Laboratory Validation of Two Wearable Sensor Systems for Measuring Head Impact Severity in Football Players

GUNTER P. SIEGMUND [ID],[1,2] KEVIN M. GUSKIEWICZ,[3,4] STEPHEN W. MARSHALL,[3,4] ALYSSA L. DEMARCO,[1] and STEPHANIE J. BONIN[5]

[1]MEA Forensic Engineers & Scientists, 11-11151 Horseshoe Way, Richmond, BC V7A 4S5, Canada; [2]School of Kinesiology, University of British Columbia, Vancouver, BC, Canada; [3]Department of Exercise and Sport Science, Matthew A. Gfeller Sport-Related Traumatic Brain Injury Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; [4]Injury Prevention Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; and [5]MEA Forensic Engineers & Scientists, Laguna Hills, CA, USA

Associate Editors Dan Elson and Stefan Duma oversaw the review of this article.

Abstract—Wearable sensors can measure head impact frequency and magnitude in football players. Our goal was to quantify the impact detection rate and validity of the direction and peak kinematics of two wearable sensors: a helmet system (HITS) and a mouthguard system (X2). Using a linear impactor, modified Hybrid-III headform and one helmet model, we conducted 16 impacts for each system at 12 helmet sites and 5 speeds (3.6–11.2 m/s) ($N = 896$ tests). Peak linear and angular accelerations (PLA, PAA), head injury criteria (HIC) and impact directions from each device were compared to reference sensors in the headform. Both sensors detected ~96% of impacts. Median angular errors for impact directions were 34° for HITS and 16° for X2. PLA, PAA and HIC were simultaneously valid at 2 sites for HITS (side, oblique) and one site for X2 (side). At least one kinematic parameter was valid at 2 and 7 other sites for HITS and X2 respectively. Median relative errors for PLA were 7% for HITS and -7% for X2. Although sensor validity may differ for other helmets and headforms, our analyses show that data generated by these two sensors need careful interpretation.

Keywords—Head impact, Wearable sensors, Validity, Acceleration, Concussion, Football.

INTRODUCTION

Head impacts in football can cause concussions.[10,13,25,32,35] Research on the mechanics of these head impacts can improve player safety through rule changes, behavior changes, better equipment, and

Address correspondence to Gunter P. Siegmund, MEA Forensic Engineers & Scientists, 11-11151 Horseshoe Way, Richmond, BC V7A 4S5, Canada. Electronic mail: gunter.siegmund@meaforensic.com

better on-field identification of potentially injurious impacts. To achieve these goals, valid methods of measuring the direction and severity of on-field head impacts are needed.

Concussion is a transient disturbance of brain function induced by biomechanical forces from a direct or indirect blow to the head.[16] Impact-induced head motion generates neural tissue strains[3,15] that increase with greater impact severity.[19,40,43,45] Strain magnitude also varies with impact direction,[40,43] likely because of the shape of the brain/cranial cavity and stress concentrations created by structures like the blood vessels and dural folds.[12,30,40] If high enough, these neural tissue strains can cause transient changes in neuronal function.[2,11,22] Helmet design, shape, and construction affect head kinematics[37] and thus can also affect strain magnitude.

The magnitude and direction of impact-induced head motions are determined by the magnitude, direction and location of both the external force applied to the helmet and the internal reaction forces and moments applied by the neck to the base of skull.[41] Although these reactions are hard to measure, the resulting linear and angular head accelerations uniquely describe their net effect on the head, and thus are useful for quantifying concussion risk.

Wearable devices exist for measuring the head's linear and angular acceleration during an impact. These devices vary in design, but typically consist of multiple accelerometers and/or angular rate sensors built into or adhered to a helmet, mouthguard, earplug, skullcap, chinstrap or skin.[5,9,17,21,29,35] Validation data for some of these devices exist,[1,4,5,18,24,35] but each

of these studies had one or more of the following limitations: the helmet was smaller than recommended by the helmet manufacturer; only 4–6 primarily centroidal impact sites were used; some impact sites associated with concussion (e.g., crown)[14,36] were not tested; impact speeds were below the average associated with concussion (9.3 ± 1.9 m/s)[32]; the number of repeated tests was relatively low and only one sample of the device was tested; and the analyses generally focused on amplitude data pooled across multiple impact sites. Our goal was to quantify the validity of two wearable devices, a helmet-based system and a mouthguard-based system, using a wide range of impact sites and speeds, and multiple samples of each device. In addition to assessing the validity of both the direction and amplitude reported by the devices, we also quantified the number of times the sensors missed impacts. For our purposes, validity is defined as the degree to which a test system measures what the reference system measures.

METHODS

Two wearable sensors that were commercially available at the time and made direct contact with the head were tested: the Head Impact Telemetry System (HITS, Simbex, Lebanon, NH; Riddell, Rosemont, IL) and the X2 mouthguard (X2 Biosystems, Seattle, WA). HITS has 6 linear accelerometers within an MxEncoder (mass 256 g) that fit into a football helmet liner and computes head impact kinematics.[4,7] For this study, HITS was installed into Riddell Revolution Speed helmets (size large) fitted with a facemask (Riddell model 94914LW-V) and chinstrap (Riddell model R45747 3P). Total helmet mass including HITS was 2.07 kg. HITS computes various parameters, including peak resultant linear acceleration (PLA) and angular acceleration (PAA) of the head's center of gravity (COG), the head injury criterion (HIC), and the azimuth and elevation of the impact site. HITS does not measure angular acceleration about the vertical axis and estimates angular acceleration about the horizontal axes from the horizontal linear acceleration components and an assumed rotation point 0.1543 m below the head COG.[36] A 40 ms, 1 kHz trace of the resultant linear acceleration can be exported.

The X2 system has a 3-axis linear accelerometer and a 3-axis angular rate sensor embedded in a plastic mouthguard (mass 21.5 g). Mouthguards were custom-fit to the headform's upper dentition using the same vacuum-forming process used for fitting players. Accelerations are sampled at 1 kHz, whereas angular rates are sampled at 800 Hz, interpolated to 1 kHz, filtered, and differentiated to generate angular acceleration.

Linear acceleration is resolved to the head's COG using rigid body kinematics. X2 computes various parameters, including PLA, PAA, HIC and the azimuth and elevation of the direction of PLA. It provides 100 ms traces for each linear and angular sensor channel. This specific version of the X2 mouthguard is no longer commercially available; however, we understand that X2 and other companies are developing similar mouthguards and therefore these tests provide general insight into how these devices can perform.

Both devices were set to trigger when any linear accelerometer exceeded 10 g. Both devices wirelessly transmit data to dedicated base stations, display it on a computer screen, and then upload it to dedicated central servers for backup and cleansing using proprietary algorithms. In this study we focused exclusively on a cleansed data set. The parameters generated by HITS and the X2 were compared to those calculated from a 3-2-2-2 array of linear accelerometers (Endevco 7264B-2000 g, San Juan Capistrano, CA) installed in a compact cluster ($r_x = r_y = 34$ mm, $r_z = 27$ mm) in a modified load-sensing headform (MLSH) based on the 50th percentile male Hybrid III headform.[42,44] Accelerometer data were acquired at 10 kHz with hardware anti-aliasing filters prior to digitization (SAE Channel Class 1000).

Impact tests were performed on a horizontal linear impactor (Biokinetics and Associates Ltd., Ottawa, ON). The MLSH was mounted on a 50th percentile male Hybrid III neck, which was mounted to a sliding table (combined head/neck/table mass 23.5 kg). The helmeted headform was struck by a 14 kg impactor fitted with a 36 mm thick, 127 mm diameter vinyl nitrile pad and a rigid plastic cap with a spherical radius of 127 mm to simulate the striking player and helmet.[34] Helmets were fit to the MLSH according to the manufacturer's instructions over 2 layers of nylons. The crown and side/back liner pads were vented to atmosphere and then re-inflated to ~14 kPa (2 psi) prior to each test. Large cheek pads (32 mm thick compared to the regular 25 mm thickness) were used to account for the narrower cheeks of the Hybrid III headform compared to the NOCSAE headform[6] and these pads remained at atmospheric pressure. The MLSH's mandible was designed to more closely match the normal human chin location,[44] thereby correcting the jaw differences between the Hybrid III and NOCSAE headforms.[6] X2 mouthguards were fit over the upper dentition and held in place by the chinstrap force transmitted through the mandible. The mandible was not further clamped or restricted from its side-to-side or protraction/retraction range of motion.[44] No fluid was added at the mouthguard/dentition interface to simulate saliva. A custom jig was used to identically
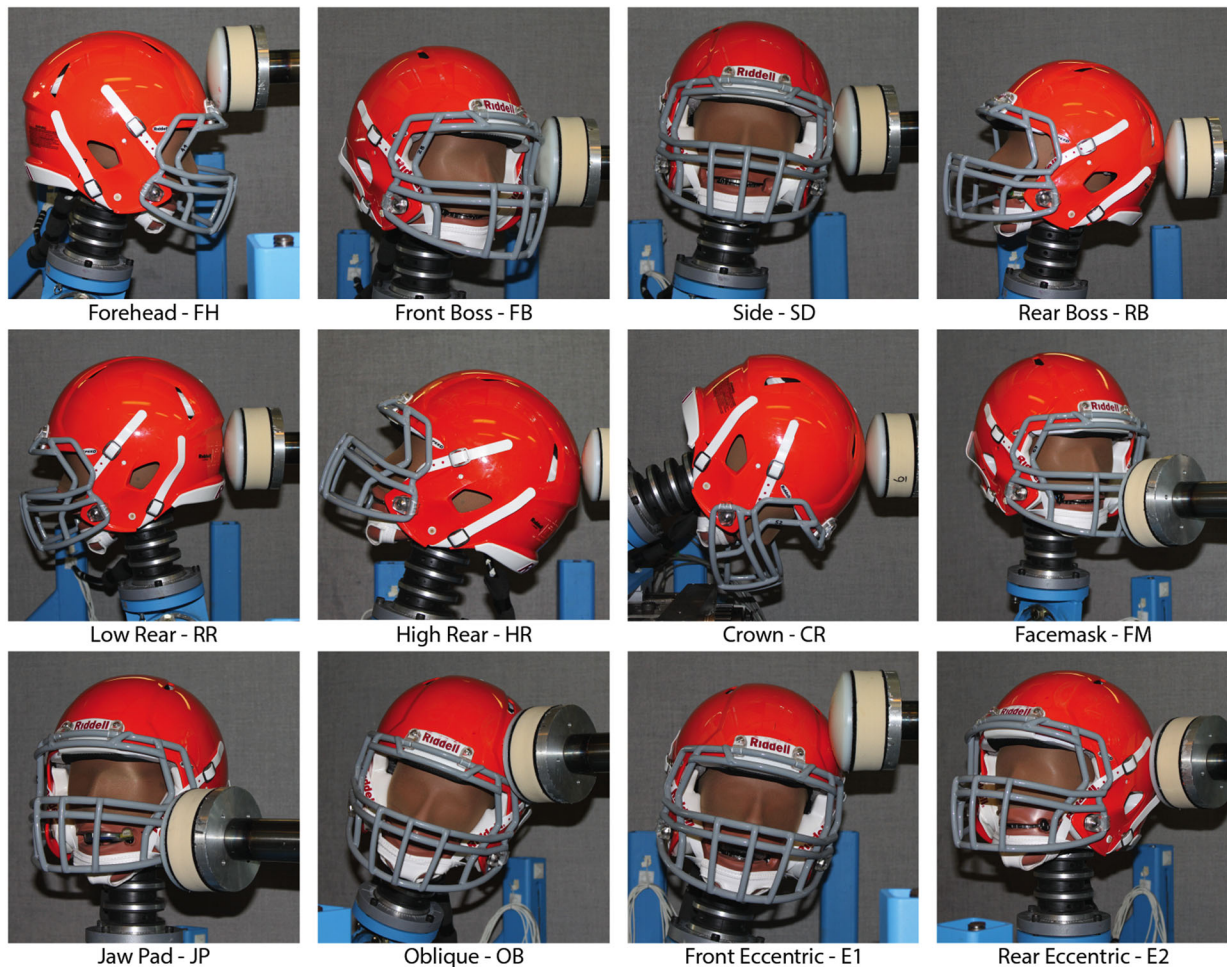
FIGURE 1. Photographs of the twelve impact sites.

set all chinstraps to a length that provided a snug fit. When fully installed, inflated and positioned relative to the nose using an index gage, the helmet did not move relative to the headform when the facemask was vigorously pulled sideways. During testing, the helmet came off during one 11.2 m/s facemask impact. None of the mouthguards were dislodged from the upper dentition during testing.

