, the reliability (and consequently validity) of many landing studies which have used too few trials could be in question. Moreover, the method for establishing performance stability should be objective and not based on arbitrary criteria. While several statistical methods have been used to determine stability during gait and other activities, there appears to be no comparisons between methods. Therefore, the purpose of the current study was to answer the following questions: (1) How many trials are necessary to achieve performance stability during landing? (2) How do the results obtained from different methods of calculating performance stability compare to one another? It was hypothesized that similar to other locomotor tasks several trials would be necessary to achieve performance stability during landing. Additionally, it was hypothesized that different methods for determining stability would provide dissimilar results.

Methods

Experimental design

A test-retest design was used to examine the stability within a single testing session of selected ground reaction force variables during nonconsecutive landing trials.

Subjects

Data from ten recreationally-active college students (age range 22-31 yr; mass 73.6 ± 11.8 kg) who had no known pathologic ankle, knee, hip, or spinal conditions were included in the study. Five subjects (four men and one woman; Group 1) volunteered expressly for the current study, while the data from five other subjects (five men; Group 2) were originally obtained in a previous investigation (James et al., 2006). Each subject read and signed a written informed consent statement approved by the Institutional Review Board at the affiliated university. Subjects wore non-restrictive athletic clothing (i.e., shorts, t-shirt) and standard laboratory shoes.

Protocol

Two landing protocols that are commonly used in landing research were used to determine the stability of selected ground reaction force variables obtained from multiple nonconsecutive landing trials. The two protocols were step-off and drop landings, each performed using different subjects (Groups 1 and 2), as previously described. In each protocol, subjects completed a brief warm-up followed by a few practice landings designed to acquaint the subjects to the task. The specific details of each protocol are given below, but data were obtained and analyzed beginning with the first non-practice trial and in the order performed by the subjects.

Group 1

Prior to testing, individuals in landing Group 1 completed a five-minute warm-up on a cycle ergometer at a self-selected speed and resistance. Following the warm-up, subjects participated in a 60 minute session consisting of three to five practice and 100 experimental step-off landing trials. Subjects received a one-minute rest following every 10 trials. Only the first 20 usable trials were included in the stability analysis in the current investigation. To perform the experimental landings, an adjustable wooden platform (0.61 m) was positioned and aligned so that subjects would step-off with their right foot and land bilaterally with their right foot on a force platform (Model OR-6-7-2000 and SGA-6 bridge amplifier; Advanced Mechanical Technology Inc., Watertown, MA, USA) inset into the ground and their left foot on the adjacent floor. Subjects were not given explicit instructions about their landing technique, but were encouraged to land consistently and in a sport-ready position. The force platform was interfaced to an Ariel Performance Analysis System (analog module; San Diego, CA, USA) and data were sampled at 1000 Hz for 10 ms prior to and 190 ms following initial ground contact. Both the force platform and surrounding floor were individually covered with a 2.5 cm thick artificial turf surface that was affixed to the under-floor by adhesive tape.

Group 2

The protocol for subjects in landing Group 2 has been described elsewhere (James et al., 2006) and the relevant parts are presented here briefly. Prior to testing, subjects performed a self-directed warm-up consisting of light calisthenics and stretching. Following the warm-up, subjects participated in a testing session in which approximately 30 drop landing trials (0.60 m landing height) were performed in each of two experimental conditions, non-fatigued and fatigued. Only the first 20 usable non-fatigued landings were included in the stability analysis in the current investigation. In the earlier study (James et al., 2006), the first 10 trials were designated as practice trials for the purposes of that protocol. However, in the current study the first 10 trials were included as trials of interest. Drop landings were initiated by the subjects after hanging by their hands from an adjustable overhead bar. Subjects landed bilaterally with each foot on separate but adjacent force platforms (Model OR-6-5-1, Advanced Mechanical Technology, Inc., Watertown, MA). Only left side ground reaction force data were used in the current study. Similar to the subjects in Group 1, subjects in Group 2 were not given explicit instructions about their landing technique, but were encouraged to land consistently and in a sport-ready position. Ground reaction force data were sampled at 1000 Hz and recorded for 400 ms prior to and 200 ms following initial ground contact. Data were sampled and stored using a Modular Data Acquisition System (Model 7000, TransEra Corp., Orem, UT) interfaced to a laptop computer.

