




Development and Evaluation of a Test Method for Assessing the Performance of American Football Helmets

ANN M. BAILEY ¹, ERIN J. SANCHEZ,¹ GWANSIK PARK,¹
LEE F. GABLER,¹ JAMES R. FUNK,¹ JEFF R. CRANDALL,¹ MICHAEL WONNACOTT,²
CHRIS WITHNALL,² BARRY S. MYERS,³ and KRISTY B. ARBOGAST⁴

¹Biomechanics Consulting and Research, LLC, Charlottesville, VA, USA; ²Biokinetics and Associates, Ltd., Ottawa, ON, Canada; ³Department of Biomedical Engineering, Duke University, Durham, NC, USA; and ⁴Center for Injury Research and Prevention, Children's Hospital of Philadelphia, Pennsylvania, USA

(Received 20 August 2020; accepted 17 September 2020; published online 6 October 2020)

Associate Editor Stefan M Duma oversaw the review of this article.

Abstract—As more is learned about injury mechanisms of concussion and scenarios under which injuries are sustained in football games, methods used to evaluate protective equipment must adapt. A combination of video review, videogrammetry, and laboratory reconstructions was used to characterize concussive impacts from National Football League games during the 2015–2017 seasons. Test conditions were generated based upon impact locations and speeds from this data set, and a method for scoring overall helmet performance was created. Head kinematics generated using a linear impactor and sliding table fixture were comparable to those from laboratory reconstructions of concussive impacts at similar impact conditions. Impact tests were performed on 36 football helmet models at two laboratories to evaluate the reproducibility of results from the resulting test protocol. Head acceleration response metric, a head impact severity metric, varied 2.9–5.6% for helmet impacts in the same lab, and 3.8–6.0% for tests performed in a separate lab when averaged by location for the models tested. Overall inter-lab helmet performance varied by $1.1 \pm 0.9\%$, while the standard deviation in helmet performance score was 7.0%. The worst helmet performance score was 33% greater than the score of the best-performing helmet evaluated by this study.

Keywords—Concussion, Helmet, Kinematics, Biomechanics, Football.

INTRODUCTION

A variety of test methods have been developed to assess the ability of American football helmets to mitigate the severity of head impacts. In the 1970s, the National Operating Committee on Standards for Athletic Equipment (NOCSAE) developed a headform drop test standard for American football helmets with a focus on preventing catastrophic injuries (e.g., skull fracture) by limiting translational head acceleration.^{20,25,26} More recent concerns about concussion have motivated the inclusion of both linear and rotational head kinematics in the evaluation of helmets.^{20,25,35} Pellman *et al.* developed a test fixture and procedure that evaluated a helmet's ability to mitigate both translational and rotational head motion.²² In their test methodology, a pneumatic impactor arm struck a helmeted Hybrid III dummy head and neck that were mounted to a sliding table. Viano *et al.* developed a test matrix to assess the performance of football helmets over a range of impact speeds and orientations that represented a set of concussions and severe impacts occurring in the NFL.^{34,35} NOCSAE recently adopted a modified version of that test methodology as an additional test standard and now requires newly manufactured helmets to limit both the Severity Index and peak rotational acceleration measures.²⁰ Likewise, Virginia Tech's Summation of Tests for the Analysis of Risk (STAR) evaluation system for varsity football helmets has transitioned from a drop test based on the original NOCSAE test methodology to a pendulum impactor test using a NOCSAE head-

Address correspondence to Ann M. Bailey, Biomechanics Consulting and Research, LLC, Charlottesville, VA, USA. Electronic mail: abailey@bicorellc.com

form with a flexible neck.^{25,32} All of these methodologies currently utilize evaluation metrics that include rotational acceleration as well as translational kinematic measures. The NOCSAE standards are pass/fail to ensure all helmets exceed a minimum level of protection. In addition to NOCSAE, Virginia Tech has ranked football helmets for consumer information, using impact data obtained from play at the collegiate level to inform the scoring criteria.^{7,25,32}

Recent advances in data collection and characterization of professional football,^{2,3,13,17} created an opportunity to update the existing impactor testing method used to assess the performance of helmets worn by NFL players. The objective was to test and rank helmet models based on their ability to reduce the severity of impacts representative of those that caused concussions in NFL games between 2015 and 2017. Specifically, four main objectives were identified for the study: (1) identify test conditions based upon impact locations and velocities associated with concussion incidence in the NFL; (2) develop a laboratory test methodology to rank overall helmet performance; (3) quantify the reproducibility of that test methodology; and (4) assess the ability of the test method to mimic the on-field impact environment.

MATERIALS AND METHODS

The methods for this study have been divided into two sections. The first section outlines the development of the test methodology while the second section focuses on the evaluation of the methodology.

Test Methodology Development

A previously designed test fixture and methodology for quantifying the ability of football helmets to mitigate translational and rotational head kinematics was modified to better represent the current on-field impact environment during NFL games.^{22,35} A combination of video review,¹⁷ videogrammetry,^{2,3} laboratory reconstructions,¹³ and Next Generation Stats (NGS) field position and velocity data were employed to characterize concussion-causing impacts from NFL game concussions from the 2015-2017 seasons documented using the NFL Game Day Concussion Diagnosis and Management Protocol.¹²

The test fixture was comprised of a pneumatic impactor with a slider table and carriage oriented parallel to the ram to allow translation in the direction of impact. A Hybrid III head and neck attached to a pivot that, in turn, connected to the carriage and could be rotated about a horizontal axis perpendicular to the direction of travel. The head and neck could also be

rotated about the long axis of the neck. These two angular adjustments allowed the head to be positioned in a wide range of orientations relative to the impactor ram. The entire table could be raised and lowered as well as adjusted laterally. The impactor ram interface (henceforth “end cap”) was constructed from a cylindrical vinyl nitrile puck with a spherical nylon face attached to the ram face using Velcro.^{22,35}

