Journal of Biomechanics 45 (2012) 2019–2022

Contents lists available at SciVerse ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech
www.JBiomech.com

Short communication

Within- and between-day reliability of trunk mechanical behaviors estimated using position-controlled perturbations

Brad D. Hendershot [a], Babak Bazrgari [b], Maury A. Nussbaum [a,c,*], Michael L. Madigan [a,d]

[a] Virginia Tech—Wake Forest School of Biomedical Engineering and Sciences, Virginia Tech, Blacksburg, VA 24061, USA
[b] Center for Biomedical Engineering, University of Kentucky, Lexington, KY 40506, USA
[c] Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA
[d] Engineering Science and Mechanics, Virginia Tech, Blacksburg, VA 24061, USA

ARTICLE INFO

Article history:
Accepted 13 May 2012

Keywords:
Reliability
Trunk
Stiffness
Reflex
Position perturbation

ABSTRACT

Recent applications of position-controlled perturbation techniques to the human trunk have allowed separate estimation of intrinsic and reflexive trunk mechanical behaviors. These mechanical behaviors play an important role in spinal stability and have been associated with low back pain risk, yet the reliability of these measures remains unknown. Therefore, the objective of the current study was to assess within- and between-day reliability of several measures of trunk mechanical behaviors obtained from position-controlled trunk perturbations. A secondary objective was to assess if different harness designs, used to connect a participant to the perturbing device, influenced reliability. Data were analyzed from baseline measurements obtained from two previously published studies, and a third unpublished study. The total combined subject pool included 33 healthy young adults (17 M, 16 F). Relative and absolute reliability was quantified using intraclass correlation coefficients (ICCs) and standard errors of measurement (SEM), respectively. Within-day ICCs of intrinsic trunk stiffness (0.84–0.90) and effective mass (0.91–95) were excellent, and were generally higher than ICCs for reflex gain (0.55–0.85), maximum reflex force (0.65–0.85), and timing of maximum reflex force (0.48–0.86). Within-day ICCs (0.48–0.95) were consistently superior to between-day values (0.19–0.72). Improvements in harness design increased both within- and between-day reliability and reduced SEMs for most measures.

1. Introduction

Intrinsic and reflexive mechanical behaviors of the human trunk play an important role in spinal stability and are associated with low back pain (LBP) risk (Panjabi, 2003; Moorhouse and Granata, 2007; Shin and Mirka, 2007). Collectively termed trunk mechanical behaviors (TMB) here, these behaviors are often estimated using sudden-loading perturbations (Brown and McGill, 2009; Hodges et al., 2009). Measures of TMB have been used to assess the effects of exposures to LBP risk factors (e.g., static or repetitive trunk flexion) and to characterize individual differences (e.g., young vs. old and healthy vs. LBP patients). Most existing sudden-loading methods are not suitable for distinguishing the relative contributions of intrinsic and reflexive behaviors to dynamic trunk stiffness, since a mechanical force is applied to the trunk with little constraint on response time or position. Position-controlled perturbations, however, are a relatively new method in their application to the trunk, and allow separate estimation of intrinsic and reflexive behaviors through the use of rapid and controlled perturbations (Moorhouse and Granata, 2007). Assessing the reliability of TMB obtained by this new method is therefore important to determine the clinical relevance and to support its use as an outcome measure. Given the lack of existing evidence regarding the reliability of such measures, the primary aim of this study was to evaluate the within- and between-day reliability of TMB derived from position-controlled trunk perturbations. As a secondary aim, reliability was compared between two harness designs that were used to connect a participant to the perturbing mechanism. Results are intended to facilitate the selection of measurement and analysis procedures in future studies.

2. Methods

2.1. Study design and procedures

Reliability analyses were performed on data obtained from two previous studies investigating the acute effects of prolonged trunk flexion exposures on TMB (Bazrgari et al., 2011; Hendershot et al., 2011). Additional unpublished data were obtained from a third study using similar methods. Prior to enrollment, all participants were given a practice session (on a separate day prior to testing),

* Corresponding author at: Department of Industrial and Systems Engineering, Virginia Tech, 250 Durham Hall (0118), Blacksburg, VA 24061, USA.
Tel.: +1 540 231 6053; fax: +1 540 231 3322.
E-mail address: nussbaum@vt.edu (M.A. Nussbaum).