Data reduction

Vertical ground reaction force-time histories were reduced to five discrete variables during the first 100 ms of ground contact using a custom MatLab (v. 6.0, MathWorks, Inc., Natick, MA) program. The discrete variables were the first (F1) and second (F2) peak force magnitudes that occurred following contact, average loading rates to each peak (F1LR and F2LR, respectively), and impulse (IMP; trapezoid method) from 0-100 ms post-contact (Figure 1). F1LR was calculated as the slope of the line of the ground reaction force-time history from initial contact

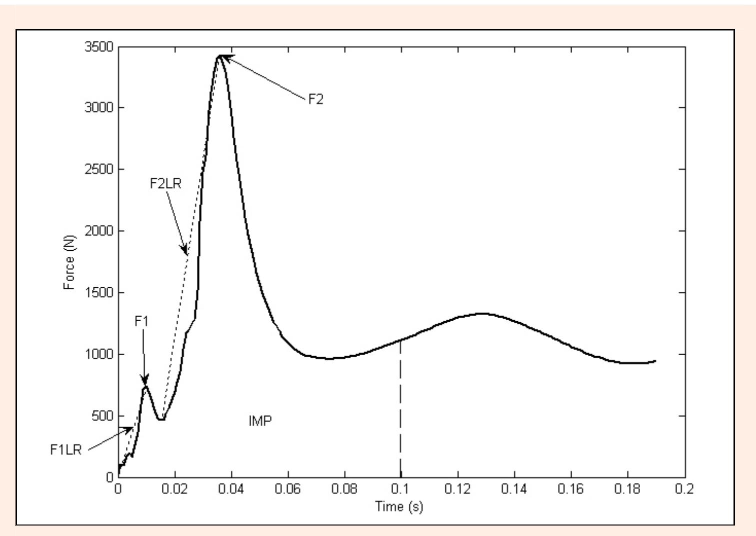


Figure 1. Example ground reaction force-time history and the five discrete dependent variables. F1 = first peak ground reaction force; F2 = second peak ground reaction force; F1LR = average loading rate from the instant of ground contact to the instant of F1; F2LR = average loading rate from the instant of the minimum force between F1 and F2 to the instant of F2. IMP = impulse (total force) calculated from 0-100 ms following initial ground contact.

Table 1. Ground reaction force 20-trial mean and (standard deviation) values for all subjects.

Subject Group	Subject Number	Sex	Mass (kg)	F1 (BW)	F2 (BW)	F1LR (BW/s)	F2LR (BW/s)	Impulse (BW·s)
1	1	male	70.8	.78 (.09)	2.33 (.58)	54.88 (7.24)	54.40 (25.82)	.11 (.01)
	2	male	70.9	1.94 (.12)	2.36 (.40)	128.50 (8.63)	37.50 (13.54)	.14 (.01)
	3	female	62.9	1.96 (.21)	3.13 (.56)	139.30 (20.19)	64.33 (26.14)	.16 (.01)
	4	male	77.8	1.18 (.13)	2.99 (.40)	82.01 (9.41)	68.11 (16.37)	.14 (.01)
	5	male	100.8	.42 (.04)	1.22 (.15)	33.58(4.78)	38.05 (12.00)	.05 (.00)
2	6	male	58.6	1.14 (.13)	3.93 (.64)	105.80 (10.77)	151.01 (57.80)	.16 (.01)
	7	male	75.0	.95 (.10)	2.92 (.34)	102.05 (13.42)	106.41 (35.19)	.14 (.01)
	8	male	78.6	.99 (.13)	1.48 (.32)	55.36 (9.05)	15.37 (9.60)	.10 (.01)
	9	male	77.3	1.12 (.09)	2.30 (.51)	86.89 (12.61)	49.83 (23.67)	.12 (.01)
	10	male	63.2	1.20 (.09)	5.73 (.53)	112.14 (10.74)	237.99(47.70)	.20 (.02)

Notes: Subject Group 1 performed 0.61 m step-off landings; Group 2 performed 0.60 m drop landings. Units of BW represent multiples of body weight.

to F1. F2LR was calculated as the slope of the line of the ground reaction force-time history from the minimum force value that occurred between F1 and F2 to F2. These variables were selected because they represent different characteristics of ground reaction force loading (peaks, loading rates, total force), have been previously quantified in landing studies examining ground reaction force variables (e.g., James et al., 2006), and are analogous to many variables quantified in previous studies examining the stability of ground reaction force data during running (Bates et al., 1983) and walking (Hamill and McNiven, 1990).