Twelve helmet sites were tested (Fig. 1): forehead (FH), front boss (FB), side (SD), rear boss (RB), low rear (RR), high rear (HR), crown (CR), facemask (FM), jaw pad (JP), oblique (OB) and two eccentric sites (E1, E2). The impactor initially struck the facemask at the FM, FB and JP sites. Ten impacts were nominally centroidal, i.e., the impactor's axis passed near a vertical axis through the headform's COG. For the E1 and E2 sites, the impactor struck the helmet in the mediolateral direction ahead (E1) and behind (E2) the headform's COG. Five impact speeds were tested: 3.6, 5.5, 7.4, 9.3 and 11.2 m/s. These represent the mean, ± 1, -2 and -3 standard deviations of the

closing speed for reconstructed NFL impacts $(9.3 \pm 1.9$ m/s).[41] Only the 4 lower speeds were tested at CR (to limit axial neck compression) and JP (to limit a direct blow to the mandible). Initial tests to HR at 9.3 m/s broke the HITS sensors, and thus only the 3 lower speeds were tested at this site. Tests outside a speed tolerance of ± 0.1 m/s were repeated.

Sixteen impacts (8 per lab) were performed at each impact site ($n = 12$) and speed ($n = 3$ for HR; $n = 4$ for CR/JP; $n = 5$ for other sites) yielding a total of 896 impacts. Eight different samples of each sensor and 8 different helmets were used in each of two laboratories (Biokinetics and Associates Ltd., Ottawa, ON and Southern Impact Research Center, Rockville, TN). Facemasks and chinstraps were replaced after FM, FB and JP impacts of 7.3 m/s and higher. Tests were blocked by impact site and speed to improve repeatability. Tests were separated by ≥ 5 min, and repeat tests on the same helmet were separated by ≥ 20 min. Tests were repeated for laboratory errors (e.g., over/under speed, no trigger signal, leaking helmet liner pads) and
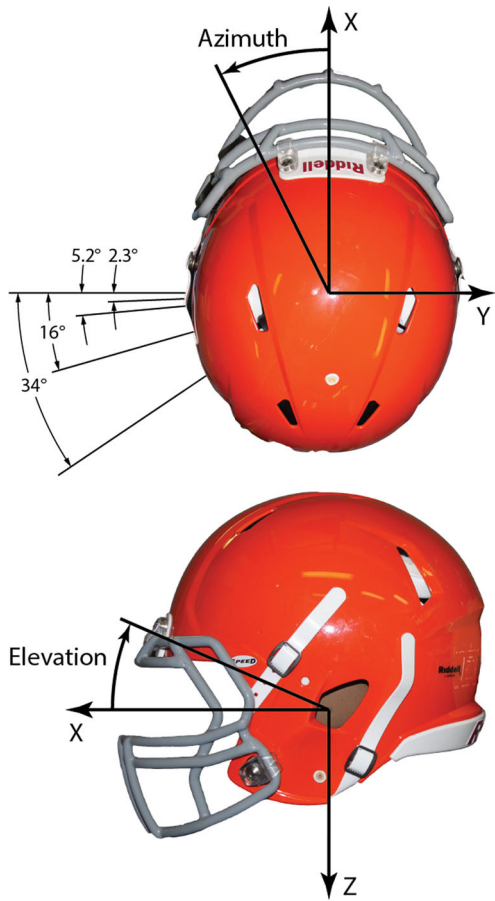
FIGURE 2. Definition of the reference frame and impact direction angles. Also shown are graphical depictions of the angular differences that were deemed significantly different for HITS (5.2°) and X2 (2.3°) in our statistical analyses, and the median difference between the sensor's impact direction and the MLSH impact direction for HITS (34°) and X2 (16°) across all sites (see "Results" and "Discussion" sections for more details).

for experimenter-related sensor issues (e.g., dead batteries, damaged sensors). Tests were not repeated for impact related issues (e.g., chinstrap unbuckled during impact or helmet came off during impact).

The set of first valid trials, i.e., no laboratory errors, was designated Database D_0 and was used to calculate the detection rate for both devices. To maximize the data for the validation analysis, tests were repeated if a device did not detect an impact or if an impact was cleansed. The sets of valid trials with the most HITS data and the most X2 data were designated D_{HITS} and D_{X2} respectively. D_{HITS} and D_{X2} were used to validate the impact directions and magnitudes reported by HITS and X2 respectively.

The following parameters were exported from both systems: PLA, PAA, HIC, azimuth (Az°) and elevation (El°). HITS reportedly estimates the impact location, whereas X2 estimates the direction of PLA. All direc-

FIGURE 3. Impact directions and cumulative distributions of ▶ error for HITS (red data) and X2 (blue data) as a function of impact site (code at top right of each pair of graphs). The polar plots show the individual test data plotted on a top view of the helmet, nose towards the top of the page, and the portion of the helmet below the horizontal plane (0° elevation, indicated by the dark equator) is projected outside the equator (see Legend at top of figure). Below each polar plot is a cumulative distribution of the absolute angle error (|Angle|) of HITS relative to the contact point (red dashed line), HITS relative to the MLSH (solid red line) and X2 relative to the MLSH (blue line). The dark and light gray shaded regions in the cumulative distributions correspond to 22.5° and 45° errors.

tions were transformed into a common reference frame (Fig. 2). The reference accelerometers were low-pass filtered (300 Hz, zero-lag, 4-pole Butterworth; Newman et al.[28]) before calculating the 6 degree of freedom (DOF) accelerations at the headform's COG.[31] Impact direction (Az°, El°) was calculated from the 3 components of the resultant PLA vector pointing towards the COG. Data from each trial for PLA, PAA, HIC, Az° and El° from the MLSH, HITS and X2 are provided in the Supplementary Materials. Representative linear and angular acceleration vs. time graphs for each impact site are also included in the Supplementary materials.

Each sensor's detection rate was computed as the percentage of impacts with valid data (as classified by the manufacturer) within the D_0 dataset. To evaluate impact direction, we calculated the mean direction, standard deviation and focus for the MLSH and both sensors at each impact site using spherical methods.[20,23] Since the distribution of the direction data about the mean was asymmetrical (Kent distribution) for most sites, we reported the major and minor semi-axes of standard deviation ellipses (SDEs). Validity of the azimuth and elevation data was assessed using the mean spherical error (δ_{AZ}, δ_{EL}) between HITS and impactor contact location and between both devices and MLSH directions at each impact site. If the 95th-percentile confidence ellipse (CE) of the mean spherical error (MSE) at each site included the origin ($\delta_{AZ} = 0$, $\delta_{EL} = 0$), then the device direction was not significantly different from the reference direction.[20] We also calculated the arithmetic mean and standard deviation (SD) of the absolute angular errors, i.e., the spatial angle between the sensor and reference directions, at each impact site. Spherical analyses were done in Matlab (2013b, MathWorks, Natick, MA) using SPAK library functions.[23]

To quantify the validity of impact magnitude (PLA, PAA and HIC), we used general linear mixed models[26,39] (see Supplementary Materials for more information) and arbitrarily defined 3 levels of validity. Level I validity was defined as an average intercept and