In the impactor test method, the ram represents the collision partner in a helmet-to-helmet collision, while the test dummy mounted to a sliding table exemplifies the player experiencing an impact. To characterize these impacts, impact locations for concussive impacts in the NFL were studied qualitatively through video review by visually classifying the location according to nine impact regions (i.e. side (upper), side (lower), facemask (side edge), facemask (central), facemask (upper edge), front, rear (upper), rear (lower), top).¹⁶ The concussive impacts were also assessed quantitatively through videogrammetry, with model-based image matching used to estimate helmet kinematics from NFL game footage.³ Quantitative helmet impact locations for 57 concussive impacts in NFL games were obtained using model-based image matching in conjunction with three-dimensional computer models of the helmets to find the location of first contact on the helmet.³ Three-dimensional heat maps based on the density of impacts within a given area on the helmet’s surface (Fig. 1), coupled with the coarser data on overall distribution of concussive impacts by region from video review were used to define new impact locations for the test protocol.^{3,17}

Closing velocity vectors (i.e., the difference between the concussed player and collision partner velocity vectors just prior to impact) from the videogrammetry study³ were calculated and then grouped based on impact source (i.e., the ground, or another player’s helmet or body) and helmet impact regions from the video review.¹⁷ An initial set of impact directions was selected by averaging the unit vectors from these groups. A pilot study was performed on four helmet models using the impactor test setup to assess the test fixture’s ability to achieve and to replicate these impact vectors. These preliminary tests revealed four shortcomings, each described in detail below: (1) inability of the test fixture to replicate impacts to the top of the helmet, (2) non-biofidelic behavior of the Hybrid III neck in axial loading, (3) unrealistic interactions between the helmet and the end cap, and (4) potential for test variability due to inconsistency in head and helmet positioning.

The first shortcoming of the test fixture was that its limited range of adjustability to permit testing at all possible impact locations. For example, it was not possible to rotate the headform sufficiently forward in

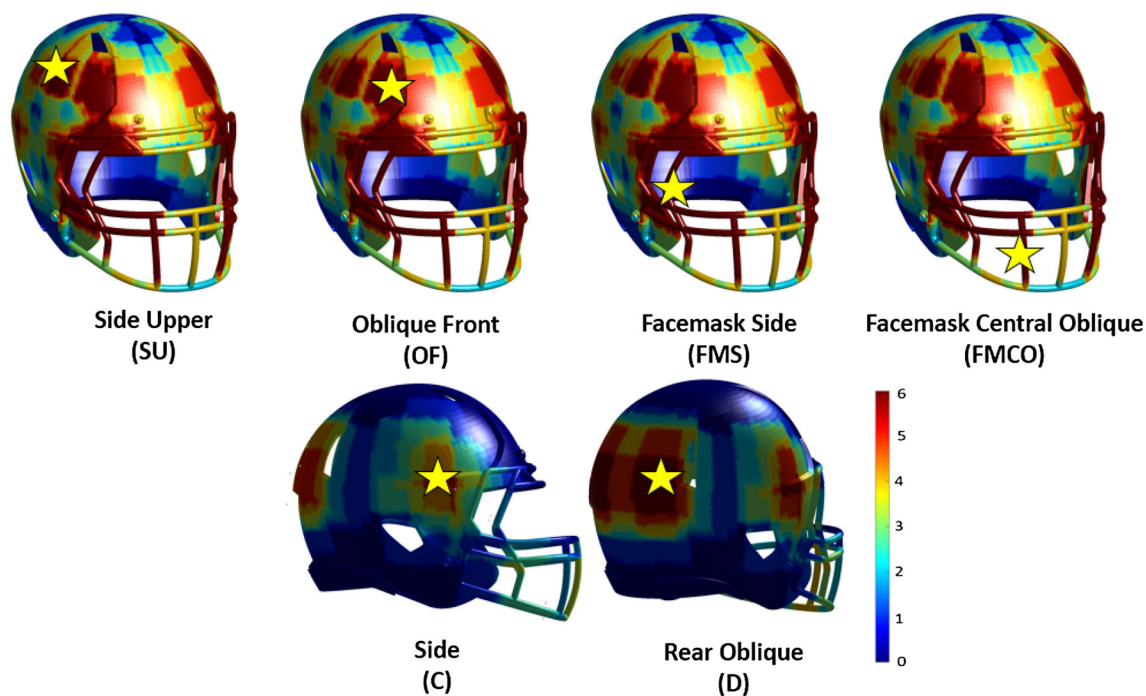


FIGURE 1. Heat maps of impact locations used to define impact locations, based on results from a previous videogrammetry study.³ The top row was based on helmet-to-helmet concussion impact locations and bottom row was based on helmet-to-ground concussion impact locations, with the color scale indicating the number of impacts within 25° of azimuth or elevation on the helmet's surface.

order for the impactor to strike the top area of the helmet without the facemask interacting with the slider table. Second, the Hybrid III neck responded in a non-biofidelic manner to impacts with a significant downward component (+z axis of the neck). In such impacts, the rubber discs in the dummy neck were compressed axially during loading by the impactor but unloaded as the head rebounded often in a different direction. At the end of the unloading phase, the rebounding dummy head was abruptly stopped by tension in the steel neck cable. Since this problem proved intractable for any appreciable vertical load components, impact vectors were modified to reduce axial neck loading, while remaining within the range of impact locations identified in the videogrammetry study. A third problem occurred under impact scenarios where the edge of the end cap occasionally “caught” the edge of the helmet or facemask. Since this behavior was not considered representative of the on-field conditions, this problem was addressed by redesigning the end cap with a larger face.⁵ The fourth shortcoming related to variations in head and helmet positioning was addressed by adding headform reference lines to verify head position relative to the ram, including supplemental helmet positioning references, and generating system response corridors.⁵