Statistical analyses

Mean and standard deviation values were calculated for each ground reaction force variable. Descriptive variables were reported in multiples of body weight (BW). However, for the stability analyses, ground reaction force variables were not normalized to body weight because: (1) all stability evaluations were made within subject, and (2) the ICC analysis is more sensitive when there is greater variation between subjects in the numerical values of the dependent variables (Portney and Watkins, 2000).

In order to test the experimental hypotheses, the performance stability of each non-normalized ground reaction force variable across trials was quantified for each subject group using two methods: (1) test-retest intra-class correlation coefficient for single measures (ICC; Model 3, 1) (Portney and Watkins, 2000), and (2) sequential averaging technique. The ICC was selected as a traditional method for determining stability, while the sequential averaging technique was used in order to facilitate comparison with previous research that has reported the stability of selected variables during running (Bates et al., 1983), walking (Hamill and McNiven, 1990), and vertical jumping (Rodano and Squadrone, 2002). Both stability analysis methods were used to test the first (number of trials necessary for stability) and second (comparison of methods) hypotheses.

Intra-class correlation coefficient analysis

The stability of each ground reaction force variable initially was calculated for each subject group by using ICC (Model 3, 1) applied to the first two landing trials. The ICC calculation was then iteratively repeated in increments of one trial for combinations of trials ranging from three to 20. The maximum ICC value for all iterations, the number of trials needed to achieve the maximum ICC value, the probability that the maximum ICC value was significantly different from zero, and the ICC 95% confidence interval upper and lower limits were determined. Additionally, the number of trials necessary to reach ICC values of 0.80, 0.85, and 0.90 were calculated. All ICC procedures were completed using SPSS v12. The criterion alpha-level for establishing statistical significance was set to 0.05.

Sequential averaging analysis

The ground reaction force variables also were examined for performance stability across the first 20 landing trials for each subject in each group using a sequential averaging technique as described in the literature for running (Bates et al., 1983), walking (Hamill and McNiven, 1990), and vertical jumping (Rodano and Squadrone, 2002). The process involved computing the mean, standard deviation, and 0.25 standard deviation values for the first 20 trials. Then, the cumulative mean and mean deviation values were computed for each of the 20 trials in the order that they were obtained experimentally. A cumulative mean was calculated as the average of each trial with all previous trials. This calculation was repeated in succession for all trials from one to 20 for each subject and variable. Using this procedure, the final cumulative mean value was identical to the overall 20 trial mean. A mean deviation value was calculated as the absolute difference between the cumulative mean of a corresponding trial and the mean of all 20 trials. Finally, stability was estimated as one greater than the smallest trial number for which all successive mean deviations were smaller than the 0.25 standard deviation criterion value for that particular subject and variable (Bates et al., 1983). The 20 trial and 0.25 standard deviation criterion values were chosen to facilitate comparison to previous results reported for running (Bates et al., 1983) and walking (Hamill and McNiven, 1990).

The sequential averaging procedure was repeated using a 10 reference trial data set and a 0.25 standard deviation criterion value in order to examine differences in stability that might result from using a different number of reference trials. Additionally, the sequential averaging procedure was repeated for the 10 reference trial data set by iteratively varying the standard deviation criterion value until the number of trials necessary for stability approximated the results obtained from the ICC analysis.

Table 2. **Summary of ICC analyses for subject Group 1 (step-off landing).**

ICC Value	Statistic	F1	F2	F1LR	F2LR	IMP	Mean (SD)
ICC Maximum	n-trials	5	8	2	2	2	3.8 (2.7)
	ICC	.71	.91	.77	.87	.98	
	p	.000	.000	.036	.013	.000	
	ICC 95%CI upper	.958	.988	.974	.985	.998	
	ICC 95%CI lower	.337	.739	-.106	.183	.803	
ICC 0.80	n-trials	--	2	--	2	2	2.0 (.00)
ICC 0.85	n-trials	--	2	--	2	2	2.0 (.00)
ICC 0.90	n-trials	--	7	--	--	2	4.5 (3.5)

Notes: n-trials is the number of trials needed to reach the indicated ICC value; -- indicates that the ICC value was never achieved; p is the probability (rounded to three decimal places) that the ICC value is significantly different from zero; F1 and F2 are the first and second peak ground reaction force variables, respectively; F1LR and F2LR are the loading rates (average slope of the line) preceding these peaks; and IMP is the impulse for the first 100 ms of the landing; SD is the standard deviation.