Front

Left · Right

Rear

±45° and ±22.5° about the Reference mean

±45° and ±22.5° about the Contact point

Contact point

Vertex

+45°
0°
−45°
−90°

MLSH
HITS
X2

FH · FB · SD · RB

RR · HR · CR · FM

JP · OB · E1 · E2

Cumulative Proportion
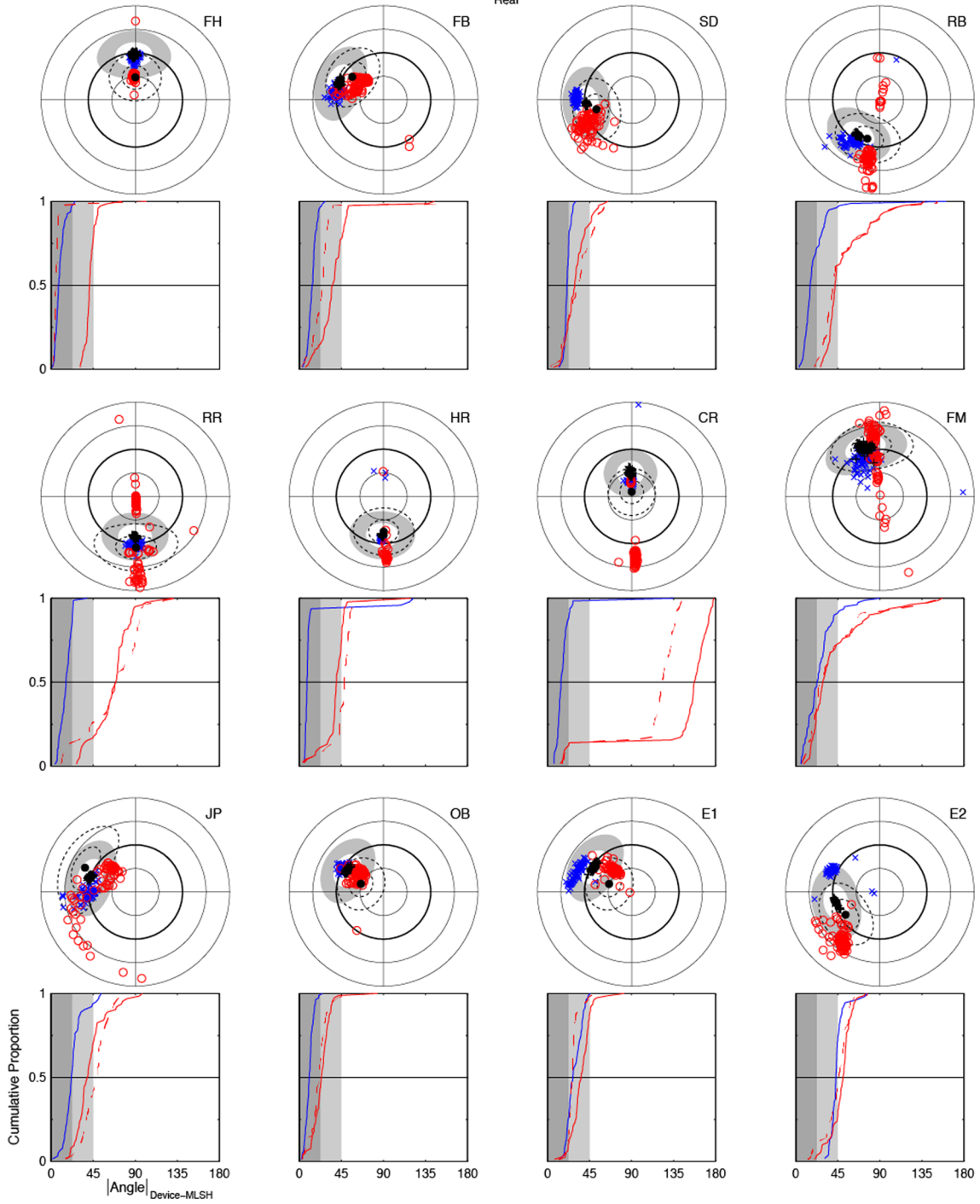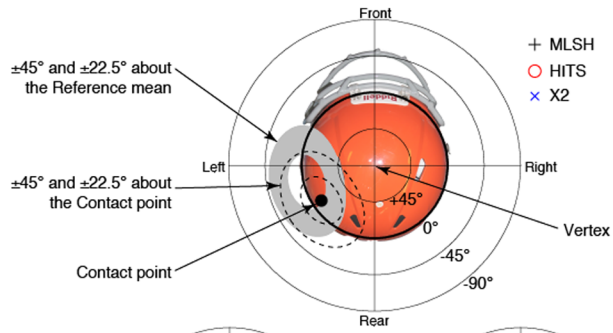
0 · 45 · 90 · 135 · 180

|Angle|Device−MLSH

TABLE 1. Mean direction (azimuth, elevation), standard deviation ellipse SDE (semi-major axis, semi-minor axis) and focus for the reference (MLSH) and sensors at each impact site within each dataset.

| | Dataset D_{HITS} | | | | | | Dataset D_{X2} | | | | | |
| | MLSH | | | HITS | | | MLSH | | | X2 | | |
	Mean (Az°, El°)	SDE (°)	Focus	Mean (Az°, El°)	SDE (°)	Focus	Mean (Az°,El°)	SDE (°)	Focus	Mean (Az°, El°)	SDE (°)	Focus
FH	(2, 3)	(3, 2)[a]	1.00	(4, 44)	(8, 1)[a]	0.98	(2, 3)	(3, 2)[a]	1.00	(1, 11)	(6, 3)[a]	0.99
FB	(69, 0)	(3, 1)[a]	1.00	(61, 32)	(15, 10)	0.91	(69, 0)	(3, 1)[a]	1.00	(81, 0)	(8, 5)[a]	0.99
SD	(95, 3)	(1, 1)[a]	1.00	(120, −7)	(17, 9)[a]	0.94	(95, 3)	(1, 1)[a]	1.00	(88, −16)	(4, 3)[a]	1.00
RB	(149, 11)	(2, 1)[a]	1.00	(169, −20)	(24, 3)[a]	0.71	(149, 11)	(2, 1)[a]	1.00	(146, −7)	(10, 6)	0.95
RR	(179, 13)	(2, 1)[a]	1.00	(186, 15)	(50, 6)[a]	0.38	(179, 13)	(2, 1)[a]	1.00	(180, −2)	(6, 5)	0.99
HR	(178, 16)	(2, 1)[a]	1.00	(183, −22)	(12, 1)[a]	0.94	(178, 16)	(2, 1)[a]	1.00	(177, 10)	(13, 3)	0.90
CR	(3, 45)	(4, 2)[a]	1.00	(183, −21)	(18, 2)	0.73	(3, 45)	(5, 2)[a]	1.00	(3, 60)	(6, 4)	0.97
FM	(17, −7)	(7, 4)[a]	0.99	(6, −9)	(32, 3)[a]	0.70	(17, −7)	(7, 4)[a]	0.99	(24, 12)	(16, 10)[a]	0.93
JP	(72, −1)	(3, 2)[a]	1.00	(69, 9)	(34, 11)[a]	0.73	(72, −1)	(3, 2)[a]	1.00	(90, −2)	(16, 9)[a]	0.94
OB	(59, 9)	(3, 1)[a]	1.00	(56, 30)	(10, 10)	0.96	(59, 9)	(3, 1)[a]	1.00	(59, −1)	(5, 4)[a]	0.99
E1	(57, 3)	(4, 2)[a]	1.00	(41, 35)	(12, 6)[a]	0.97	(56, 3)	(4, 2)[a]	1.00	(70, −22)	(9, 5)[a]	0.98
E2	(104, 4)	(4, 1)[a]	1.00	(140, −29)	(14, 7)[a]	0.96	(105, 4)	(4, 1)[a]	1.00	(65, −8)	(13, 6)[a]	0.95
Avg		2.5[b]	1.00		10.9[b]	0.83		2.5[b]	1.00		7.4[b]	0.97
SD			0.00			0.18			0.00			0.03

Average (Avg) and standard deviation (SD) of the SDE axes and focus values are given at the bottom of the table.
[a] Data distributed on the sphere in a Kent distribution (general bivariate normal distribution) rather than in a Fisher distribution (isotropic bivariate normal distribution).
[b] Represents the radius, in degrees, of a circle with the average area of the standard deviation ellipses.

TABLE 2. Pairwise angular differences (errors) between the sensor and reference directions at each impact site for HITS with respect to the impactor contact point, HITS with respect to the MLSH, and X2 with respect to the MLSH.

| | HITS wrt contact point | | | HITS wrt MLSH | | | X2 wrt MLSH | | |
	Mean (°)	SDE (maj,min)[b]	|Angle| (°)	Mean (°)	SDE (maj,min)[b]	|Angle| (°)	Mean (°)	SDE (maj,min)[b]	|Angle| (°)
FH	4	(8, 1)[a]	7 (12)	41	(9, 2)[a]	42 (6)	8	(6, 3)[a]	9 (5)
FB	17	(15, 10)	26 (22)	33	(15, 10)	37 (20)	12	(7, 4)[a]	14 (5)
SD	31	(17, 9)[a]	33 (15)	26	(17, 9)[a]	30 (12)	21	(4, 3)[a]	21 (3)
RB	33	(24, 3)[a]	50 (27)	36	(24, 4)[a]	53 (25)	18	(10, 5)	20 (19)
RR	24	(50, 6)[a]	66 (29)	6	(50, 6)[a]	66 (20)	15	(6, 4)[a]	16 (6)
HR	45	(12, 1)[a]	47 (15)	38	(12, 1)[a]	40 (16)	5	(13, 3)	15 (27)
CR	120	(18, 2)	111 (39)	155	(19, 3)[a]	141 (51)	15	(6, 3)	17 (16)
FM	4	(32, 3)[a]	39 (33)	12	(32, 5)[a]	42 (32)	20	(16, 11)[a]	25 (16)
JP	26	(34, 11)[a]	48 (13)	*10*	*(34, 11)[a]*	42 (17)	17	(16, 8)[a]	23 (10)
OB	18	(10, 10)	21 (11)	21	(10, 9)	24 (11)	10	(5, 4)[a]	11 (4)
E1	25	(12, 6)[a]	27 (7)	35	(12, 7)[a]	36 (11)	28	(9, 5)[a]	29 (8)
E2	43	(14, 7)[a]	44 (13)	48	(14, 6)[a]	49 (11)	41	(13, 7)[a]	44 (8)
Avg	32.5	10.9[c]	42	38.4	11.3[c]	49	17.5	7.3[c]	20
SD	30.4		33	39.1		35	9.8		15
Med	26		33	34		40	16		17
10 percentile			8			21			7
90 percentile			92			85			38

Data are given as the mean spherical error (MSE) and its standard deviation ellipse SDE (semi-major axis, semi-minor axis), and the mean (standard deviation) of the absolute angular error (\|Angle\|). Average (Avg), standard deviation (SD) and median (Med) of the mean, SDE axes and absolute angles are given at the bottom of the table. The italicized cells show sites where the sensor and reference directions were not significantly different.
[a] Data distributed on the sphere in a Kent distribution (general bivariate normal distribution) rather than in a Fisher distribution (isotropic bivariate normal distribution).
[b] (maj, min) means the angular arc of the semi-major and semi-minor axes of the standard deviation ellipse.
[c] Represents the radius, in degrees, of a circle with the average area of the standard deviation ellipses.

slope that were not statistically different from zero and one respectively for all impact sites combined. This meant that on average the device correctly measured the parameter of interest despite potentially incorrect measurements at some or all sites. Level II validity was defined as Level I validity and site-specific deviations in

intercept and slope that were not statistically different from zero for the 7 centroidal impact sites where the impactor contacted the helmet shell, i.e., FH, SD, RB, RR, HR, CR, and OB. This meant the device correctly measured the parameter of interest for centroidal helmet impacts. Level III validity was defined as Level II validity and no significant deviations in site-specific intercepts and slopes for the facemask and eccentric impacts, i.e., FB, FM, JP, E1, and E2. Level III validity meant the device measured the parameter of interest correctly at all 12 sites.

The models were implemented in SAS (SAS Institute Inc, Cary, NC) and Matlab using maximum likelihood estimation (ML) and the Satterthwaite method of approximating degrees of freedom. Separate analyses were carried out for PLA, PAA and HIC for each device and a significance level of $p < 0.05$ was used for all tests. PLA, PAA and HIC were log-transformed before fitting the models. To compare with prior validation studies, we computed the mean slope (β) and coefficient of determination (R^2) for a linear regression with zero-intercept between the untransformed reference and sensor data across all sites and speeds.[4,5] We also computed the relative errors [(RE, Eq. (1)] and absolute errors [AE, Eq. (2)] for each site.[18] Mean RE represents the expected device error over a population of impacts, whereas mean AE represents the expected device error for an individual impact.

$$\text{RE}_{ij} = \frac{(x_{ij} - y_{ij})}{y_{ij}} \quad i = \text{trial}, j = \text{site} \quad (1)$$

$$\text{AE}_{ij} = \frac{|x_{ij} - y_{ij}|}{y_{ij}} \quad i = \text{trial}, j = \text{site} \quad (2)$$

RESULTS

HITS detected and classified as valid 861 of the 896 impacts (96.1%) and X2 detected and classified as valid 845 of the 896 impacts (95.4%) (see the Supplementary Materials for a detailed description of impacts classified as invalid). After including repeated tests, the D_{HITS} dataset for the validation analysis contained 878 impacts (2.0% missing) and the D_{X2} dataset contained 863 tests (3.7% missing). The two validation datasets shared 825 common trials.