After extensive pilot testing, six impact locations were chosen (Fig. 1) and are further described in the

supplemental material. Four impact locations (i.e. side upper (SU), oblique front (OF), facemask side (FMS), and facemask central oblique (FMCO)) were selected to represent the range of common impact sites in concussive helmet-to-helmet impacts. Two other impact locations, the side (C) and oblique rear (D) locations, were chosen to represent the locations of common helmet-to-ground impact sites.^{3,35} Since there is currently no accepted test methodology to assess helmet-to-ground impacts, these additional impact locations were included with the aim of providing a more complete assessment of overall helmet performance relative to on-field usage.

The impactor speed is meant to replicate the closing velocity between two players' helmets. In a previous laboratory study of NFL impacts,²² the impactor test speeds were chosen based on the mean and standard deviation of the closing speed for players who sustained concussions. The same approach was used in the present study. A summary of head and helmet kinematics observed in studies of NFL concussive impacts was tabulated (Table 1). All studies indicated similar average closing velocities for the concussive impacts (Fig. 2a). One previous study of concussive and severe impacts in the NFL from 1996 to 2001 used videogrammetry to determine an average impact closing velocity of $9.3 \pm 1.9 \text{ m s}^{-1}$.^{21,22,35} Based on a more recent videogrammetric study of 57 cases of NFL

TABLE 1. Summary of data used to develop test conditions.

Data Source	Impact Source	# Cases	Impact Velocity $m s^{-1}$	Change in Translational Helmet Velocity $m s^{-1}$	Change in Rotational Helmet Velocity $rad s^{-1}$	Peak Head Translational Acceleration g	Peak Head Rotational Velocity $rad s^{-1}$	Peak Head Rotational Acceleration $rad s^{-2}$
2015–2017 video reconstructions	Helmet	30	9.6 ± 1.9	5.9 ± 1.5	33 ± 11	–	–	–
	Body	16	8.4 ± 2.0	6.1 ± 2.9	21 ± 7	–	–	–
	Ground	11	8.3 ± 1.9	8.1 ± 1.7	28 ± 17	–	–	–
2015–2016 NGS tracking	Helmet	101	8.7 ± 2.6	–	–	–	–	–
	Body	132	8.2 ± 2.7	–	–	–	–	–
	All	233	8.4 ± 2.6	–	–	–	–	–
1996–2001 laboratory reconstructions ^{21,28}	All	22	9.3 ± 1.9	–	–	94 ± 27	43 ± 11	6788 ± 1840
2015–2017 laboratory reconstructions ¹³	Helmet	18	9.6 ± 1.5	6.3 ± 1.5	34 ± 10	84 ± 31	39 ± 11	6186 ± 2345

Averages and standard deviations for head and helmet kinematics measures for concussed players are presented.

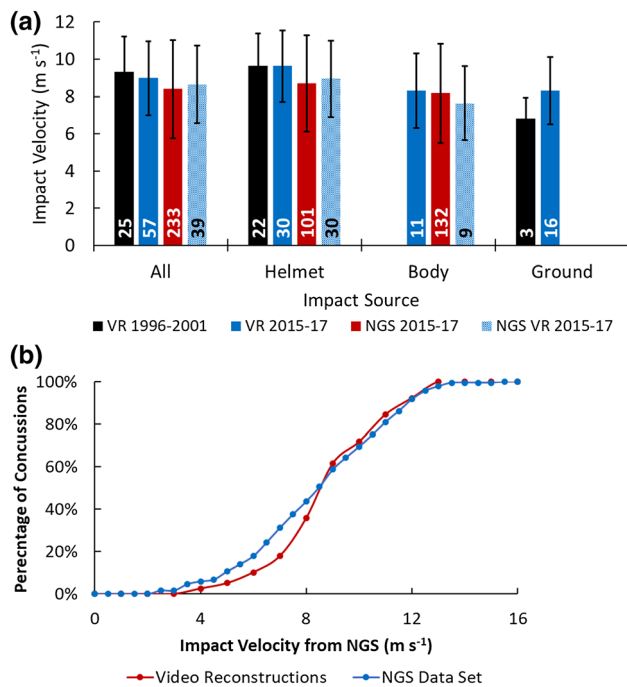


FIGURE 2. (a) Comparison of average and standard deviation of impact velocities by impact source and data source for NFL concussions. Data is provided for video reconstructions performed by Pellman *et al.* (VR 1996–2001) and by Bailey *et al.* (VR 2015–17), as well as velocities calculated from NGS data for all concussions for which NGS data existed (NGS 2015–17) and for the subset of cases studied in the video reconstruction data set (NGS VR 2015–17). Number of data points within each category is noted. (b) Cumulative density plots for impact velocity for concussive helmet-to-helmet and helmet-to-body impacts from the 2015–2016 NGS data set and those included in the videogrammetry study.³

