Sensor-based 9-week Serial Balance Data Show Need for Individualized Baseline Profiles: Implications on Concussion Diagnosis

D. Al-Mfarej[1], D. Gonzalez[2] and J. Tung[1]

[1]Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, Canada
[2]School of Health and Human Performance, Dalhousie University, Halifax, Canada

Abstract— **Objective: The ability to accurately identify concussions and assess recovery is essential to protect individuals from experiencing negative consequences regarding premature return-to-play. To date, there is no "gold" standard of concussion diagnosis nor method to track recovery. Instead, clinicians rely on symptom checklists and neuropsychiatric tests to inform clinical decisions. Balance is one of several commonly assessed motor capabilities to screen for concussion, including sensor-based assessments of balance using normative data as baseline to track changes. However, the timing and frequency of balance measurements to screen for impairment and monitor recovery remains underexamined. This study examines the utility of a rapid (5-min) sensor-based balance measurement on a habitual (i.e., pre/post-practice) schedule to screen for concussions. The primary objective of the study is to determine the factors that affect baseline sway measures. Factors considered include individual differences, assessment timing (pre/post activity), balance condition, and longitudinal changes. Additionally, this study examines whether normative data gathered from a large sample of healthy controls are reflective of individual balance profiles. Design: A pilot study using a repeated observation design. Methods: Five varsity hockey players (3 males, 2 females) were recruited for a 9-week study. Each athlete was tested prior to and after practice using an IMU, performing a modified Balance Error Scoring System (BESS) test. Results: Sampled data used to estimate individual beta distributions indicates significant individual differences in balance behaviour across a range of metrics. Additionally, preliminary results show that normative values drawn from a large cross-sectional sample are not reflective of individual balance profiles drawn from our longitudinal sample. Conclusions: This study supports the need for individualized baseline profiles for balance in order to achieve higher accuracy and sensitivity in concussion detection. Serial, habitual testing is recommended to enable concussion detection from objective measures with higher accuracy and sensitivity during sideline assessments.**

Keywords— **Individualized baselines, concussion, sideline assessment, balance assessment, serial data**

I. Introduction

Objectively determining the occurrence of concussion, understanding the possible consequences thereafter, and evaluating recovery from the injury are top priorities for concussion research. This is a difficult task for medical professionals due to the lack of a gold-standard tests and reliance on subjective symptom scores. Providers are left with limited data upon which to make decisions about diagnoses, and return to play [1]. Gathering timely data to inform whether a person can safely return to the game can have a major impact on recovery: athletes continuing to play after a suspected concussion can have prolonged recovery compared to those immediately removed [2,3]. Therefore, providing clinicians with sensitive and objective tools that can determine if someone has suffered a concussion, and monitor recovery following the injury, is essential for the health and safety of individual athletes.

Current screening, diagnosis, and assessment of sport related concussion is conducted using one of 3 approaches: 1) monitoring symptoms and their resolution, 2) comparing patient performance on specific tests to normative data, and 3) comparing patient performance to a pre-injury "baseline". The most common approach is on self- and observer-based symptom monitoring. Considering symptoms are highly subjective and often influenced by circumstances, symptom self-report has been plagued by inconsistencies [4]. An alternative is to apply a nomothetic approach by using normative values based on sampling comparable populations. However, applying normative values to individuals also has drawbacks, highlighted by Capitani [5]. For example, is it reasonable to compare a 29- and a 21-year-old because they are in the 20-29 age range? Would it be more practical to compare a 29- to a 31-year-old? To account for individual factors, many clinics run a model that includes individual baselines, popularized by using computerized software, such as ImPACT [6]. Cost and feasibility of widespread adoption of individual baselines remains a challenge [7], especially in repeated testing [1].

A more economical approach is to use readily available tools such as the SCAT-5 (Sport Concussion Assessment Tool v5). Although not the intended use (as stated in the SCAT-5 manual), there is some evidence to suggest that such a tool may warrant use for baseline testing [8]. However, subjectivity in self- and observer- based reporting remain concerning. An amalgamation of assessment domains, such as balance (mBESS; modified Balance Error Scoring System), cognition (SAC; Standardised Assessment of Concussion), symptoms checklists, and other questions (e.g., the Glascow Coma Scale), the SCAT-5 is used on the sidelines to indicate whether further medical attention is warranted. Accurate administration of these tools (e.g., BESS) requires trained individuals to make complex judgements. Decision-makers need knowledge and objective tools to make informed decisions, as opposed to sending athletes to the emergency room with every hit that occurs.