Results

Ground reaction force-time histories and the values of the selected discrete variables normalized to body weight were typical of a vertical landing task (Table 1).

Relative to the first hypothesis, results indicated that performance stability of the ground reaction force variables was achieved after several nonconsecutive discrete landing trials. For the ICC analysis, the number of trials required for stability varied by group and variable (Tables 2 and 3). For the sequential averaging analysis, the number of trials required for stability varied by subject, variable, and number of reference trials used for the estimation (Tables 4 and 5).

For subject Group 1, the maximum ICC value was 0.98 for the IMP variable, with an ICC range of 0.71 to 0.98 for all variables (Table 2). All maximum ICC values were significantly ($p \leq 0.05$) different from zero and the number of trials needed to reach the maximum ICC value ranged from two to eight (mean 3.8 ± 2.7 trials; Table 2). The F1 and F1LR variables never achieved an ICC of 0.80, but F2 achieved an ICC of 0.85 after two trials and 0.90 after seven trials; F2LR achieved an ICC of 0.85 after two trials; and IMP reached an ICC of 0.90 after two trials (Table 2). For the sequential averaging analysis that used 20 reference trials and a 0.25 standard deviation criterion value, subject Group 1 exhibited a five to 17 trial range (mean 11.7 ± 3.1 trials) to achieve stability across all subjects and ground reaction force variables (Table 4). For the sequential averaging analysis that used 10 reference trials and a 0.25 standard deviation criterion value, subject Group 1 exhibited a five to 10 trial range (mean 7.9 ± 1.5 trials) to achieve stability across all subjects and ground reaction force variables (Table 5).

For subject Group 2, the maximum ICC value was 0.99 for the IMP variable, with an ICC range of 0.55 to 0.99 for all variables (Table 3). All maximum ICC values were significantly ($p \leq 0.05$) different from zero and the number of trials needed to reach the maximum ICC value ranged from two to six (mean 3.6 ± 1.7 trials; Table 3). The F2 and F2LR variables never achieved an ICC of 0.80, but F1 achieved an ICC of 0.90 after two trials; F1LR achieved an ICC of 0.80 after two trials, 0.85 after three trials, and 0.90 after four trials; and IMP reached an ICC of 0.90 after two trials (Table 3). For the sequential averaging analysis that used 20 reference trials and a 0.25 standard deviation criterion value, subject Group 2 exhibited a three to 17 trial range (mean 11.5 ± 4.5 trials) to achieve stability across all subjects and ground reaction force variables (Table 4). For the sequential averaging analysis that used 10 reference trials and a 0.25 standard deviation criterion value, subject Group 2 exhibited a three to 9 trial range (mean 6.6 ± 1.9 trials) to achieve stability across all subjects and ground reaction force variables (Table 5).

Relative to the second hypothesis, results indicated that the number of trials necessary to achieve performance stability of the ground reaction force variables differed between analysis methods. Fewer trials were required for stability using the ICC analysis when compared to the sequential averaging analysis (Tables 2, 3, 4 and 5). On average, the ICC analysis required 3.7 ± 2.1 trials for stability when data were collapsed across variables and subject groups (Tables 2 and 3). Conversely, the sequential averaging analysis required 11.6 ± 1.2 trials for stability when using 20 reference trials and the 0.25 standard deviation criterion value (Table 4) and 7.2 ± 0.9 trials when using 10 reference trials and the 0.25 standard

Table 3. **Summary of ICC analyses for subject Group 2 (drop landing).**

ICC Value	Statistic	F1	F2	F1LR	F2LR	IMP	Mean (SD)
ICC Maximum	n-trials	2	4	6	4	2	3.6 (1.7)
	ICC	.97	.76	.91	.55	.99	
	p	.001	.000	.000	.008	.000	
	ICC 95%CI upper	.997	.967	.988	.926	.999	
	ICC 95%CI lower	.755	.366	.723	.095	.928	
ICC 0.80	n-trials	2	--	2	--	2	2.0 (.00)
ICC 0.85	n-trials	2	--	3	--	2	2.3 (.6)
ICC 0.90	n-trials	2	--	4	--	2	2.7 (1.2)