concussions, the average closing velocity was $8.9 \pm 2.0 m s^{-1}$. In addition, we analyzed data from Next Generation Stats (NGS), which provided 10 Hz position data in the horizontal plane derived from RFID chips mounted in the players' shoulder pads. The vertical velocity was not measured but in a previous videogrammetry study was found to contribute less than $1 m s^{-1}$ on average for helmet-to-helmet impacts, which is considerably less than either the horizontal or vertical components.³ In addition, the NGS data represented upper torso velocity rather than the helmet velocity measured with videogrammetry but permitted a much larger data set of impact conditions. According to the NGS data, the average closing velocity was $8.4 \pm 2.6 m s^{-1}$ for 233 concussive helmet-to-helmet ($n = 101$) and helmet-to-body impacts ($n = 132$) (Fig. 2b). The recent results were sufficiently similar to the previous test conditions that a decision was made to retain three of the test speeds used in the earlier study: 9.3 , 7.4 , and $5.5 m s^{-1}$, which represented the average speed, and one and two standard deviations below the average closing speed, respectively.³⁵ For perspective, the average closing velocity calculated from NGS data for all helmet impacts (i.e., injurious and non-injurious) identified by video review of 16 NFL games was $3.4 \pm 1.6 m s^{-1}$. Exploratory impactor tests at $11.2 m s^{-1}$, the highest speed used in the earlier study, produced unrealistic motion of the dummy head and helmet which had the potential to damage the test equipment. Therefore, further testing at $11.2 m s^{-1}$ was abandoned.

Laboratory Performance Metric

Currently, there is no consensus as to which mechanical parameters of head response best predict concussion risk. Brain strain, which is generally assumed to be a measure of concussion risk, has demonstrated correlation with the magnitude, direction, and duration of rotational head acceleration.^{11,15,31} Many researchers have postulated multiple mechanisms of concussion and have suggested that both translational and rotational head kinematics be used to predict injury.²⁶ In order to combine these parameters into a single parameter which could be used to evaluate the severity of an impact, the Head Acceleration Response Metric (HARM) was created. Diffuse axonal multi-axis general evaluation (DAMAGE) was chosen as the rotational injury metric to be included in HARM. DAMAGE predicts brain strain from angular head acceleration time histories relative to head coordinate directions using a lumped-parameter model. It was developed as a simpler alternative to finite element modeling (FEM) of the brain, and its output has been shown to be highly correlated with FEM results and is dependent upon both direction and duration of loading.¹⁵ To select a translational injury metric for inclusion in HARM, a regression analysis was performed in which the matched sets of injury and non-injury head kinematics from dummy reconstructions of on-field football impacts were used to fit coefficients for a linear combination of DAMAGE and a selection of existing translational metrics.^{13,14,28} This analysis showed that Head Injury Criterion (HIC) and DAMAGE provided the best ability to distinguish between injurious and non-injurious head kinematics (Eq. 1).

$$HARM = C_1 HIC + C_2 DAMAGE \quad (1)$$

where $C_1 = 0.0148$ and $C_2 = 15.6$ are constants that were determined from fits to head kinematics measured in physical dummy reconstructions.^{13,14,21,28} Note that a lower HARM value indicates a less severe impact and a better helmet performance.

In order to assess overall helmet performance across all tests, a metric called the Helmet Performance Score (HPS) was developed as the weighted sum of HARM scores at each test condition:

$$HPS = \sum_{i=1}^{18} M_i (HARM)_i \quad (2)$$

The weights M_i were based upon the incidence of concussive impacts in NFL games by impact location and speed. Further details are provided by Gabler *et al.*¹⁴ and are provided in the helmet test protocol.⁵

The formulation of HPS involves calculating the HARM for each test condition by using a weighting

coefficient (M_i) and then summing across 18 test conditions based on the combination of the six impact locations and three velocities (Eq. 2, Table 2). A weight was assigned to each test condition to reflect the on-field incidence of concussion by location and speed (See supplemental material). In order to normalize the severity of individual test conditions and ensure that the desired weighting effect was achieved, the weighting coefficients (M_i) were derived by dividing the test condition weight by the average HARM from a sample of previously tested helmets at that test condition.

Evaluation of the Test Method

Laboratory testing was performed on 36 helmet models used by NFL players. The smallest size which encompassed the dimensions of the Hybrid III headform based on manufacturer-provided sizing charts was tested for each model so long as the pressure between the helmet and headform did not exceed 70 kPa.⁵ All helmets were size large except for the Vicis Zero1, which was a size B. Helmets with customizable fits were customized to fit the Hybrid III head geometry.

Testing was performed in two laboratories (Biokinetics and Associates Ltd., Ottawa, ON, Canada and Biocore LLC, Charlottesville, VA, USA) using the aforementioned test apparatus and test conditions. The pneumatic impactor (Biokinetics and Associates Ltd., Ottawa, ON, Canada) with a total ram mass of 15.6 ± 0.5 kg was propelled toward a Hybrid III 50th male head and neck attached to the slider table (Biokinetics and Associates Ltd., Ottawa, ON, Canada) (Fig. 3a). A 6-axis Hybrid III femur load cell (N6ACC11A, MG Sensor, Rheinmuenster, Germany) was attached in line with the impactor shaft to measure impact forces. The newly-designed impactor end cap was spherically shaped (140 mm radius sphere, 203 mm diameter cap) and was fabricated from extruded nylon 6/6. The end cap was attached with Velcro to a cylindrical (41.3 mm thick, 127 mm diameter) piece of vinyl nitrile foam (VN600, DerTex, Inc., Saco, Maine), which in turn was attached with Velcro to a circular (127 mm diameter) metal backing plate on the ram (Fig. 3a). The 50th percentile male Hybrid-III test dummy head was instrumented using a nine-accelerometer package (NAP) (Endevco 7264B-2000, Meggitt, Irvine, CA) in Lab 1, and a 6DX Pro with an array of six collinear accelerometers (Endevco 7264B-2000, Meggitt, Irvine, CA; 6DX-Pro-2000-18000, Diversified Technical Systems, Seal Beach, CA) was used in Lab 2. These sensor packages were previously shown to provide statistically similar results.⁹ Post-impact travel of the ram was limited to

TABLE 2. Coefficients (M_i) used to weight the contribution of each test condition in the calculation of the Helmet Performance Score (HPS).