Senor-based technology may provide the means to improve concussion assessment, including mounting evidence supporting the utility of balance assessment to evaluate patients with concussion [1,9]. In a cross-sectional study, King et al. utilized wearable sensors for assessment of sway metrics capable of distinguishing impaired balance due to concussion [10].

The present study proposes an alternative assessment paradigm utilizing sensor-based sway measurements to evaluate an athlete's balance immediately prior to and following a sporting event. Assessing balance on a routine basis will permit personalized comparisons in the event of a suspected injury. The primary research question asks, "Can normative condition-specific sway ranges be used when screening for concussion?" The hypothesis of this pilot study was that objective, sensor-based testing of balance prior to (and following) sport participation over a 9-week period will show significant individual variations, indicating the need for individual specific sway baselines.

II. Methodology

A. Data Collection

Five volunteers (3 males, 2 females; average age: 22.8 +/- 3.7 years) from the men's and women's varsity hockey teams at the University of Waterloo with no history of previous concussion, brain injury, or balance impairment were recruited. Following pilot tests, 5 was chosen as a maximum feasible sample size for a researcher (DG) to conduct while minimizing the length of time required to sustain compliance. Following written informed consent, each participant required ~5 minutes per session. All procedures were approved by the University of Waterloo's Office of Research Ethics.

Based on the modified BESS (mBESS), participants stood with eyes closed on a firm surface in 3 conditions: i) double leg stance, ii) single leg stance on their non-dominant foot, and iii) tandem stance with their non-dominant foot at back, for twenty seconds in each position. Shimmer3 ExG sensors (Shimmer, Ireland), including 3D accelerometer, gyroscope, and magnetometer, were attached to the sternum to provide objective measures of postural sway.

A researcher (DG) attended one practice per week for each of the men's and women's varsity hockey team. Testing was conducted immediately before and after practice, once per week for nine consecutive weeks. In total, 90 testing sessions were planned. Participants were tested in random order. Over the study period, 33-51 independent samples were obtained for each participant.

B. Data Processing

MATLAB (version 2020a, Mathworks Inc., Natick, USA) was used for data processing and analysis. 3D raw sensor data were transformed to obtain accelerations in the medial-lateral (ML), anterior-posterior (AP), and vertical (VT) directions. The data was then filtered using a 3rd order low-pass Butterworth filter with a cut-off frequency of 3.5 Hz. Drawing on King et al., the five most sensitive metrics for concussion sway assessment were selected, all of which were in the ML direction [10]: root mean square (RMS), total power, range of acceleration (RoA), path length (PL), and mean distance (MD).

C. Hypothesis Testing

A linear mixed effects model (LME) was fit to the data for hypothesis testing. Since the data is comprised of repeated measures from 5 participants, it exhibits non-independencies. The LME model controls the random variability between groups around the fixed effects, accounting for non-independencies in the data. LME models also produce fairly unbiased predictions when data points are missing completely at random (MCAR), which is the case for our dataset.

The response variable in the model was the respective sway metric, with all five sway metrics modelled separately. The fixed effects were test condition, pre/post practice, week number, and their two-way interactions. Random intercepts and slopes with participant as the grouping variable were introduced. The data was initially log transformed, and normality of the response variable was inspected visually. Model fit statistics like the Akaike information criterion (AIC), Bayesian information criterion (BIC), and adjusted R-squared were calculated to ensure optimal model selection. Homoscedasticity and normality of the error terms, normality of random effects, as well as linearity of the relationship were assessed visually and considered acceptable. It is recommended that the random effects factor has a minimum of 5 groups [11]. While on the lower bound of the advised range, the dataset is acceptable for preliminary study.

III. Results

Following model fitting, analysis of variance for LME models was performed. The p-values for the fixed effect terms are shown in Table 1 below to indicate significant effects on the response, even if for a single participant or condition. The condition variable showed a statistically significant p-value for all five sway metrics. All other fixed effect terms, including their interaction terms, showed no significant effect on the sway metrics performance.