MLSH impact directions were highly focused (>0.99), had low variability (average SDE semi-axis $\pm 2.5°$), and aside from CR, were concentrated along the helmet equator ($-7° \leq \text{El}° \leq 16°$) (Fig. 3; Table 1). HITS impact directions were less focused (range 0.38–0.98) and more variable (average SDE semi-axis $\pm 10.9°$). HITS performed best when estimating impactor contact location at FH (MSE = 4° SDE major semi-axis $\pm8°$, Table 2) and worst when estimating both impactor contact location and MLSH direction at CR, where it reported most impacts as being from behind and below (MSE = 120° and 155° respectively, Table 2; Fig. 3). Excluding CR, MSE's averaged $25 \pm 13°$ relative to the contact point and $28 \pm 14°$ relative to the MLSH impact direction. Based on the 95th-percentile confidence ellipses, HITS impact directions were significantly different from all contact points, and all MLSH impact directions except for JP.

X2 impact directions were also less focused (range 0.93–1.00) and more variable (average SDE semi-axis $\pm7.4°$) than the reference directions (Fig. 3; Table 1). X2 performed best at HR and FH (MSE = 5° and 8° respectively) and worst at E2 (MSE = 41°) (Table 2). Across all sites, MSE averaged $18° \pm 10°$. Based on the 95th-percentile confidence ellipses, all X2 impact directions were significantly different from the MLSH impact directions.

Average absolute angular errors were similar to mean spherical errors (MSEs) except for HITS at RR, FM and JP, where low MSEs belied large variability in the underlying data (Table 2). Low variability in X2 directions yielded average absolute angular errors within 10° of MSE at all sites. Cumulative distributions of the absolute angular errors at each site are given in Fig. 3.

TABLE 3. Average intercepts (β_{00}) and slopes (β_{10}) from the general linear mixed model (GLMM) for log-transformed peak linear acceleration (PLA), peak angular acceleration (PAA), and the head injury criterion (HIC).

	HITS				X2			
	Intercept (β_{00})	p value	Slope (β_{10})	p value	Intercept (β_{00})	p value	Slope (β_{10})	p value
PLA	0.591	0.0064	0.829	0.0014	0.200	0.0072	0.972	0.1445
PAA	2.368	0.0039	0.706	0.0029	1.697	0.0027	0.775	0.0027
HIC	1.170	0.0002	0.794	0.0007	0.232	0.0450	0.969	0.2640

The p values evaluate the probability that the average intercept (β_{00}) and average slope (β_{10}) are significantly different from 0 and 1 respectively.
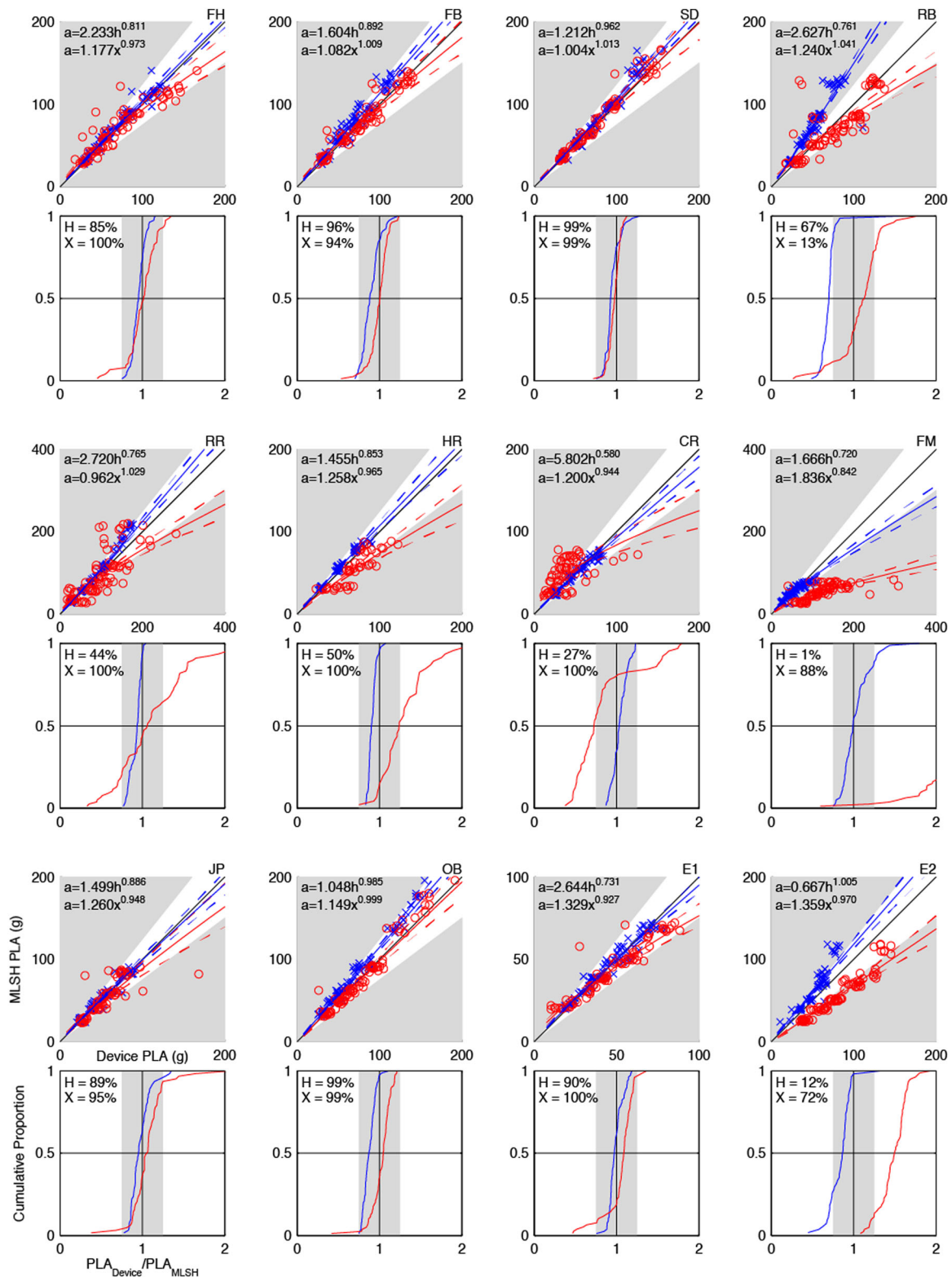
FIGURE 4. Peak linear accelerations (PLA) and cumulative distributions of error for HITS (red data) and X2 (blue data) as a function of impact site (site code at top right of each pair of graphs). In the upper panel of each pair the diagonal black line represents the unity slope with a surrounding white region representing ±25%. Equations in the top left corner show the systematic log-transformed GLMM results for predicting the MLSH data (a) from the HITS (h) and X2 (x) data. The solid and dashed red and blue lines show the equation and 95th percentile confidence interval for each sensor. The lower panel of each pair shows the cumulative distribution of the ratio between the device and the MLSH reference for all tests at each site. The percentages of tests that fall within the ±25% shaded corridor are shown in the top left corner for HITS (H) and X2 (X).
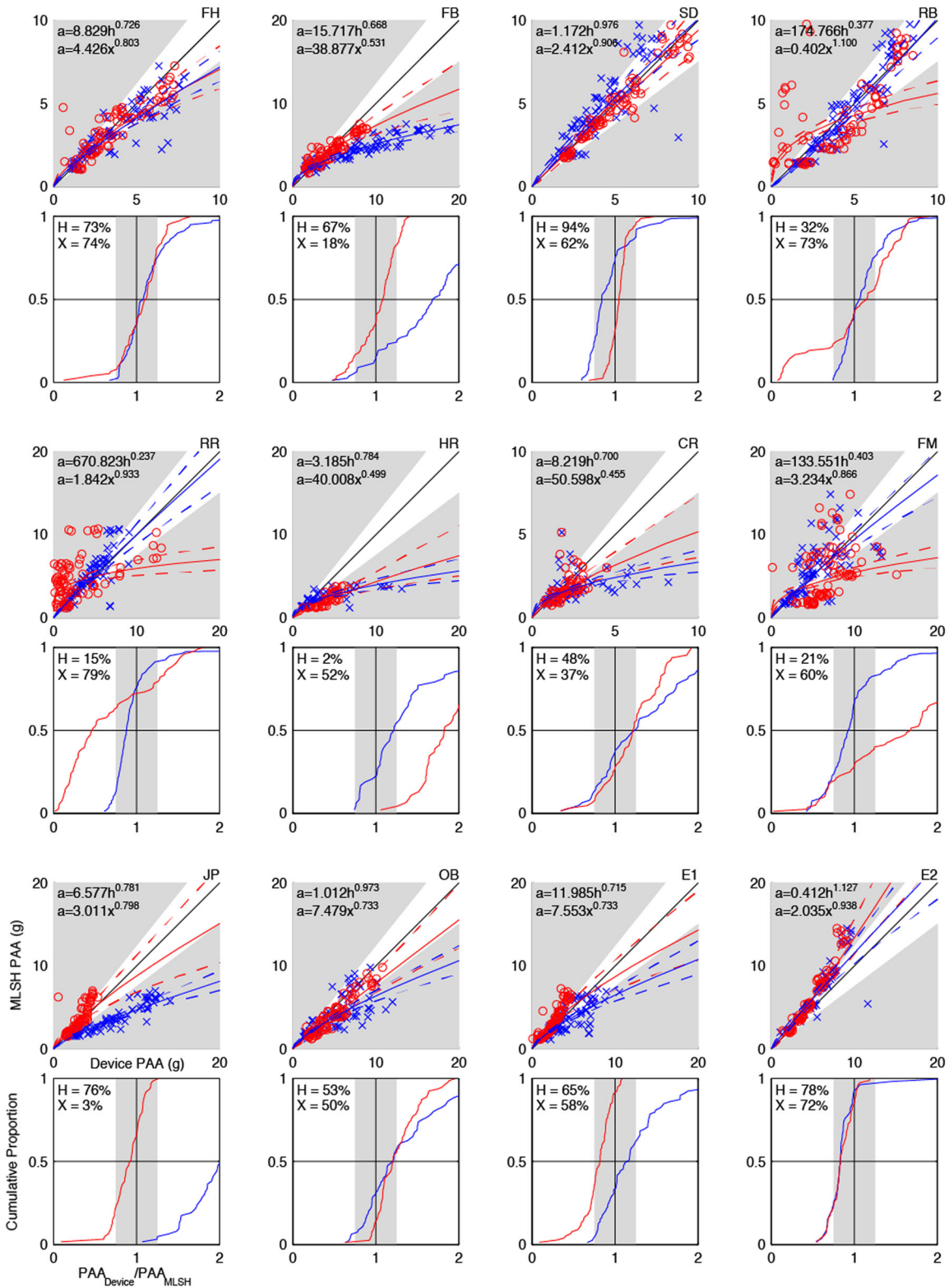
FIGURE 5. Peak angular accelerations (PAA) and cumulative distributions of error for HITS (red data) and X2 (blue data) as a function of impact site (site code at top right of each pair of graphs). In the upper panel of each pair the diagonal black line represents the unity slope with a surrounding white region representing ± 25%. Equations in the top left corner show the systematic log-transformed GLMM results for predicting the MLSH data (a) from the HITS (h) and X2 (x) data. The solid and dashed red and blue lines show the equation and 95th percentile confidence interval for each sensor. The lower panel of each pair shows the cumulative distribution of the ratio between the device and the MLSH reference for all tests at each site. The percentages of tests that fall within the ±25% shaded corridor are shown in the top left corner for HITS (H) and X2 (X).
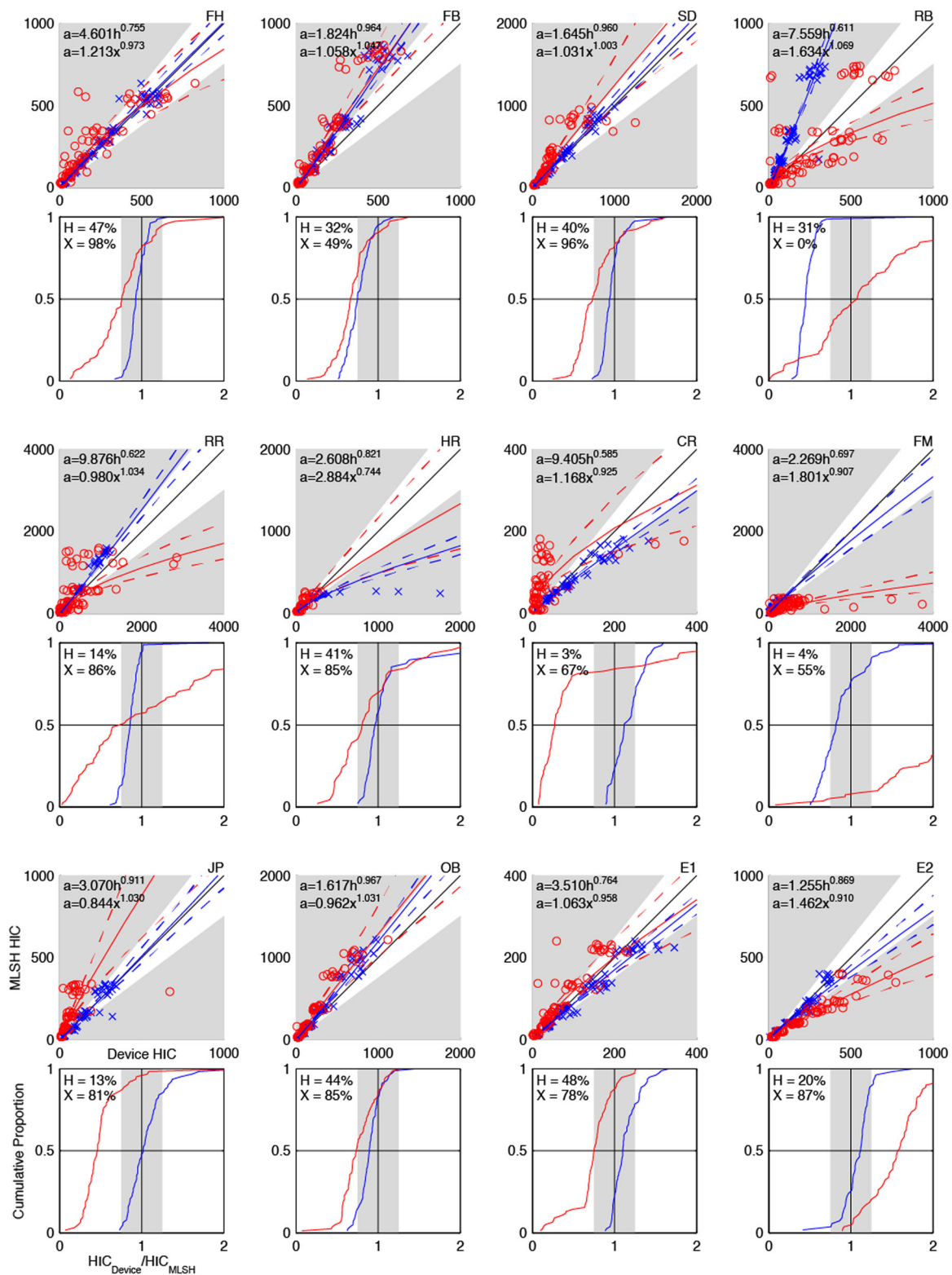
FIGURE 6. Head injury criterion (HIC) and cumulative distributions of error for HITS (red data) and X2 (blue data) as a function of impact site (site code at top right of each pair of graphs). In the upper panel of each pair the diagonal black line represents the unity slope with a surrounding white region representing ±25%. Equations in the top left corner show the systematic log-transformed general linear mixed model results for predicting the MLSH data (a) from the HITS (h) and X2 (x) data. The solid and dashed red and blue lines show the equation and 95th percentile confidence interval for each sensor. The lower panel of each pair shows the cumulative distribution of the ratio between the device and the MLSH reference for all tests at each site. The percentages of tests that fall within the ±25% shaded corridor are shown in the top left corner for HITS (H) and X2 (X).