Notes: n-trials is the number of trials needed to reach the indicated ICC value; -- indicates that the ICC value was never achieved; p is the probability (rounded to three decimal places) that the ICC value is significantly different from zero; F1 and F2 are the first and second peak ground reaction force variables, respectively; F1LR and F2LR are the loading rates (average slope of the line) preceding these peaks; and IMP is the impulse for the first 100 ms of the landing; SD is the standard deviation.

Table 4. Results of the sequential averaging analysis using 20 reference trials and a 0.25 standard deviation criterion value.

Subject Group	Subject Number	F1	F2	F1LR	F2LR	IMP	Mean (SD)
1	1	16	14	11	15	17	14.6 (2.3)
	2	12	16	7	16	17	13.6 (4.2)
	3	10	12	10	12	12	11.2 (1.1)
	4	8	9	10	9	5	8.2 (1.9)
	5	11	12	11	10	11	11.0 (0.7)
	Mean (SD)	11.4 (3.0)	12.6 (2.6)	9.8 (1.6)	12.4 (3.0)	12.4 (5.0)	11.7 (3.1)
2	6	4	14	15	14	11	11.6 (4.5)
	7	3	9	7	8	14	8.2 (4.0)
	8	15	3	16	5	16	11.0 (6.4)
	9	12	14	17	14	13	14.0 (1.9)
	10	14	17	7	15	11	12.8 (3.9)
	Mean (SD)	9.6 (5.7)	11.4 (5.5)	12.4 (5.0)	11.2 (4.4)	13.0 (2.1)	11.5 (4.5)

Notes: Values are the number of trials to reach stability. F1 and F2 are the first and second peak ground reaction force variables, respectively; F1LR and F2LR are the loading rates (average slope of the line) preceding these peaks; and IMP is the impulse for the first 100 ms of the landing; SD is the standard deviation.

deviation criterion value when data were averaged across subjects, variables, and groups (Table 5).

Further dissimilarities between stability analysis methods were revealed by the results of the sequential averaging analysis that used 10 reference trials and iteratively varied the standard deviation criterion value. These results indicated that standard deviation criterion values less stringent than 0.25 were required to approximate the ICC results. For Group 1, a standard deviation criterion value of 0.60 resulted in an average of 3.8 ± 1.8 trials to reach stability, compared to an average of 3.8 ± 2.7 trials for the ICC analysis. For Group 2, a standard deviation criterion value of 0.49 (rounded to two decimal places; actual value 0.4875) resulted in an average of 3.6 ± 2.1 trials to reach stability, compared to an average of 3.6 ± 1.7 trials for the ICC analysis.

Discussion

Performance stability is the test-retest reliability of a variable measured repeatedly over time. Stability is necessary for both the reliability of the data the ability to generalize to a greater population of trials. The number of trials obtained from an individual in an experiment is thought to influence stability (Bates et al., 1983; Salo et al., 1997) and thus is an important methodological consideration in the design of landing experiments. The number of trials needed for stability of ground reaction force variables during nonconsecutive landings has not been examined and is important for assessing the quality of previous landing studies and planning future studies. Therefore, one purpose of the study was to determine how many trials were necessary to achieve performance stability during landing. Another purpose was to compare two different methods of determining stability.

The first hypothesis was supported. Several trials were necessary to achieve performance stability of the selected ground reaction force variables during nonconsecutive discrete landing trials. Using the ICC method, four trials (mean 3.8 ± 2.7 trials for Group 1; 3.6 ± 1.7 trials for Group 2) were needed to achieve the maximum ICC values, which ranged from 0.55 to 0.99 across ground reaction force variables. Additionally, four of ten variables across the two subject groups failed to achieve an ICC value of 0.80, but only one variable (Group 2, F2LR; ICC = 0.55) failed to reach an ICC value 0.70. In comparison, the ICC value for the F2LR variable for Group 1 was 0.87. The results of the ICC analysis suggest that the test-retest reliability (stability) of landing trials is relatively strong for most of the selected ground reaction force variables and can be achieved within two to eight trials.