0021-9290/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jbiomech.2012.05.026

during which they were exposed to all testing protocols, including strength measurements and maintaining a submaximal "preload" (see below). In each study, participants then completed six experimental sessions, with ≥ 3 days between consecutive sessions (mean=median=6 days; SD=3 days). Pre- and post-exposure measures of TMB were obtained using a position-controlled sudden perturbation paradigm. Only pre-exposure, baseline measurements were used in the current analysis, and which were obtained with participants in a consistent upright posture. Briefly, a pseudorandom sequence of 12 small (± 5 mm) anterior–posterior position perturbations was applied to the trunk ($\sim$T8) via a servomotor, rigid rod, and chest harness (see Bazrgari et al., 2011 and Hendershot et al., 2011 for a complete description). During a sequence, participants stood upright with the pelvis and lower extremities constrained, while maintaining a constant extensor effort or preload (10% of maximal trunk extensor muscle activity). Pseudorandom delays between each perturbation were used to prevent anticipation of perturbation timing and to reduce confounding from anticipatory changes in muscle activation. The total length of a sequence of 12 perturbations was $\sim$40 s, with each perturbation completed within 40 ms (which is less than typical erector spinae reflex delays). As such, the entire perturbation sequence generated 12 replications that are used for TMB estimation.

TMB were quantified by relating trunk kinematics to trunk kinetics during each perturbation pulse and by modeling the trunk as a single degree-of-freedom mass-spring-damper system (Hendershot et al., 2011). Five measures were obtained, using methods described previously (Bazrgari et al., 2011; Hendershot et al., 2011): intrinsic trunk stiffness, effective mass, reflex gain (relationship between reflex force and perturbation velocity), maximum reflex force, and timing of maximum reflex force (with respect to perturbation onset).

A combined total of 33 healthy young adults with no self-reported history of low back pain participated in the three studies. This included 17 males with mean (SD) age, stature, and body mass of 23 (3) yr, 180.5 (5.6) cm, and 73 (8.8) kg, respectively; and 16 females with corresponding values of 24 (3) yr, 165.3 (6.4) cm, and 59 (4.1) kg. All participants completed informed consent procedures approved by the Virginia Tech Institutional Review Board.

2.2. Data analysis

Within-day (i.e., within each sequence of 12 perturbations) and between-day (n=6) relative reliability of each measure was quantified using the intraclass correlation coefficient (ICC), while absolute reliability was quantified using the standard error of measurement (SEM). ICCs and SEMs were calculated according to a repeated-measures analysis of variance (ANOVA) model (Eliasziw et al., 1994), with day as a random effect. This model allows concurrent estimation of within-day and between-day reliability, which these authors indicated as providing a better estimate of reliability because both sources of error are estimated concurrently. For each ICC, a 95% one-sided lower-limit confidence interval was also calculated (Eliasziw et al., 1994). Comparisons were made to assess whether procedural changes influenced reliability. Specifically, results were compared for data obtained using a previous, wooden chest harness (Hendershot et al., 2011) vs. a newer, plastic version (Bazrgari et al., 2011) and unpublished data (Fig. 1).

Improvements in the harness design were intended to better (more widely) distribute the perturbation to the trunk and to improve comfort. Any resulting effects on reliability were unknown, though it was expected that the latter harness would provide more repeatable perturbations and hence more reliable measures of TMB. In addition, between-day reliability estimated using mean values (across all 12 perturbations in a sequence) was compared to the single "best" perturbation (i.e., having the highest correlation between measured and model-predicted responses). Note that for this comparison, in which a single measurement is used, the statistical model of Eliasziw et al. (1994) reduces to the Shrout and Fleiss (1979) case 2.