Impact Location	Impact Velocity		
	5.5 m s ⁻¹	7.4 m s ⁻¹	9.3 m s ⁻¹
Side upper (SU)	2.93e-2	1.82e-2	2.33e-2
Oblique front (OF)	9.49e-3	6.01e-3	7.82e-3
Rear oblique (D)	5.83e-3	3.71e-3	4.72e-3
Side (C)	9.96e-3	6.24e-3	8.13e-3
Facemask side (FMS)	3.92e-3	2.58e-3	3.58e-3
Facemask central oblique (FMCO)	6.87e-3	4.93e-3	6.87e-3

170 ± 5 mm and laboratory temperature was kept constant at 22 ± 2 °C.

Impact locations (Fig. 3b) were achieved through a series of translations and rotations of the components of the slider table.⁵ Additional procedures aimed at improving the repeatability and reproducibility of tests results included checks on head position and the use of two reference dimensions for positioning the helmet (Fig. 3c).⁵ Consistency verifications of the recorded head kinematics used the redundant sensors in the accelerometer arrays to cross check the calculation of kinematics resulting from the different sensor arrays.^{28,30}

Two helmet samples, henceforth samples A and B, from each helmet model were tested in Lab 1. Due to the potential for deformation with impacts to the FMS and FMCO locations, the facemask and chin strap were replaced after the 7.4 and 9.3 m s⁻¹ impacts to the facemask. For each impact location, the helmet model test order was randomized for the A samples. The B samples were tested in reverse order from the A samples. Testing was repeated at the second laboratory using a portion of the overall test matrix. One sample of each of the 36 helmet models was tested at all six impact locations at the 7.4 m s⁻¹ impact speed. Similar to the testing in Lab 1, the facemasks of the samples tested in Lab 2 were replaced between the FMS and FMCO impacts.

Data Processing and Statistics

Sensor data was collected at 10 kHz in both labs and filtered using a channel frequency class (CFC) 180 filter.²⁹ Head sensor arrays measured six degree-of-freedom head kinematics that were transformed to the head center of gravity.¹⁰ Peak head kinematics, HIC, DAMAGE, and HARM were calculated for each test.^{14,15,33} The time at which the ram force returned to less than 20 N of compression was used as the truncation time for all peak kinematics and metric calculations except for HIC, which was truncated at 15 ms. The truncation times were less than 30 ms in nearly all

tests. HPS was calculated for each of the two helmet model samples tested at Lab 1 and then averaged to rank the overall performance of the helmet models tested.

The coefficient of variation in HARM was calculated for samples of the same model tested in the same lab, as well as those tested in a different lab. Helmet model variation (HMV), defined as the standard deviation of a metric across all helmet models tested divided by the average for all models tested, was calculated as a point of comparison to intra- and inter-lab test variation. To compare the performance of helmets tested in Lab 1 and Lab 2, a linear regression model was used to regress HARM against the fixed effects of helmet model and impact location. A random effect comprised of the interaction between test lab and impact location was included since inter-lab variation in test results was expected to be location-dependent given that each impact location involves a unique setup of the test device. Analysis of variance (ANOVA) was performed to estimate the significance of each fixed effect, and a log likelihood ratio test was used to estimate the significance of the random effect using R version 3.4.0 (R Foundation for Statistical Computing, Vienna, Austria). Significance levels were set to $\alpha = 0.05$. Since helmets were only tested at 7.4 m/s in Lab 2, only the 7.4 m/s tests were included in the data set for this analysis.

Evaluation of Test Fixture Response

The test fixture used in this study approximates the conditions of a helmet-to-helmet impact by incorporating several assumptions and simplifications. First, the test fixture limits the degrees of freedom (DOF) at the base of the neck to one DOF of translation. The impactor is also limited to one DOF translation. Other assumptions related to the stiffness of the end cap, the effective mass of the impactor and biofidelity of the Hybrid III head and neck further affect the test fixture's ability to represent the on-field impact environment. To benchmark