Table 1: p-values for the fixed effect terms of the LME model. * indicates p<0.05

	RMS	Power	RoA	PL	MD
Condition	0.001*	0.004*	<0.001*	0.003*	0.003*
Pre/Post	0.278	0.271	0.346	0.241	0.204
Week	0.668	0.673	0.799	0.659	0.701
Condition:Pre/Post	0.819	0.497	0.905	0.667	0.625
Condition:Week	0.419	0.558	0.124	0.500	0.500
Pre/Post:Week	0.168	0.111	0.583	0.094	0.090

Descriptive statistics are shown in Table 2, including RMS median and IQR range measures for all 5 participants, and 3 conditions, for pre-practice sessions.

Table 2: RMS median and IQR ranges (m/s^2) grouped by the participants and conditions, for pre- practice session.

	Condition 1	Condition 2	Condition 3
1	1.06E-02 (4.04E-03)	2.87E-02 (1.34E-02)	2.36E-02 (3.07E-03)
2	4.51E-02 (2.55E-02)	7.03E-02 (2.48E-02)	6.19E-02 (9.99E-03)
3	9.40E-03 (3.39E-03)	4.67E-02 (4.13E-02)	3.31E-02 (1.88E-02)
4	7.80E-03 (3.60E-03)	1.48E-02 (2.18E-02)	5.73E-02 (3.00E-02)
5	6.83E-03 (9.73E-04)	1.16E-02 (6.78E-03)	1.95E-02 (4.88E-03)

Including the participant random effects increased adjusted R^2 value from 0.26 to 0.63, and lead to reduction in AIC from 160 to 53. To formally test the significance of the random effects, LME models with and without the random intercept and slope were compared using theoretical likelihood ratio test. For all 5 metrics analyzed, theoretical likelihood ratio test p-values were $p<0.001$ indicating that the introduction of the random intercept and slope significantly improves the model fit. AIC values confirm significant differences in model fit.

A. Building Individualized Balance Profiles

To model each participant's individualized baseline balance profiles, a probability density function (PDF) was fit to their RMS data. The data for condition 1 was used and pre/post data was combined as our previous analysis showed no significant difference in performance for the variable. As such, a maximum of 18 datapoints were available for each participant. Bootstrap sampling using the means was then applied to obtain 50 samples per participant.

The goal was to fit a univariate, parametric, bounded PDF to the data for each athlete. The beta distribution was selected, and goodness of fit was assessed using a Q-Q plot and comparison of cumulative distribution functions. As it can flexibly take on a range of shapes, the beta distribution was chosen to model individual differences. Using the maximum likelihood method, fitted PDFs for all participants were generated and plotted in Figure 1. The dashed black line and grey shaded area on the figure illustrate published normative data for the median and IQR range for the RMS ML values of 76 athletes with no concussion [10]. As a side note, the sensor was placed on the lumbar spine (versus the sternum in the current study). Using the inverted pendulum theory, placing the sensor on the sternum should yield higher accelerations than the lumbar spine due to a longer moment arm.

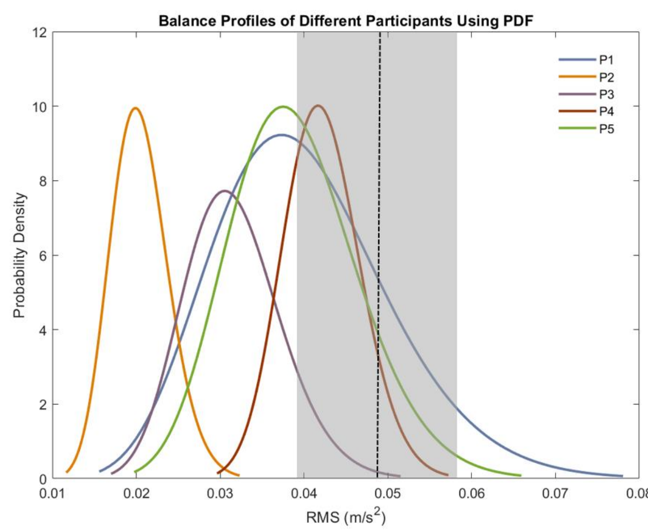

Figure 1: Baseline balance profiles of the 5 participants. Dashed black line and grey shaded area on the figure illustrate published normative data for the median and IQR range for the RMS ML values of 76 athletes with no history of concussion.

IV. Discussion

Statistical analysis of the fixed effects in the model, as shown in Table 1, indicates the time of testing (i.e., pre vs post practice), and the week number do not play significant roles on the participant's sway. The observed significance of test condition (i.e., double leg vs single-leg vs tandem) is already well known. Analysis of the random intercepts in the model indicated significantly different individual baselines, largely attributable to participant 2.