TABLE 4. HIT system: site-specific intercepts ($\beta_{00} + \beta_{0j}$) and slopes ($\beta_{10} + \beta_{1j}$) for HITS from the general linear mixed model (GLMM) for peak linear acceleration (PLA), peak angular acceleration (PAA), and the head injury criterion (HIC).

	PLA		PAA		HIC	
	Intercept ($\beta_{00} + \beta_{0j}$)	Slope ($\beta_{10} + \beta_{1j}$)	Intercept ($\beta_{00} + \beta_{0j}$)	Slope ($\beta_{10} + \beta_{1j}$)	Intercept ($\beta_{00} + \beta_{0j}$)	Slope ($\beta_{10} + \beta_{1j}$)
CR—crown	1.758*	0.580*	2.106*	0.700*	2.241*	0.585*
E1—front Eccentric	0.972*	0.731*	2.484*	0.715*	1.256*	0.764*
E2—rear eccentric	− 0.405	1.005	− 0.888	1.127	0.227	0.869*
FB—Front boss	0.473*	0.892*	2.755*	0.668*	0.601*	0.964
FH—Forehead	0.803*	0.811*	2.178*	0.726*	1.526*	0.755*
FM—Face mask	0.510*	0.720*	4.894*	0.403*	0.820*	0.697*
HR—High rear	0.375	0.853*	1.159	0.784	0.958*	0.821*
JP—jaw pad	0.405	0.886*	1.884*	0.781*	1.122*	0.911
OB—oblique	0.046	0.985	0.012	0.973	0.480	0.967
RB—rear boss	0.966*	0.761*	5.163*	0.377*	2.023*	0.611*
RR—low rear	1.001*	0.765*	6.509*	0.237*	2.290*	0.622*
SD—side	0.192	0.962	0.159	0.976	0.498	0.960

The asterisk (*) indicates intercepts and slopes that are significantly different from 0 and 1 respectively. Italic pairs highlight those sites where both the intercept and slope were not significantly different from 0 and 1 respectively.

TABLE 5. X2 mouthguard system: Site-specific intercepts ($\beta_{00} + \beta_{0j}$) and slopes ($\beta_{10} + \beta_{1j}$) for X2 from the general linear mixed model (GLMM) for peak linear acceleration (PLA), peak angular acceleration (PAA), and the head injury criterion (HIC).

	PLA		PAA		HIC	
	Intercept ($\beta_{00} + \beta_{0j}$)	Slope ($\beta_{10} + \beta_{1j}$)	Intercept ($\beta_{00} + \beta_{0j}$)	Slope ($\beta_{10} + \beta_{1j}$)	Intercept ($\beta_{00} + \beta_{0j}$)	Slope ($\beta_{10} + \beta_{1j}$)
CR—crown	0.182	0.944*	3.924*	0.455*	0.156	0.925*
E1—front eccentric	0.285*	0.927*	2.022*	0.733*	0.061	0.958*
E2—rear eccentric	0.307*	0.970	0.710	0.938	0.380*	0.910*
FB—front boss	0.079	1.009	3.660*	0.531*	0.056	1.047*
FH—forehead	0.163	0.973	1.488*	0.803*	0.193*	0.973
FM—face mask	0.607*	0.842*	1.174*	0.866*	0.588*	0.907*
HR—high rear	0.229*	0.965	3.689*	0.499*	1.059*	0.744*
JP—jaw pad	0.231*	0.948*	1.102*	0.798*	− 0.170	1.030
OB—oblique	0.139	0.999	2.012*	0.733*	− 0.038	1.031
RB—rear boss	0.215*	1.041	− 0.912	1.100	0.491*	1.069*
RR—low rear	− 0.039	1.029	0.611	0.933	− 0.021	1.034*
SD—side	0.004	1.013	0.881	0.906	0.030	1.003

The asterisk (*) indicates intercepts and slopes that are significantly different from 0 and 1 respectively. Italic pairs highlight those sites where both the intercept and slope were not significantly different from 0 and 1 respectively.

Neither HITS nor X2 achieved Level I validity for PLA, PAA or HIC, although for X2 the average slopes for PLA and HIC met the criterion but the average intercepts did not (Table 3). Based on these results, we did not test further for Level II and III validity. Despite neither sensor achieving Level I validity, we computed site-specific intercepts and slopes to assess how validity varied across sites (Figs. 4, 5, and 6; Tables 4 and 5). HITS was valid for all three variables (PLA, PAA, HIC) at the SD and OB sites, and valid for at least one of the outcomes at 2 other sites (SD, HR, OB and E2; Table 4). At many sites, fewer than half of the tests fell within ±25% of the reference value (4 sites for PLA, 5 sites for PAA,

and all sites for HIC) (see percentages in top left corner of error panels in Figs. 4, 5, and 6). X2 was valid for all three variables at the SD site only, and valid for at least one variable at another 7 sites (Table 5). At only a few sites did more than half of the tests fall outside ±25% (PLA: 1 site; PAA: 3 sites; HIC: 2 sites). The average coefficients of variation across all sites were lowest for the MLSH (PLA 4.4 ± 1.2%, PAA 8.0 ± 3.5%) and two or more times larger for both HITS (PLA 15 ± 9%, PAA 21 ± 21%) and X2 (PLA 8.0 ± 3.5%, PAA 23 ± 9%) (see Supplementary Materials for site-specific data).

Simple linear regressions on the pooled data yielded slopes significantly different from one for both sensors

TABLE 6. Results of simple linear regression ($y_n = \beta x_n$) for peak linear acceleration (PLA), peak angular acceleration (PAA) and head injury criteria (HIC) pooled across all impact sites and speeds for the HITS and X2 sensors.