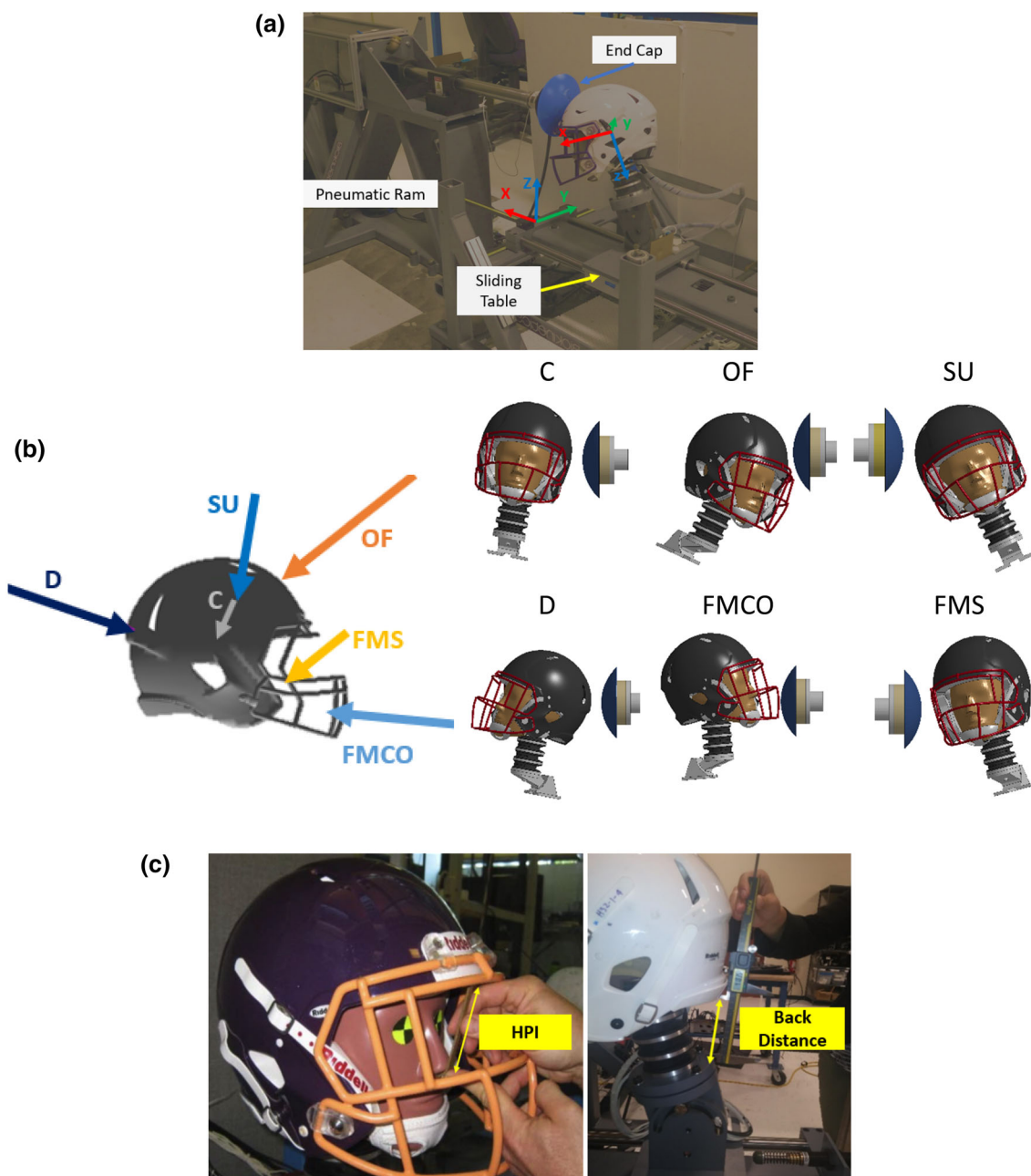


FIGURE 3. (a) Pneumatic ram and slider table used for laboratory testing. (b) Impact location vectors derived from videogrammetry data.³ (c) Demonstration of the use of the nose gauge and helmet positioning index (left). A secondary measurement (back distance) between the center of the lower back of the helmet shell and the neck base may be measured using calipers to ensure that the helmet is seated on the head repeatably (right).

test fixture response, head kinematics from the linear impactor test methodology were compared to head kinematics produced in laboratory reconstructions of on-field concussive impacts. In the dummy reconstructions, two partial 50th percentile Hybrid III test dummies were propelled into one another using electric belt-driven sleds.¹³ Peak head kinematics were plotted against closing velocity for both types of tests and compared. In addition, the range of peak head

kinematics at each impact location in the 9.3 ms^{-1} impactor tests was compared to reconstructions in which the dummy was struck in approximately the same helmet location at approximately the same closing velocity ($7.4\text{--}11.2 \text{ m s}^{-1}$). Two-sided, heteroscedastic Student's t-tests were used to compare HIC, DAMAGE, and peak translational and rotational accelerations by impact location to a significance level of 0.05.

RESULTS

A total of 1296 helmet tests were performed in Lab 1 (36 helmet models, 6 impact locations, and 3 impact velocities, and 216 tests were performed at Lab 2 (36 helmet models, 6 impact locations, and 1 impact velocity). In general, variation in helmet performance across helmet models was similar for different impact locations and speeds. The largest range in HARM for the models tested at these conditions occurred for the side (C) location at 9.3 m s^{-1} with a difference of 6.2 (HVM = 14%) between the best- and worst-performing helmet model. This was largely due to the variation of HIC across different models (HVM = 19%) (Fig. 4). The largest variation in DAMAGE across helmet models occurred for the face-mask central oblique location at 9.3 m s^{-1} (HVM = 16%) (Fig. 4).

The HPS for the worst-performing helmet tested was 33% greater than that of the best-performing helmet, which had an HPS 15% less than the average HPS for all helmets tested ($\text{HPS} = 1.03 \pm 0.07$) (Fig. 5a). Variation in HPS between samples of the same model tested at Lab 1 was $1.1 \pm 0.9\%$ on average, with a maximum difference of 4.3% between samples (Fig. 5b). Variation in HARM between samples *A* vs. *B* averaged $4.0 \pm 3.6\%$ across all test conditions, and was largest for the facemask impact locations FMS ($4.6 \pm 3.9\%$) and FMCO ($5.5 \pm 5.5\%$) (Fig. 5c).

The HARM difference for the average of samples tested in Lab 1 vs. the sample tested in Lab 2 was larger on average than the variation for samples of the same model tested in the same lab (Figs. 5d and 6). When HARM variation was averaged across all helmet models and all 7.4 m s^{-1} test conditions, inter-lab variation for HARM was $4.6\% \pm 3.5\%$ compared to intra-lab variation of $3.9 \pm 3.2\%$. The mixed regression model comparing HARM from helmets tested at 7.4 m/s performed at each of the two labs showed that helmet model ($p < 0.001$) and test condition ($p < 0.001$) were significant predictors of HARM, while the random effect of the interaction between test lab and impact location was not significant ($p = 0.19$). This result indicated that while inter-lab variation was higher on average than intra-lab variation, the difference in variation could not be systematically explained by test lab, even when assessed on an impact location basis.