However, in contrast to the ICC analysis the sequential averaging technique suggested that as many as 12

Table 5. Results of the sequential averaging analysis using 10 reference trials and a 0.25 standard deviation criterion value.

Subject Group	Subject Number	F1	F2	F1LR	F2LR	IMP	Mean (SD)
1	1	6	8	6	7	8	7.0 (1.0)
	2	6	7	7	7	7	6.8 (.4)
	3	9	9	9	9	9	9.0 (.0)
	4	9	9	9	9	5	8.2 (1.8)
	5	8	10	10	10	5	8.6 (2.2)
	Mean (SD)	7.6 (1.5)	8.6 (1.1)	8.2 (1.6)	8.4 (1.3)	6.8 (1.8)	7.9 (1.5)
2	6	4	9	8	9	6	7.2 (2.2)
	7	7	9	7	9	9	8.2 (1.1)
	8	4	3	7	3	5	4.4 (1.7)
	9	8	7	6	5	7	6.6 (1.1)
	10	9	6	6	6	5	6.4 (1.5)
	Mean (SD)	6.4 (2.3)	6.8 (2.5)	6.8 (.8)	6.4 (2.6)	6.4 (1.7)	6.6 (1.9)

Notes: Values are the number of trials to reach stability. F1 and F2 are the first and second peak ground reaction force variables, respectively; F1LR and F2LR are the loading rates (average slope of the line) preceding these peaks; and IMP is the impulse for the first 100 ms of the landing; SD is the standard deviation.

nonconsecutive landing trials (11.7 ± 3.1 for Group 1; 11.5 ± 4.5 for Group 2) might be necessary to achieve performance stability when using 20 reference trials and a 0.25 standard deviation criterion value. Although landing, running, and walking were not compared statistically, results of the sequential averaging analysis from the current study suggest that more trials might be needed to achieve stability during landing than during either running or walking. In running, it was reported that eight nonconsecutive trials were necessary to achieve stability of the mean values of selected ground reaction force variables when using the 0.25 standard deviation criterion value and either 10 or 20 reference trials (Bates et al., 1983). During walking, stability of selected ground reaction force variables were reported following 10 nonconsecutive trials when using 20 reference trials and the 0.25 standard deviation criterion value (Hamill and McNiven, 1990). The current results for landing are similar to a previous report for vertical jumping, which used the sequential averaging technique (25 reference trials and a 0.30 standard deviation criterion value) to conclude that 12 trials were necessary to achieve performance stability in lower extremity joint kinetic parameters (Rodano and Squadrone, 2002). Collectively, results from the investigations which have used the sequential averaging technique suggest that 8-12 trials might be necessary to achieve performance stability in ground reaction force and lower extremity kinetic variables during various locomotor tasks, but results differ slightly among activities. Results from the current study suggest that the number of trials required for stability during landing is slightly greater than the number of trials reported for running and walking, but within the range of values reported for running, walking, and vertical jumping when using similar criterion values.

The second hypothesis also was supported. Different methods for determining stability provided dissimilar results. The ICC and sequential averaging methods were compared, but other methods for assessing reliability, stability, and variability as reported in the literature (Belli et al., 1995; Kadaba et al., 1989; Owings and Grabiner, 2003) were not examined. As previously stated, the ICC analysis on the current data revealed that an average of four trials were necessary to achieve stability, while 12 trials were needed using the sequential averaging technique when using the selected criteria. These values differ substantially from each other and the decision to follow one recommendation over the other could have important implications relative to the time and financial investment in an experiment. Logically, the best method would minimize the need to arbitrarily determine criteria for establishing stability, would be easy to implement, and would be familiar to most researchers. Therefore, the ICC method would seem to provide an objective means for determining performance stability. In comparing the two methods, the sequential averaging technique (using the selected values) would appear to provide a conservative estimate of the number of trials to achieve stability. Hamill and McNiven (1990) characterized 0.25 as a conservative standard deviation criterion value. Greater standard deviation criterion values would result in a fewer number of trials to achieve stability. In the current study, the standard deviation criterion value necessary to approximate the ICC results also was investigated. Using 10 reference trials, standard deviation criterion values of 0.60 (Group 1) and 0.49 (Group 2) provided results comparable to the ICC analysis. Therefore, the smaller standard deviation criterion values reported in the literature (e.g., 0.25, 0.30) appear to provide conservative estimates of stability.