All statistical analyses were done using JMP (Version 9, SAS Institute Inc., Cary, NC, USA), with the level of significance set at p=0.05. ICC values were qualitatively interpreted using the following criteria (Cicchetti and Sparrow, 1981; Fleiss, 1986): 0.00–0.39=poor, 0.40–0.59=fair, 0.60–0.74=good, and 0.75–1.00=excellent.

3. Results

Within-day ICCs were excellent for all measures obtained with the new plastic harness, fair to excellent for the wooden harness, and highest in both cases for intrinsic stiffness and effective mass (Table 1). Between-day ICCs were fair to good for most measures obtained using the new harness, poor to good with the wooden harness, and were again highest for intrinsic trunk stiffness and effective mass (Table 1). Within- and between-day reliability of

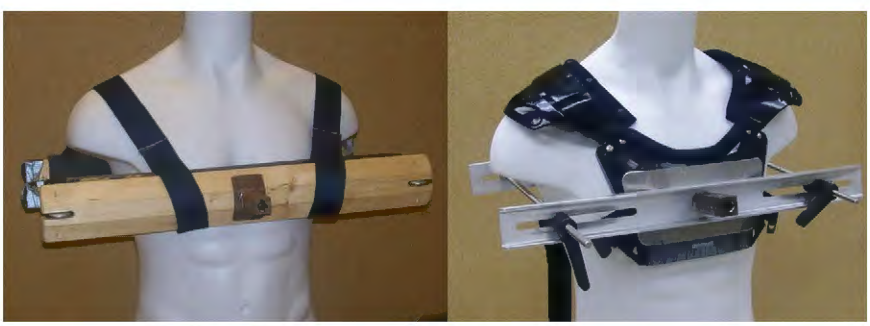

Fig. 1. (Left) Wooden chest harness and (right) plastic chest harness.

Table 1
Within- and between-day reliability of trunk mechanical behaviors using two harnesses.

	Within-day			Between-day	
	Mean	*ICC*	*SEM*	*ICC*	*SEM*
Wooden Harness (n=12)					
Intrinsic trunk stiffness (N/m)	4813.6	0.84 (0.79)	512.5	0.38 (0.32)	1519.8
Effective mass (kg)	20.2	0.95 (0.94)	0.6	0.72 (0.69)	2.1
Reflex gain (N s/m)	1066.4	0.55 (0.49)	117.2	0.19 (0.12)	314.5
Maximum reflex force (N)	220.7	0.65 (0.61)	20.6	0.22 (0.15)	53.7
Timing of max. reflex (ms)	148.2	0.48 (0.45)	5.1	0.27 (0.22)	7.9
Plastic Harness (n=21)					
Intrinsic trunk stiffness (N/m)	6406.6	0.90 (0.84)	464.6	0.67 (0.55)	963.9
Effective mass (kg)	17.7	0.91 (0.85)	0.8	0.54 (0.40)	2.2
Reflex gain (N s/m)	1270.0	0.85 (0.77)	109.2	0.37 (0.23)	264.5
Maximum reflex force (N)	210.9	0.85 (0.77)	15.8	0.36 (0.22)	48.4
Timing of max. reflex (ms)	156.0	0.86 (0.79)	2.6	0.46 (0.32)	6.9

Notes: n=number of participants. Mean values (across participants and sessions), intraclass correlation coefficients (ICC, with 95% one-sided lower-limit confidence intervals in parentheses), and standard errors of measurement (SEM) are provided. Units for each measure refer to means and SEMs.

Table 2
Between-day reliability comparing mean vs. best approaches (see text) for two harnesses.

	Mean		Best	
	ICC	*SEM*	*ICC*	*SEM*
Wooden Harness (n=12)				
Intrinsic trunk stiffness (N/m)	0.71 (0.51)	1810.1	0.61 (0.38)	1742.5
Effective mass (kg)	0.89 (0.79)	2.1	0.88 (0.76)	2.3
Reflex gain (N s/m)	0.37 (0.17)	312.4	0.34 (0.14)	326.1
Maximum reflex force (N)	0.44 (0.23)	52.6	0.40 (0.18)	58.1
Timing of max reflex (ms)	0.57 (0.36)	7.1	0.45 (0.23)	7.9
Plastic Harness (n=21)				
Intrinsic trunk stiffness (N/m)	0.81 (0.70)	908.8	0.82 (0.70)	912.1
Effective mass (kg)	0.69 (0.53)	2.2	0.67 (0.52)	2.4
Reflex gain (N s/m)	0.52 (0.33)	253.7	0.51 (0.32)	285.6
Maximum reflex force (N)	0.49 (0.32)	48.3	0.49 (0.33)	50.6
Timing of max reflex (ms)	0.62 (0.45)	6.8	0.59 (0.41)	8.1