	HITS ($N = 878$)				X2 ($N = 863$)			
	Slope β	SE	p value	R^2	Slope β	SE	p value	R^2
PLA	0.814	0.010	<0.0001	0.626	1.108	0.006	<0.0001	0.897
PAA	0.936	0.013	<0.0001	0.594	0.768	0.012	<0.0001	0.560
HIC	0.698	0.021	<0.0001	0.483	1.128	0.013	<0.0001	0.817

Values given are the slope estimate (β), the standard error (SE) of the slope estimate, the p value for whether the slope is significantly different from 1, and coefficient of determination (R^2).

TABLE 7. HIT system: average and standard deviation (SD) of the percent (%) relative error (RE) and absolute value of the relative error (AE) for peak linear acceleration (PLA), peak angular acceleration (PAA) and head injury criteria (HIC) at each site for the HITS sensor.

HITS Site	PLA		PAA		HIC	
	RE (SD) (%)	AE (SD) (%)	RE (SD) (%)	AE (SD) (%)	RE (SD) (%)	AE (SD) (%)
FH	−0 (18)	13 (12)	7 (26)	21 (17)	−23 (35)	34 (24)
FB	−1 (12)	8 (8)	3 (23)	19 (13)	−31 (22)	34 (17)
SD	−3 (7)	6 (5)	5 (12)	9 (8)	−21 (27)	30 (16)
RB	8 (26)	21 (18)	4 (48)	39 (28)	18 (77)	57 (54)
RR	13 (45)	35 (31)	−33 (52)	56 (26)	12 (106)	78 (72)
HR	30 (30)	32 (28)	92 (42)	92 (42)	−10 (44)	36 (27)
CR	−16 (38)	37 (18)	23 (35)	33 (24)	−47 (64)	76 (24)
FM	136 (60)	137 (57)	62 (78)	80 (59)	208 (266)	214 (262)
JP	7 (22)	14 (17)	−10 (19)	16 (14)	−49 (30)	54 (21)
OB	3 (12)	9 (8)	26 (26)	28 (24)	−23 (21)	26 (17)
E1	5 (17)	14 (11)	−21 (18)	22 (17)	−25 (25)	29 (21)
E2	48 (17)	48 (17)	−15 (13)	17 (11)	56 (33)	56 (32)
Avg (SD)	20 (50)	32 (43)	9 (49)	35 (36)	8 (117)	62 (99)
Median	7	15	3	23	−21	42
10 percentile	−22	2	−41	4	−71	10
90 percentile	79	79	67	82	106	106

Average (Avg), SD, median, 10th and 90th percentiles for data and pooled across all impact sites.

and all three variables (Table 6). The coefficients of determination (R^2) varied from 48 to 63% for HITS and from 56 to 90% for X2. The average REs for the pooled data varied from 8 to 20% for HITS and from −8 to 26% for X2, but large variability in the data yielded AEs that varied from 32 to 62% for HITS and from 12 to 41% for X2 (Tables 7 and 8).

DISCUSSION

This study quantifies the validity of the HITS and X2 mouthguard against reference sensors mounted to a laboratory headform. Both systems are designed to measure head impact kinematics in football players and report the direction and peak kinematics of impacts they detect and deem valid. In contrast to prior validation studies,[1,4,5,18,24,35] our protocol used more impact sites, a wider range of impact speeds, and multiple samples of each device. Overall, we found that both devices detected >95% of impacts and generated graded response data and consistent impact directions at most impact sites tested. Using our relatively strict validity criteria, we found that both devices have limitations: (i) both sensors did not detect ~4% of impacts, (ii) both sensors reported impact directions that were significantly different from the reference directions, and (iii) neither sensor accurately measured PLA, PAA or HIC at all impact sites. Overall, our laboratory findings indicate that users should not rely solely on these devices to accurately measure the direction and magnitude of single impacts to a football player's head at all impact sites. Despite these limitations, both devices generated data that are otherwise unavailable and can be used to study football head impacts if the data

TABLE 8. X2 mouthguard system: Average and standard deviation (SD) of the percent (%) relative error (RE) and absolute value of the relative error (AE) for peak linear acceleration (PLA), peak angular acceleration (PAA) and head injury criteria (HIC) at each site for the X2 mouthguard.

X2 Site	PLA		PAA		HIC	
	RE (SD) (%)	AE (SD) (%)	RE (SD) (%)	AE (SD) (%)	RE (SD) (%)	AE (SD) (%)
FH	−5 (8)	8 (5)	15 (36)	24 (31)	−5 (11)	10 (7)
FB	−10 (11)	13 (8)	67 (56)	73 (47)	−24 (14)	25 (12)
SD	−5 (9)	9 (5)	−6 (33)	23 (24)	−3 (14)	10 (10)
RB	−30 (11)	32 (7)	12 (30)	21 (24)	−53 (17)	55 (9)
RR	−8 (7)	8 (6)	2 (65)	27 (60)	−14 (14)	16 (11)
HR	−9 (5)	9 (4)	50 (92)	58 (87)	27 (110)	39 (107)
CR	4 (9)	8 (6)	41 (72)	56 (61)	16 (18)	19 (15)
FM	3 (17)	13 (12)	4 (56)	31 (46)	−11 (33)	27 (21)
JP	−2 (12)	10 (7)	101 (43)	101 (43)	6 (25)	17 (20)
OB	−12 (7)	13 (6)	32 (49)	41 (41)	−10 (14)	14 (10)
E1	−1 (8)	7 (5)	27 (54)	36 (48)	14 (16)	15 (15)
E2	−16 (13)	17 (11)	−14 (23)	20 (17)	9 (19)	16 (14)
Avg (SD)	−8 (14)	12 (10)	26 (61)	41 (51)	−6 (37)	22 (31)
Median	−7	9	5	22	−8	14
10 percentile	−25	2	−25	4	−42	3
90 percentile	8	28	103	103	22	49

Average (Avg), SD, median, 10th and 90th percentiles for data pooled across all impact sites.

are interpreted properly in the context of their validity.

Head impacts must be detected to be measured. Both HITS and X2 either missed or misclassified ~4% of impacts. Particularly worrying are missed impacts (9 by HITS, 21 by X2), which could go unnoticed by clinicians relying solely on these devices. A broken battery wire, possibly from repeatedly removing and reinstalling the battery, explained one of the missed HITS impacts. Fifteen of the missed X2 impacts were at E2, a site that generated head rotations about an axis near X2's accelerometer and thus may not have reached its 10 g trigger threshold. Triggering on angular rate as well could potentially fix this problem. The reasons for the remaining missed impacts, particularly those at 9.3 and 11.2 m/s that were missed by HITS (see Supplementary Materials), remain unclear and require further investigation.

Impacts that are detected and classified as invalid are a separate issue. Both devices have post-processing algorithms to evaluate signal characteristics indicative of impact or artifact, and some nominal rate of misclassification is inevitable. Users should be aware of this issue and review invalid trials when they witness an impact without valid data. Moreover, our laboratory protocol could not assess false positives (non-impacts classified as valid head impacts) and future on-field work should explore this phenomenon.

Impact direction is a potentially important parameter because tissue strain and brain injury tolerance appear to vary with the direction of impact-induced skull motion.[12,30,40,43] Aside from the JP site for HITS,

both systems reported impact directions that were significantly different from the reference (Table 2). These significant differences were likely due to the paired nature of the statistical analysis and the large number of samples ($N = 48$ to 80) at each impact site. Indeed an a posteriori analysis showed that on average our analysis identified significant differences when the MSEs were $> 5.2°$ for HITS and $> 2.3°$ for X2 (these angles are depicted graphically in Fig. 2). These small angular differences likely exceed the accuracy needed for measuring on-field impacts or calculating brain strain in a model, and therefore we also calculated the average absolute angular errors and evaluated the cumulative distribution of these absolute angular errors (see lower graph within each panel of Fig. 3). Here we found that across all tests HITS directions were $42° \pm 33°$ (range 7°–111°) different from the impactor contact point and $49° \pm 35°$ (range 24°–141°) different from MLSH directions (Table 2). Ignoring the CR site, the average absolute angular errors for HITS decreased to $37° \pm 25°$ (range 7°–66°) relative to the impactor contact point and $42° \pm 21°$ (range 24°–66°) relative to the MLSH direction. These absolute differences are similar, though less variable, than reported previously ($31° \pm 46°$)[4] for data pooled from four sites similar to our FM, FB, SD and RB sites. The cumulative distributions (Fig. 3) showed that four sites (RB, RR, CR and FM) accounted for most absolute angular errors larger than 45° for HITS.

For HITS, the CR site was particularly challenging: 55 of 64 CR impacts were reported as nearly opposite to the actual impact direction. HITS also reported
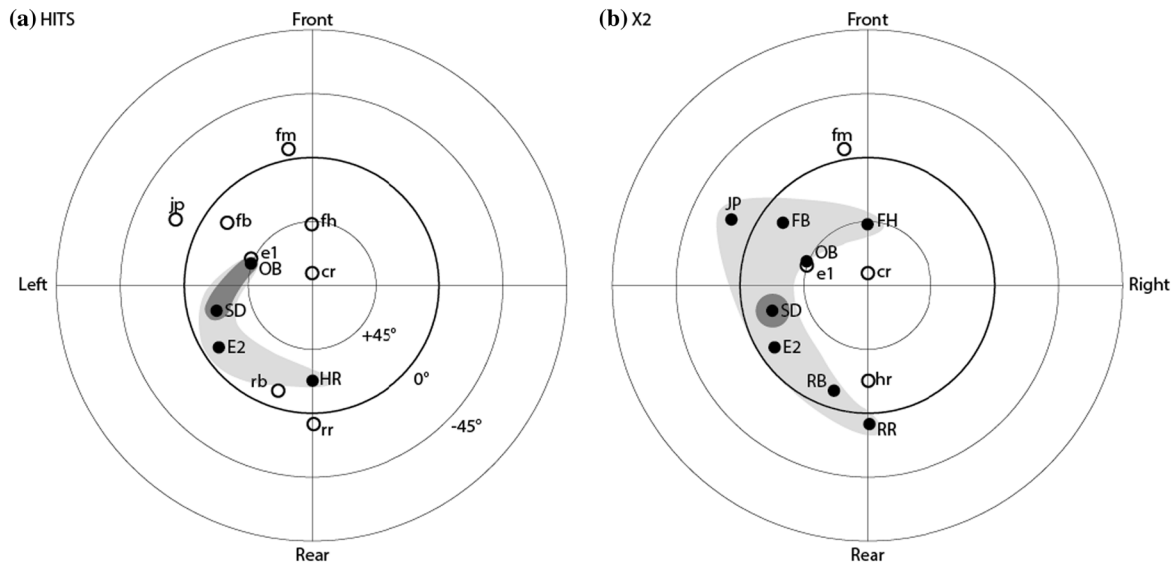
FIGURE 7. Schematic of impact sites for (a) HITS and (b) X2 mouthguard where all three magnitude parameters (PLA, PAA and HIC) were classified as ''valid'' (dark gray shading) and where at least one of the three parameters was classified as ''valid'' (light gray shading). Closed circles and capital letters indicate at least one valid parameter at this site; open circles and lower case letters indicate no valid parameters at the site. Two letter codes correspond to the impact sites in Fig. 1 and the legend for the coordinate frame is shown in Fig. 3.