While a primary result of the current study suggests that a minimum of four trials (ICC analysis) might be necessary to achieve performance stability during nonconsecutive landing trials, this result should be evaluated in context with the delimitations and limitations of the study. Delimitations included the age (22-31 yrs), sex (nine male, one female), and activity level (recreationally active) of the subjects. Because it is unknown how widely the current results can be generalized to subjects with different characteristics, generalization to different populations should be made with caution. Future research could examine if the current results are robust and applicable to subjects who have different characteristics.

The main limitations of the study involved factors associated with the landing task, variables selected for analysis, and interpretation of the ICC values. First, the landing task (0.60 and 0.61 m discrete landing) was not necessarily representative of landings that occur during most functional activities. The landing height was greater than typically would be performed. Additionally, in most functional activities a landing would be preceded by a jump and followed by another task such as a jump or a cutting maneuver thus potentially altering feed forward or feedback control of the landing. Another limitation was the absence of kinematic and joint kinetic variables in the analysis. While the aim of the current study was to examine the stability of ground reaction force variables, inclusion of kinematic and joint kinetic variables could have provided additional insight about the stability of trials during the selected landing task. Future research could address these issues. Finally, determining the number of trials necessary for stability using the ICC analysis is more objective than using the sequential averaging technique because it requires fewer arbitrary decisions. However, the interpretation of the quality of an ICC value also is arbitrary. Portney and Watkins suggested that reliability coefficients below 0.50 represent poor reliability, values between 0.50 and 0.75 suggest moderate reliability, and values above 0.75 indicate good reliability (Portney and Watkins, 2000). Furthermore, they suggest that these categories are arbitrary and that the reliability tolerance should be determined based on the precision of the measured variables and the application of the results (Portney and Watkins, 2000). In the current study, no ICC values were less than 0.50 (poor reliability), two ICC values were between 0.50 and 0.75 (moderate reliability), and eight ICC values were greater than 0.75 (good reliability). Moreover, of the eight ICC values greater than 0.75, five of these were greater than 0.90. Additionally, in the current study the number of trials needed for stability was based on the greatest ICC value observed and not one of the arbitrary criterion values or qualitative categories.

Conclusion

In conclusion, landing performance stabilized after several trials. Additionally, the methods for determining stability provided different results. An average of four trials was required for stability when using the ICC analysis and 12 trials was required when using the sequential averaging technique with the selected criteria. As determined by the ICC, some of the ground reaction force variables achieved their maximum test-retest reliability (stability) after only two trials, while other variables required six to eight trials each. Additionally, some variables never achieved an ICC of 0.80, regardless of the number of trials performed, while other variables achieved ICC values greater than 0.95. The ICC analysis provided a traditional statistical method for objective determination of stability, while the sequential averaging technique was more subjective, but conservative when using the criteria previously reported in the literature. Based on the current results, it is recommended that a minimum of four trials, and possibly as many as eight trials (the upper limit of the ICC analysis), should be obtained from each subject in a single session during an experiment involving 0.60 m drop or 0.61 m step-off landings. Additionally, researchers should be aware of the reliability of landing data in their investigations and could easily calculate the ICC using traditional methods and report these values post hoc.

References

Bates, B.T., Dufek, J.S. and Davis, H.P. (1992) The effect of trial size on statistical power. *Medicine and Science in Sports and Exercise* **24**, 1059-1065.

Bates, B.T., Osternig, L.R., Sawhill, J.A. and 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**, 181-191.

Belli, A., Lacour, J.R., Komi, P.V., Candau, R. and Denis, C. (1995) Mechanical step variability during treadmill running. *European Journal of Applied Physiology and Occupational Physiology* **70**, 510-517.

Griffin, L.Y., Agel, J., Albohm, M.J., Arendt, E.A., Dick, R.W., Garrett, W.E., Garrick, J.G., Hewett, T.E., Huston, L., Ireland, M.L., Johnson, R.J., Kibler, W.B., Lephart, S., Lewis, J.L., Lindenfeld, T.N., Mandelbaum, B.R., Marchak, P., Teitz, C.C. and Wojtys, E.M. (2000) Noncontact anterior cruciate ligament injuries: risk factors and prevention strategies. *Journal of the American Academy of Orthopaedic Surgeons* **8**, 141-150.