Notes: n=number of participants. Intraclass correlation coefficients (ICC, with 95% one-sided lower-limit confidence intervals in parentheses) are given along with standard errors of measurement (SEM, with units given for each measure). Values here differ from those in Table 1 due to the use of a distinct ANOVA model.

the plastic harness was superior to the wooden harness in terms of ICCs and SEMs for all measures of TMB except effective mass. Between-day reliability (both ICCs and SEMs) was comparable using either mean or best values, and was again superior using the plastic harness for all measures except effective mass (Table 2). Relative reliability of effective mass was higher among males than females both within-day (ICC=0.92 vs. 0.65) and between-day (ICC=0.65 vs. 0.35). There were no substantial gender differences among the other reliability measures.

4. Discussion

Intrinsic trunk stiffness and effective mass were generally more reliable than the three reflexive measures (reflex gain, maximum reflex force, and timing of maximum reflex force). Intrinsic trunk stiffness and effective mass are estimated while the trunk is in motion (~40 ms), and are only affected by passive trunk properties and a fixed level of background muscle activity. Reflex gain, maximum reflex force, and timing of maximum reflex force measures, in contrast, all rely on excitation of stretch reflexes. Variability in reflex sensitivity (i.e., the detection threshold to applied strain within a muscle), or complete absence of a stretch reflex, may thus contribute to the decreased reliability of reflexive measures (Granata et al., 2004).

Within-day reliability was consistently superior to between-day reliability. Throughout the perturbation sequences, participants maintained a constant preload effort, with this preload based on strength measured on a given day. There was no evidence for learning effects in strength (p=0.11 from a one-way, repeated-measures ANOVA), and between-day reliability of the strength measures was excellent (ICC=0.87 (0.79), SEM=73 N). As noted earlier, all participants were given an initial practice session, and this practice likely reduced such learning effects. Learning was not expected within a given day, because each individual perturbation within a sequence was separated by pseudorandom delays to prevent anticipation of perturbation timing, and to reduce confounding from anticipatory changes in muscle activation. Despite this, variations in preload effort can influence estimated TMB (Stokes et al., 2000; Moorhouse and Granata, 2005), and could explain the poorer reliability between days. Further, because participants were provided visual feedback of extensor EMG to maintain the preload, variations in EMG preparations, as well as in muscle recruitment patterns, could also influence the between-day reliability.

Improvements in the harness design increased reliability, which was probably a result of changes in contact surface/pressure at the trunk-harness interface. The wooden harness (Hendershot et al., 2011) created a relatively small interface region, while the plastic harness (Bazrgari et al., 2011) had a more distributed contact area. This increased contact area facilitated more consistent harness placement, better distributed the perturbation to the trunk, and was more comfortable, all of which likely led to more consistent perturbations. Variability in the tightness of the harness, however, likely contributed to poorer between-day vs. within-day reliability (e.g., a looser harness may lead to decreases in TMB estimation). The plastic harness has adjustable physical stops that were used to ensure consistent harness tightness for a given participant, within a given experimental session. However, additional methods may be required for standardization across participants and between sessions, which would likely further improve between-day reliability.

Inferior reliability in trunk effective mass among females was likely related to additional and more variable soft tissue at the trunk-harness interface. Differences in the mean values between the two harnesses were also likely a result of the differences in contact areas. Finally, use of mean vs. best values yielded comparable relative and absolute reliability, supporting the use of either method.