42% of RR impacts, 10% of RB impacts and 4% of FM impacts as CR impacts (defined here as El° > 60°).[27] Side-by-side comparisons of test video for CR and RR impacts that produced correct and reflected HITS directions did not reveal obvious differences that explained these errors (see Supplementary Materials). Although large azimuth errors at the rear boss occur in HITS for hockey helmets,[1] none of the prior HITS validation studies for football reported these kinds of reflected impact directions for helmet shell impacts. To further explore this issue, we conducted follow-up tests using large Riddell Revolution helmets (a model preceding the Revolution Speed used here) since it was used in prior validation tests.[4,18] We found no direction errors in 5 CR impacts and one direction error in 12 RR impacts. These findings suggest that our direction issues were somehow related to the helmet models, although it remains unclear if they are due to design differences between the helmets or differences in how they fit the headform. The Speed helmet has a different outer shell than its predecessor, has an interior comfort fit overliner and inflatable liners. These findings also suggest that the HITS validation data presented here may be specific to the Speed helmet model. Whatever the explanation for the inter-helmet differences, validating a sensor for each helmet model is not optimal, and further work is needed to identify why the Speed helmets generated this problem. Since previous HITS data acquired mostly from Revolution helmets showed that concussions occur relatively frequently for impacts to the top of the head,[14,36] resolving this issue is important for interpreting HITS data acquired from Speed helmets.

The absolute angular errors in impact direction between X2 and the reference averaged $20° \pm 15°$ (range 9°–44°) and outliers nearly opposite to the true impact direction were occasionally seen (Fig. 3). Like with HITS, these reflected directions may indicate rebounding of the sensors relative to the headform or possibly some laboratory-related difference from *in vivo* sensor performance. Nevertheless, the mean X2 direction errors fell within a 45° cone ($\pm22.5°$) and are likely sufficiently accurate for many analyses of head impact data.

Neither sensor achieved Level 1 validity for PLA, PAA, or HIC, but X2 was close for both PLA and HIC. The 95th percentile confidence intervals (CI) for the average slopes for both parameters spanned one (PLA: 0.932–1.011; HIC: 0.912–1.027), but both intercept CIs just excluded zero (PLA: 0.066–0.335; HIC: 0.006–0.458). These findings differ from prior validation studies that showed both systems behaved well when data from multiple impact sites were pooled.[4,5] For impact sites similar to our SD, RB, FB and FM sites, Beckwith *et al.*[4] reported slopes (m) near one and significant coefficients of determination (r^2) when HITS data were regressed against reference sensors (PLA, $m = 1.009$, $r^2 = 0.903$; PAA, $m = 0.939$, $r^2 = 0.526$; HIC, $m = 1.061$, $r^2 = 0.787$). Likewise, for impact sites similar to our OB, SD, RR and FM sites, Camarillo *et al.*[5] reported similarly strong results when X2 data were regressed against

reference sensors (PLA, $m = 1.01$, $r^2 = 0.90$; PAA, $m = 0.90$, $r^2 = 0.89$). We found poorer results when we ran the same simple linear regressions on pooled data from our 12 sites (Table 8), although site-by-site comparisons between our data and these prior studies showed nominally similar slopes for PLA (difference between -0.20 and 0.12) at all of these sites except for the HITS facemask data.

Our results highlight two validation testing problems: (i) regression results from pooled data can exaggerate validity; and (ii) pooled slopes near one may be an artifact of the specific sites used for the validation. While some researchers have proposed that pooled data across all sites can be used to reach population-based conclusions even if site-specific data are not valid, this logic is only valid if the impact distribution in the study population is similar to the impact distribution used in the validation. If, for example, the population data were distributed like the impacts we tested, then our results suggest that even population-based data may not be accurate.

Head impact sensors should ideally be valid at all sites to be most useful in answering questions related to injury prevention and improving player safety. As with any study, the quality required of its instrumentation and data depends on the study goals. For studying impact number or frequency, our results suggest that both HITS and the X2 mouthguard are appropriate instruments. However, for studying impact direction and magnitude (severity), our results indicate more interpretation is needed. Our site-specific analyses showed that both sensors were valid for PLA, PAA and HIC at the SD site, with HITS also being valid at the OB site (Fig. 7). Of 25 reconstructed NFL concussions, 8 were caused by side impacts,[33] and of 57 concussions reported in NCAA players, 7 were described as primarily coronal-plane rotations caused by side impacts.[36] Combining these data, 15 of 82 (18%) concussions were caused by side impacts, leaving many concussions outside the valid areas of both sensors. Impacts to the front and rear helmet surfaces caused 36 of 82 (44%) concussions, but both sensors performed poorer at these sites: HITS was valid only for PAA at 1 of the 4 sagittal plane sites we tested, and X2 was valid for PLA at 2 sagittal plane sites and PAA at 1 sagittal plane site (Fig. 7).

As with impact direction, our validity criteria for impact magnitude was in part defined by variability within the data itself. As a result, good accuracy at one site could be classified as invalid if the variability at this site was low, whereas poor accuracy at another site might be classified as valid if variability at that site was high. For instance, PLA at the E2 site was valid for HITS but not for X2, despite 72% of the X2 data lying within ±25% of the reference data compared to only
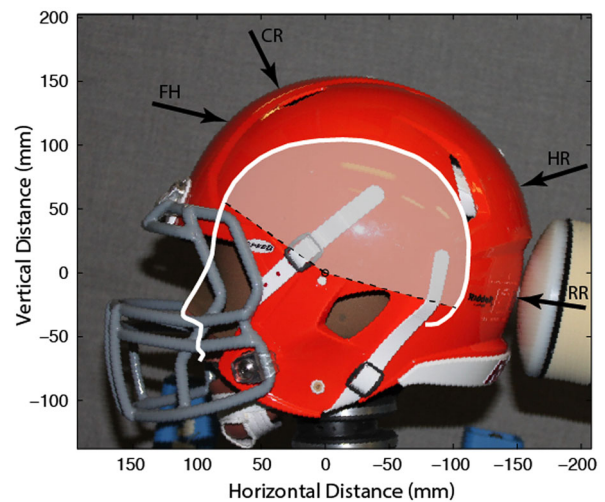

FIGURE 8. Schematic showing the headform position superimposed on a lateral view of the low rear (RR) impact site. The location and direction of the other mid-sagittal impact sites (FH, CR, and HR) are also shown. The dashed black lines and shaded region show where the geometry of the Hybrid III and NOCSAE headforms are in good agreement.[6]

12% of the HITS data (see E2 panel in Fig. 3). In recognition of the data-dependent criteria of our formal validity tests, we also examined relative and absolute errors of the impact magnitude (Tables 7 and 8). Excluding the FM site for HITS, average RE and AE were low for PLA (HITS: RE = 8%, AE = 19%; X2: RE = -8%, AE = 12%), and for HITS were similar to prior reports.[18] Since neither device rigidly couples to the head, the magnitude of these errors seems reasonable and could be included in some analyses that rely on these data. At the facemask, HITS routinely reported PLAs that were over twice the reference values. Others have observed a similar pattern at the facemask[4,18] and attributed it to decoupling of the HITS sensors from the head.[4] Field data from HITS do not suggest that front impacts, which presumably capture most facemask impacts, generate higher PLAs than other sites.[8] Thus more work is needed to assess the field relevance of this phenomenon, particularly since 14 of 25 (56%) concussions reported in NFL players resulted from facemask impacts.[33]

Despite not meeting our validity criteria, both sensors can provide useful information about the number, direction and magnitude of head impacts. Based on our results, analyses of impact data could focus on valid impact sites and exclude invalid sites. For instance, injury risk functions calculated from side impacts may be more reliable than those calculated from facemask impacts. Knowing the regional validity and errors of these devices can help researchers better interpret analyses based on their data. Moreover, the

results presented here could also be used to develop a calibration model for each sensor.

Despite a more comprehensive protocol than prior validation studies, our study has limitations. We used a modified Hybrid III head and standard neck that has no hair, sweat, or nape. The RR site was close to the lower edge of the MLSH's occiput (Fig. 8) and the absence of a bearing surface to simulate the soft tissues of the human nape may have altered helmet/head interaction compared to a human and contributed to the impact direction errors for the RR site. However, the lack of a complete bearing surface does not explain the larger proportion of impact direction errors observed at the CR site, a region remote from an edge and where the headform's geometry is humanlike (Fig. 8). The absence of a nape might also be related to the missed impacts at RR and RB sites (see Table 1 in the Supplementary Materials). Moreover, the Hybrid III neck has limited ability to translate horizontally and therefore yields impact directions that are biased toward the horizontal plane. The test-to-test repeatability of the combined MLSH and linear impactor platform (without a helmet) at six of our twelve impact sites is better for PLA (coefficients of variation = $2.0 \pm 1.2\%$) than for PAA $(8.3 \pm 5.8\%)$.[38] With the addition of a helmet and hard impactor cap in the current study, within-site variability increased for PLA ($4.4 \pm 1.2\%$), but remained similar for PAA ($8.0 \pm 3.5\%$). Variability of the MLSH's accelerometers is likely much lower than the test-to-test variability of the entire platform (with or without the helmet), and since we performed only within-test comparisons between the wearable and reference sensors, we believe the MLSH sensors provided a reliable reference for these validation tests. Nevertheless, the ability of the MLSH and linear impactor to simulate important features of on-field use and performance remains unknown. We also used only one common helmet, facemask and chinstrap. Given the different performance we saw from limited impacts to a different helmet, more work is needed to assess the validity of other equipment combinations. Our analysis focused only on PLA, PAA and HIC because both systems generated these parameters and they have been cited most widely. Other parameters, in particular angular velocity change,[40] should also be considered in future work. Also, our results were governed by variability in the data rather than specific error levels needed for biomechanical or clinical decision-making. Despite these limitations, we believe that valid sensor systems should perform well in a controlled testing environment and that the addition of many uncontrolled on-field variables is unlikely to improve their on-field performance compared to the laboratory.

In summary, we found that HITS and X2 mouthguards detected more than 95% of impacts between 3.6 and 11.2 m/s at the 12 impact sites we tested. However, neither system accurately estimated the direction and magnitude for all 12 impact sites and 5 impact speeds for the helmet and headform combination we tested. Despite these limitations, both systems provide useful data, which combined with the error data presented here, could be used to examine a number of important issues related to player safety and justify rule changes and improved helmet design.

ELECTRONIC SUPPLEMENTARY MATERIAL

The online version of this article (doi: 10.1007/s10439-015-1420-6) contains supplementary material, which is available to authorized users.