Though a limited number of laboratory reconstructions were performed at closing velocities comparable to the two lower impactor test speeds, peak kinematics for the linear impactor tests and laboratory reconstructions showed similar trends relative to closing velocity (Fig. 7). The ranges of peak kinematics

from the impactor tests at 9.3 m s^{-1} were comparable to the range of peak kinematics from laboratory reconstructions of concussive impacts that occurred within one standard deviation of the average closing velocity of the impacts studied (Fig. 8). Similarly, the components used to calculate HARM (i.e., HIC and DAMAGE) were comparable for each impact location. Student's t-tests determined that only the DAMAGE for the OF impactor tests was significantly different from the dummy reconstructions ($p = 0.004$).

DISCUSSION

This study updates an existing linear impactor test method to reflect better the characteristics of concussive impacts from contemporary NFL games. The linear impactor method supplements the NOCSAE drop test because the flexible Hybrid III neck allows the head to rotate in response to impact. Helmet testing efforts as early as the 1970s recognized the importance of the use of a flexible neck and realistic boundary conditions to replicate head impact kinematics.¹ There is a general consensus in the biomechanical literature that rotational head kinematics play an important role in generating brain strain and causing concussion, so it is important that a helmet testing method be able to mimic rotational head motion in an impact.^{15,23,27,31}

The comparison of head kinematics produced by the linear impactor and the laboratory reconstructions of helmet-to-helmet impacts provides evidence that the test fixture and conditions proposed in this study produce representative impact responses of on-field concussive events. When grouped by impact location, the linear impactor tests at 9.3 m s^{-1} produced similar peak translational and rotational accelerations to those from the reconstructions (Fig. 8). The test method aimed to simulate primarily helmet-to-helmet impacts, which account for 36% of NFL game concussions during the 2015-2016 and 2016-2017 seasons.¹⁷ More accurate simulation of concussions caused by helmet-to-body and helmet-to-ground impacts may require different test fixtures or conditions.

Since the selection of new test conditions for the linear impactor protocol relied heavily upon data obtained from the videogrammetry study, it was important to ensure that the process of selecting impacts based on video clarity did not bias the sample towards open-field impacts that might involve higher closing velocities. The finding that the distribution of closing velocities was similar for the videogrammetry study and the larger NGS data set suggests that the data used to develop the test conditions in this study

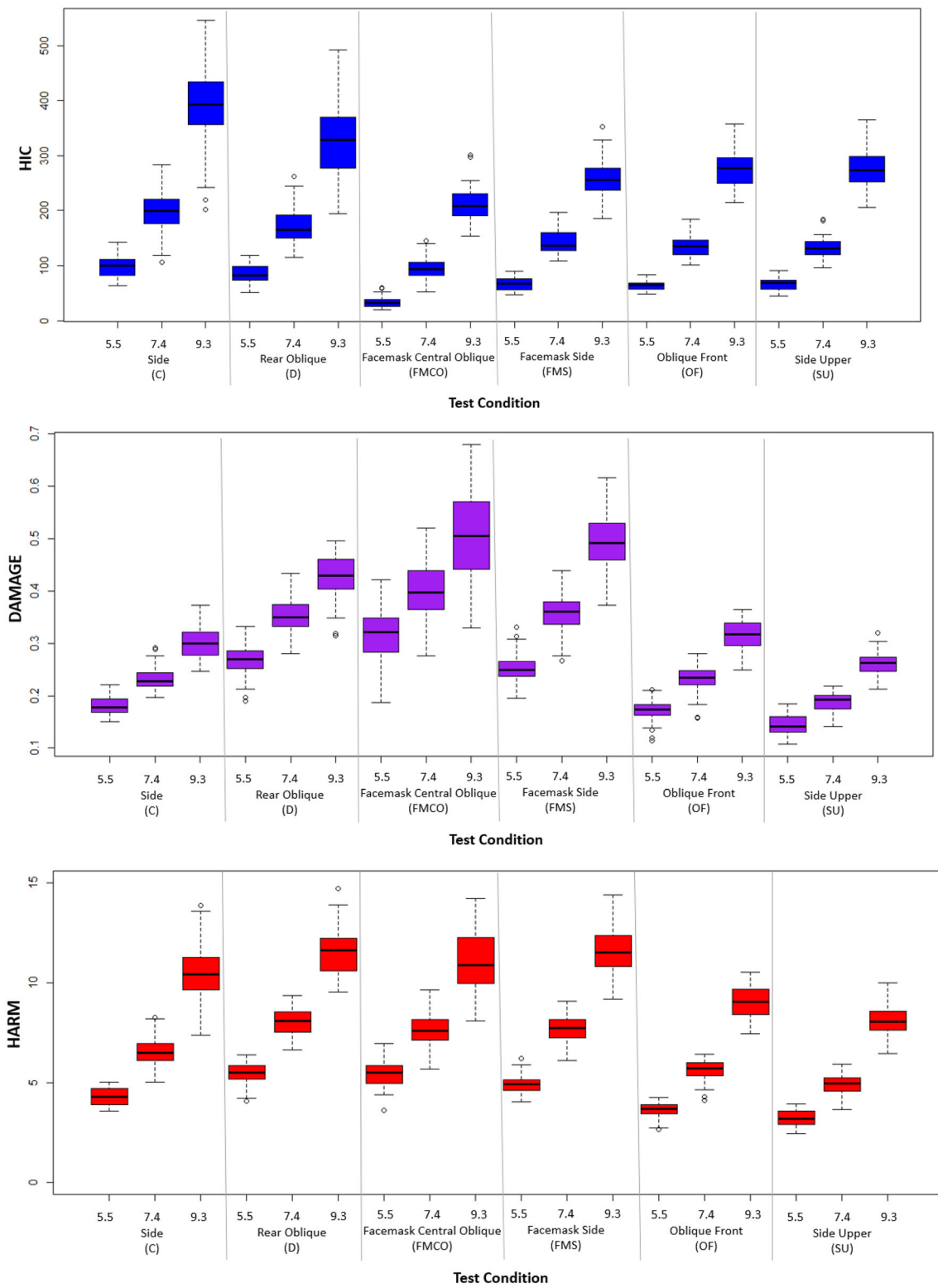


FIGURE 4. Box plots showing range of HIC15, DAMAGE, and HARM by test condition for the helmets tested in Lab 1.