Several limitations need to be considered when interpreting the current findings. Given the data that were available, formal within-day reliability was not assessed, but rather reliability within a given sequence of perturbations. However, actual within-day reliability is likely somewhere between the current within-sequence and between-day values. Similarly, between-day measurements were obtained across six sessions separated by ≥ 3 days. While participants returned at relatively consistent intervals and at similar times on separate days, it is presently unclear whether better control of inter-session intervals or the time of testing would improve between-day reliability, and the current results are probably conservative. Different subject samples were used to compare the reliability between the wooden and plastic harnesses. However, the two populations using each harness had similar characteristics ($<$5% difference in mean age, stature, and body mass), and all three studies were gender-balanced. Lastly, our reliability analyses were limited to a relatively young and healthy population, and reliability may be influenced by aging or musculoskeletal disability (e.g., LBP).

In summary, the results indicate that within-day reliability of TMB estimated using position-controlled trunk perturbations is excellent, while between-day reliability is poor to good depending on the specific measure. Intrinsic trunk stiffness and effective mass are generally more reliable than reflex gain, maximum reflex force, and timing of maximum reflex force. The mechanism used to connect a participant to the perturbing device can influence reliability. Specific positioning/tightening of such a device may influence between-day reliability, and methods for standardization may be of benefit in future work.

Conflict of interest statement

The authors declare no financial or personal relationships with other persons or organizations that might inappropriately influence our work presented herein.

Acknowledgments

This work was supported by an award (R01OH008504) from the Centers for Disease Control and Prevention (CDC). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the CDC.

References

Bazrgari, B., Hendershot, B., Muslim, K., Toosizadeh, N., Nussbaum, M.A., Madigan, M.L., 2011. Disturbance and recovery of trunk mechanical and neuromuscular behaviours following prolonged trunk flexion: influences of duration and external load on creep-induced effects. Ergonomics 54, 1043–1052.

Brown, S.H.M., McGill, S.M., 2009. The intrinsic stiffness of the in vivo lumbar spine in response to quick releases: Implications for reflexive requirements. Journal of Electromyography and Kinesiology 19, 727–736.

Cicchetti, D.V., Sparrow, S.S., 1981. Developing criteria for establishing inter-rater reliability of specific items: application to assessment of adaptive behavior. American Journal of Mental Deficiency 86, 127–137.

Eliasziw, M., Young, S.L., Woodbury, M.G., Fryday-Field, K., 1994. Statistical methodology for concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. Physical Therapy 74, 777–788.

Fleiss, J.L., 1986. The Design and Analysis of Clinical Experiments. Wiley, New York.

Granata, K.P., Slota, G.P., Bennett, B.C., 2004. Paraspinal muscle reflex dynamics. Journal of Biomechanics 37, 241–247.

Hendershot, B., Bazrgari, B., Muslim, K., Toosizadeh, N., Nussbaum, M.A., Madigan, M.L., 2011. Disturbance and recovery of trunk stiffness and reflexive muscle responses following prolonged trunk flexion: influences of flexion angle and duration. Clinical Biomechechanics 26, 250–256.

Hodges, P., van den Hoorn, W., Dawson, A., Cholewicki, J., 2009. Changes in the mechanical properties of the trunk in low back pain may be associated with recurrence. Journal of Biomechanics 42, 61–66.

Moorhouse, K.M., Granata, K.P., 2005. Trunk stiffness and dynamics during active extension exertions. Journal of Biomechanics 38, 2000–2007.

Moorhouse, K.M., Granata, K.P., 2007. Role of reflex dynamics in spinal stability: intrinsic muscle stiffness alone is insufficient for stability. Journal of Biomechanics 40, 1058–1065.

Panjabi, M.M., 2003. Clinical spinal instability and low back pain. Journal of Electromyography and Kinesiology 13, 371–379.

Shin, G., Mirka, G.A., 2007. An in vivo assessment of the low back response to prolonged flexion: interplay between active and passive tissues. Clinical Biomechechanics 22, 965–971.

Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 86, 420–428.

Stokes, I.A., Gardner-Morse, M., Henry, S.M., Badger, G.J., 2000. Decrease in trunk muscular response to perturbation with preactivation of lumbar spinal musculature. Spine 25, 1957–1964.