REFERENCES

[1]Allison, M. A., Y. S. Kang, M. R. Maltese, J. H. Bolte, and K. B. Arbogast. Validation of a helmet-based system to

measure head impact biomechanics in ice hockey. *Med. Sci. Sports Exerc.* 46:115–123, 2013.

[2]Bain, A. C., and D. F. Meaney. Tissue-level thresholds for axonal damage in an experimental model of central nervous system white matter injury. *J. Biomech. Eng.* 122:615–622, 2000.

[3]Bayly, P. V., T. S. Cohen, E. P. Leister, D. Ajo, E. C. Leuthardt, and G. M. Genin. Deformation of the human brain induced by mild acceleration. *J. Neurotrauma* 22:845–856, 2005.

[4]Beckwith, J. G., R. M. Greenwald, and J. J. Chu. Measuring head kinematics in football: correlation between the head impact telemetry system and hybrid III headform. *Ann. Biomed. Eng.* 40:237–248, 2012.

[5]Camarillo, D. B., P. B. Shull, J. Mattson, R. Shultz, and D. Garza. An instrumented mouthguard for measuring linear and angular head impact kinematics in American football. *Ann. Biomed. Eng.* 41:1939–1949, 2013.

[6]Cobb, B. R., A. MacAlister, T.J. Young, A. R. Kemper, S. Rowson, and S. M. Duma. Quantitative comparison of Hybrid III and National Operating Committee on Standards for Athletic Equipment headform shape characteristics and implications on football helmet fit. In: Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology. 229, 2015, pp. 39–46.

[7]Crisco, J. J., J. J. Chu, and R. M. Greenwald. An algorithm for estimating acceleration magnitude and impact location using multiple nonorthogonal single-axis accelerometers. *J. Biomech. Eng.* 126:849–854, 2004.

[8]Crisco, J. J., B. J. Wilcox, J. G. Beckwith, J. J. Chu, A.-C. Duhaime, S. Rowson, S. M. Duma, A. C. Maerlander, T. W. McAllister, and R. M. Greenwald. Head impact exposure in collegiate football players. *J. Biomech.* 44:2673–2678, 2011.

[9]Duma, S. M., S. J. Manoogian, W. R. Bussone, P. G. Brolinson, M. W. Goforth, J. J. Donnenwerth, R. M. Greenwald, J. J. Chu, and J. J. Crisco. Analysis of real-time head accelerations in collegiate football players. *Clin. J. Sport Med.* 15:3–8, 2005.

[10]Funk, J. R., S. M. Duma, S. J. Monoogian, and S. Rowson. Biomechanical risk estimates for mild traumatic brain injury. *Annu. Proc. Assoc. Adv. Automot. Med.* 51:343–361, 2007.

[11]Geddes-Klein, D. M., K. Schiffman, and D. Meaney. Mechanisms and consequences of neuronal stretch injury in vitro differ with the model of trauma. *J. Neurotrauma* 23:193–204, 2006.

[12]Gennarelli, T.A., L.E. Thibault, G. Tomei, R. Wiser, D. Graham and J. Adams. Directional dependence of axonal brain injury due to centroidal and non-centroidal acceleration (872197). Proceedings of 31st Stapp Car Crash Conference, 1987, pp. 49–53.

[13]Greenwald, R. M., J. T. Gwin, J. J. Chu, and J. J. Crisco. Head impact severity measures for evaluating mild traumatic brain injury risk exposure. *Neurosurgery* 62:789–798, 2008.

[14]Guskiewicz, K. M., J. P. Mihalik, V. Shankar, S. W. Marshall, D. H. Crowell, S. M. Oliaro, M. F. Ciocca, and D. N. Hooker. Measurement of head impacts in collegiate football players—relationship between head impact biomechanics and acute clinical outcome after concussion. *Neurosurgery* 61:1244–1253, 2007.

[15]Hardy, W. N., M. J. Mason, C. D. Foster, C. S. Shah, J. M. Kopacz, K. H. Yang, A. I. King, J. Bishop, M. Bey, W.

Anderst, and S. Tashman. A study of the response of the human cadaver head to impact. *Stapp Car Crash J.* 51:17–80, 2007.

[16]Harmon, K. G., J. Drezner, M. Gammons, K. Guskiewicz, M. Halstead, S. Herring, J. Kutcher, A. Pana, M. Putukian, and W. Roberts. American medical society for sports medicine position statement - Concussion in sport. *Clin. J. Sport Med.* 23:1–18, 2013.

[17]Higgins, M., P. D. Halstead, L. Snyder-Mackler, and D. Barlow. Measurement of impact acceleration: mouthpiece versus helmet accelerometer. *J. Athl. Train.* 42:5–10, 2007.

[18]Jadischke, R., D. C. Viano, N. Dau, A. I. King, and J. McCarthy. On the accuracy of the head impact telemetry (HIT) System used in football helmets. *J. Biomech.* 46:2310–2315, 2013.

[19]Ji, S., H. Ghadyani, R. P. Bolander, J. G. Beckwith, J. C. Ford, T. W. McAllister, L. A. Flashman, K. D. Paulsen, K. Ernstrom, S. Jain, R. Raman, L. Zhang, and R. M. Greenwald. Parametric comparisons of intracranial mechanical responses from three validated finite element models of the human head. *Ann. Biomed. Eng.* 42:11–24, 2013.

[20]Kent, J. T. The Fisher-Bingham distribution on the sphere. *J. R. Statist. Soc. B* 44:71–80, 1982.

[21]Knox, T, J. Pellettiere, C. Perry, J. Plaga, and J. Bonfeld. New sensors to track head acceleration during possible injurious events (2008-01-2976). SAE International Journal of Passenger Cars—Electronic and Electrical Systems 1, 2008, pp. 652–663.

[22]Laplaca, M. C., V. M.-Y. Lee, and L. W. Thibault. An in vitro model of traumatic neuronal injury - Loading rate-dependent changes in acuty cytosolic calcium and lactate dehydrogenase release. *J Neurotrauma* 14:355–368, 1997.

[23]Leong, P., and S. Carlile. Methods for spherical data analysis and visualization. *J. Neurosci. Methods* 80:191–200, 1998.

[24]Manoogian, S., D. McNeely, S. Duma, G. Brolinson, and R. Greenwald. Head acceleration is less than 10 percent of helmet acceleration in football impacts. *Biomed. Sci. Instrum.* 42:383–388, 2006.

[25]McAllister, T. W., J. C. Ford, S. Ji, J. G. Beckwith, L. A. Flashman, K. Paulsen, and R. M. Greenwald. Maximum principal strain and strain rate associated with concussion diagnosis correlates with changes in corpus callosum white matter indices. *Ann. Biomed. Eng.* 40:127–140, 2011.

[26]McCulloch, C. E., S. R. Searle, and J. M. Neuhaus. Generalized, Linear, and Mixed Models (2nd ed.). Hoboken: Wiley, p. 424, 2008.

[27]Mihalik, J. P., D. R. Bell, S. W. Marshall, and K. M. Guskiewicz. Measurements of head impacts in collegiate football players: An investigation of positional and event-type differences. *Neurosurgery* 61:1229–1235, 2007.

[28]Newman, J. A., M. C. Beusenberg, N. Shewchenko, C. Withnall, and E. Fournier. Verification of biomechanical methods employed in a comprehensive study of mild traumatic brain injury and the effectiveness of American football helmets. *J. Biomech.* 38:1469–1481, 2005.

[29]Olvey, S. E., T. Knox, and K. A. Cohn. The development of a method to measure head acceleration and motion in high-impact crashes. *Neurosurgery* 54:672–677, 2004.

[30]Ommaya, A. K. Head injury mechanisms and the concept of preventive management: a review and critical synthesis. *J. Neurotrauma* 12:527–546, 1995.

[31]Padgaonkar, A. J., K. W. Krieger, and A. I. King. Measurement of angular acceleration of a rigid body using

linear accelerometers. *ASME J. Appl. Mech.* 42:552–558, 1975.

[32]Pellman, E. J., D. C. Viano, A. M. Tucker, I. R. Casson, and J. F. Waeckerle. Concussion in professional football, reconstruction of game impacts and injuries. *Neurosurgery* 53:799–814, 2003.

[33]Pellman, E. J., D. C. Viano, A. M. Tucker, and I. R. Casson. Concussion in professional football: location and direction of helmet impacts—Part 2. *Neurosurgery* 53:1328–1341, 2003.

[34]Pellman, E. J., D. C. Viano, C. Withnall, N. Schewchenko, C. A. Bir, and P. D. Halstead. Concussion in professional football—Helmet testing to assess impact performance—Part 11. *Neurosurgery* 58:78–96, 2006.

[35]Rowson, S., J. G. Beckwith, J. J. Chu, D. S. Leonard, R. M. Greenwald, and S. M. Duma. A six degree of freedom head acceleration measurement device for use in football. *J. Appl. Biomech.* 27:8–14, 2011.

[36]Rowson, S., S. M. Duma, J. G. Beckwith, J. J. Chu, R. M. Greenwald, J. J. Crisco, P. G. Brolinson, A.-C. Duhaime, T. W. McAlister, and A. C. Maerlender. Rotational head kinematics in football impacts: an injury risk function for concussion. *Ann. Biomed. Eng.* 40:1–13, 2012.

[37]Rowson, S., S. M. Duma, R. M. Greenwald, J. G. Beckwith, J. J. Chu, K. M. Guskiewicz, J. P. Mihalik, J. J. Crisco, *et al.* Can helmet design reduce the risk of concussion in football? *J. Neurosurg.* 120:919–922, 2014.

[38]Siegmund, G. P., K. M. Guskiewicz, S. W. Marshall, A. L. DeMarco, and S. J. Bonin. A headform for testing helmet and mouthguard sensors that measure head impact severity in football players. *Ann. Biomed. Eng.* 42:1834–1845, 2014.

[39]Stroup, W. W. Generalized Linear Mixed Models: Modern Concepts, Methods and Applications. Boca Raton: CRC Press, p. 555, 2012.

[40]Takhounts, E. G., M. J. Craig, K. Moorhouse, J. McFadden, and V. Hasija. Development of brain injury criteria (Br IC). *Stapp Car Crash J.* 57:243–266, 2013.

[41]Viano, D. C., I. R. Casson, and E. J. Pellman. Concussion in professional football—Biomechanics of the struck player—Part 14. *Neurosurgery* 61:313–328, 2007.

[42]Viano, D. C., C. Withnall, and M. Wonnacott. Effect of mouthguards on head responses and mandible forces in football helmet impacts. *Ann. Biomed. Eng.* 40:47–69, 2012.

[43]Weaver, A. A., K. A. Danelson, and J. D. Stitzel. Modeling brain injury response for rotational velocities of varying directions and magnitudes. *Ann. Biomed. Eng.* 40:2005–2018, 2012.

[44]Wonnacott, M., and C. Withnall. Development of an articulating mandible headform having force sensing temporomandibular joints (2010-01-0136). Society of Automotive Engineers, 2010.

[45]Zhang, L., K. H. Yang, and A. I. King. A proposed injury threshold for mild traumatic brain injury. *J. Biomech. Eng.* 126:226–236, 2004.