Hamill, J. and McNiven, L. (1990) Reliability of selected ground reaction force parameters during walking. *Human Movement Science* **9**, 117-131.

James, C.R., Dufek, J.S. and Bates, B.T. (2006) Effects of stretch shortening cycle exercise fatigue on stress fracture injury risk during landing. *Research Quarterly for Exercise and Sport* **77**, 1-13.

Kadaba, M.P., Ramakrishnan, H.K., Wootten, M.E., Gainey, J., Gorton, G. and Cochran, G.V. (1989) Repeatability of kinematic, kinetic, and electromyographic data in normal adult gait. *Journal of Orthopaedic Research* **7**, 849-860.

Owings, T.M. and Grabiner, M.D. (2003) Measuring step kinematic variability on an instrumented treadmill: how many steps are enough? *Journal of Biomechanics* **36**, 1215-1218.

Portney, L.G. and Watkins, M.P. (2000) *Foundations of clinical research: Applications to practice.* 2nd edition. Prentice-Hall, Upper Saddle River, NJ.

Rodano, R. and Squadrone, R. (2002) Stability of selected lower limb joint kinetic parameters during vertical jump. *Journal of Applied Biomechanics* **18**, 83-89.

Salo, A., Grimshaw, P.N. and Viitasalo, J.T. (1997) Reliability of variables in the kinematic analysis of sprint hurdles. *Medicine and Science in Sports and Exercise* **29**, 383-389.

Winter, D.A. (1984) Kinematic and kinetic patterns in human gait: Variability and compensating effects. *Human Movement Science* **3**, 51-76.

Key points

- The number of trials obtained from a subject in an experiment influences the stability (test-retest reliability) and thus validity of the data.
- One trial might not be representative of a subject's more general performance.
- Multiple-trial protocols have been recommended by several researchers for a variety of activities, but the number of trials necessary to achieve stability of ground reaction force variables during landing has not been examined.
- Researchers have used different criteria and methodologies for determining stability, making comparisons among studies and activities difficult.
- In the current study, test-retest intra-class correlation coefficient revealed that on average four trials were necessary for stability, while the more conservative sequential averaging analysis suggested that 12 trials were necessary for stability.
- Researchers should be aware of the stability of landing data and collect enough trials from each subject within a single testing session to maximize reliability of their data.

AUTHORS BIOGRAPHY

C. Roger JAMES

Employment

Director of the Center for Rehabilitation Research and Associate Professor in the Department of Rehabilitation Sciences in the School of Allied Health Sciences at Texas Tech University Health Sciences Center.

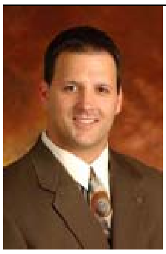
Degree

PhD

Research interests

Lower extremity biomechanics and injury mechanisms; experimental and statistical design issues in biomechanics research.

E-mail: roger.james@ttuhsc.edu



Joseph A. HERMAN

Employment

Regional Medical Practice Administrator with Abrazo Health Care located in Phoenix, AZ. Recently accepted a position as Chief Operations Officer with SMI Radiology located in Sun City West, AZ.

Degree

MS, MBA

Research interests

Biomechanics of human movement.

E-mail: jaherman@abrazohealth.com; jhermanbiomech@msn.com

Janet S. DUFEK

Employment

Visiting Research Associate Professor in the Department of Kinesiology in the School of Allied Health Sciences at the University of Nevada, Las Vegas.

Degree

PhD

Research interests

Lower extremity function, variability and human performance, single subject methodology.

E-mail: jdufek@unlv.nevada.edu

Barry T. BATES

Employment

Senior Scientist at Human Performance & Wellness, Inc. in Henderson, Nevada.

Degree

PhD

Research interests

Lower extremity function, human performance and forensic (personal injury) biomechanics.

E-mail: hpw@mail.com

✉ **C. Roger James, PhD**

Center for Rehabilitation Research, 3601 4th Street, MS 6223, Texas Tech University Health Sciences Center, Lubbock, Texas, 79430-6223 USA