The constructed baseline profiles of the different participants (Figure 1) exhibit differences in the width, location, and shape of the individual distributions. Considering the distribution for Participant 2 (P2, yellow trace) does not overlap with normative ranges (black dashed line and shaded region), using normative baselines to determine impairment may be inappropriate.

Compared to published normative ranges [10], the 5 participants in the current study demonstrated smaller RMS on average. P2 demonstrated small sway magnitudes well below the normative distribution, but very consistently across 9 weeks of testing. If the normative range (shaded region in Figure 1) represents the healthy RMS range, a concussion resulting in balance deficits in participants P2 and P3 is more likely go undetected due to their low RMS baselines. These findings suggest individual baselines are needed to account for heterogeneity, especially given how narrow the screening ranges are and the sensitivity required for reliable diagnosis.

One advantage of profiling sway as a PDF is estimating the probability of a measurement being within their individual unimpaired range. In other words, replacing hard Boolean (yes/no) decisions with soft clusters capable of outputting the probability of a participant being injured may be a more effective approach for objective concussion screening. However, this requires serial baseline data. A potential next step that extends from this idea is the utilization of Bayesian online learning wherein the distribution of the sway baseline profile is updated given new data points as data is collected serially over time. This would allow for an increasingly improved representation of participant profiles over time as more data is collected.

V. Conclusions

Individual heterogeneity in sway measures must be accounted for in order to effectively detect concussion using objective balance tests. One way to account for individual differences is by performing habitual testing in a serial fashion to build personalized balance baseline profiles. Such a method has the potential to reduce missed concussion diagnoses, thereby reducing inappropriate return to play decisions.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. McCrory P, Meeuwisse W, Aubry M, et al. Consensus statement on concussion in sport: the 4th International Conference on Concussion in Sport held in Zurich, November 2012. *Br J Sports Med*. 2013;47(5):250-258.
2. Asken BM, McCrea MA, Clugston JR, Snyder AR, Houck ZM, Bauer RM. Playing through it: Delayed reporting and removal from athletic activity after concussion predicts prolonged recovery. *J Athl Train*. 2016;51(4):329-335. doi:10.4085/1062-6050-51.5.02
3. Elbin RJ, Sufrinko A, Schatz P, et al. Removal From Play After Concussion and Recovery Time. *Pediatrics*. 2016;138(3):e20160910-e20160910. doi:10.1542/peds.2016-0910
4. Chrisman S, Quitiquit C, Health FR-J of A, 2013. Qualitative study of barriers to concussive symptom reporting in high school athletics. *Elsevier*.
5. Capitani E. Normative data and neuropsychological assessment. Common problems in clinical practice and research. *Neuropsychol Rehabil*. 1997;7(4):295-310.
6. Schatz P. Long-term test-retest reliability of baseline cognitive assessments using ImPACT. Am J Sports Med.2010;38(1):47-53. doi:10.1177/0363546509343805
7. Kirkwood MW, Randolph C, Yeates KO. Returning pediatric athletes to play after concussion: The evidence (or lack thereof) behind baseline neuropsychological testing. *Acta Paediatr Int J Paediatr*. 2009;98(9):1409-1411. doi:10.1111/j.1651-2227.2009.01448.x
8. Putukian M, Echemendia R, Dettwiler-Danspeckgruber A, et al. Prospective clinical assessment using sideline concussion assessment tool-2 testing in the evaluation of sport-related concussion in college athletes. *Clin J Sport Med*.2015;25(1):36-42. doi:10.1097/JSM.0000000000000102
9. Baker CS, Cinelli ME. Visuomotor deficits during locomotion in previously concussed athletes 30 or more days following return to play. *Physiol Rep*. 2014;2(12):e12252. doi:10.14814/phy2.12252
10. King, L. A., Mancini, M., Fino, P. C., Chesnutt, J., Swanson, C. W., Markwardt, S., & Chapman, J. C. (2017). Sensor-Based Balance Measures Outperform Modified Balance Error Scoring System in Identifying Acute Concussion. *Annals of biomedical engineering*, *45*(9), 2135–2145. https://doi.org/10.1007/s10439-017-1856-y
11. 11. Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, *6*, e4794. https://doi.org/10.7717/peerj.4794