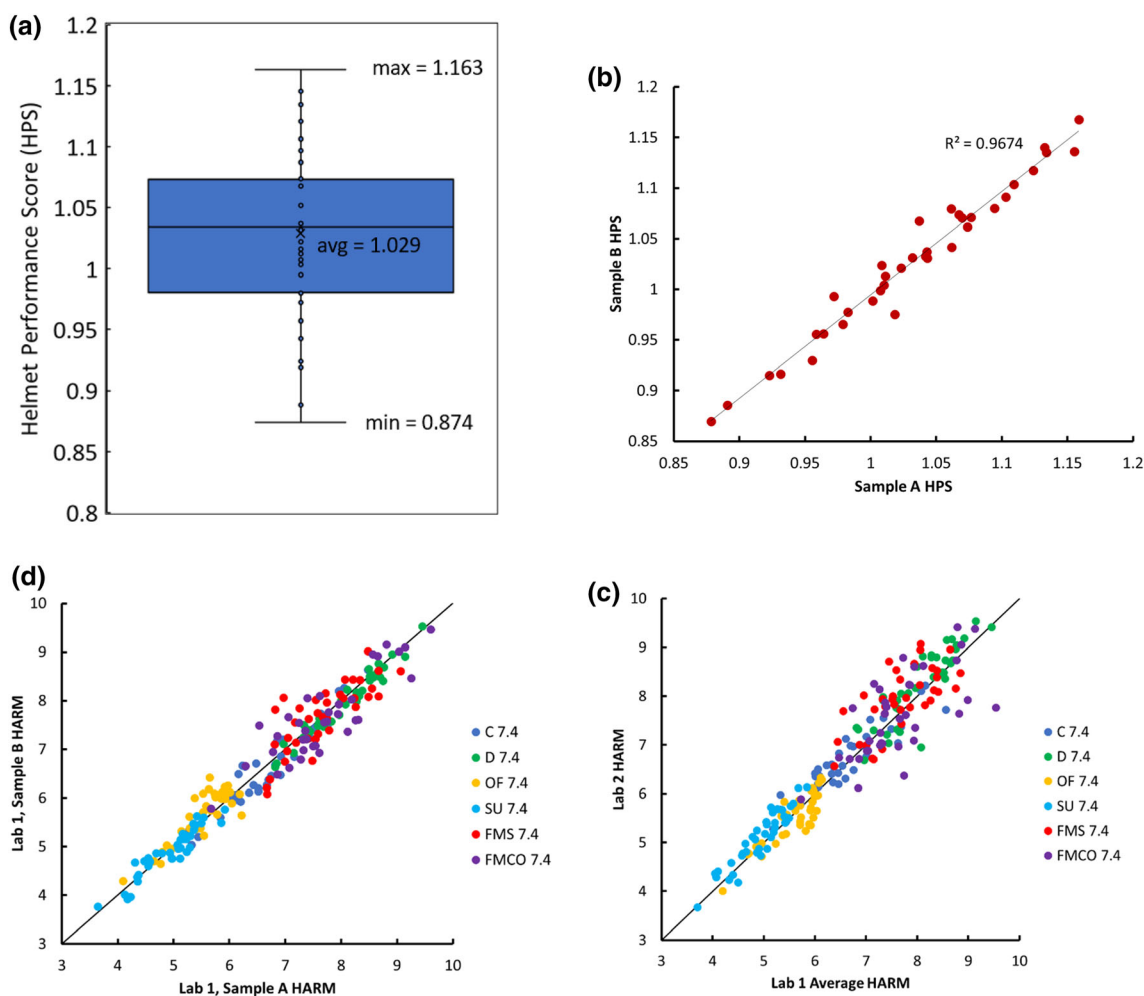


FIGURE 5. (a) Boxplot showing the average HPS for samples A and B for each helmet model. (b) Comparison of HPS for samples of the same helmet model. (c) Scatterplot comparing HARM for A vs. B helmet samples across all impact locations at 7.4 m s^{-1} . (d) Scatterplot comparing HARM for helmets of the same model tested at Laboratory 1 vs. Laboratory 2.

were representative of the larger population of concussive events. A limitation of the NGS data was that it did not capture head motion directly, but rather captured the torso velocity absent the vertical component. The average vertical component of closing velocity for helmet-to-helmet impacts included in the videogrammetry study was less than 1 m s^{-1} . Thus, it is likely that the closing velocities estimated from the NGS data were only minimally affected by this limitation. A new injury metric called the Head Acceleration Response Metric (HARM) was used to account for the translational and rotational kinematic responses of the head in each test. A new scoring system, the helmet performance score (HPS), was developed as a weighted sum of the HARM values for all tests, with each test weight derived from the proportion of concussive impacts in the NFL represented by that test condition. This weighting schematic would therefore incentivize better performance at the test condi-

tions most frequently associated with the incidence of concussion.

The test conditions and helmet performance metrics presented in this study target the locations and velocities most commonly associated with game concussions at the professional football level. The objective was to create a repeatable helmet ranking system that acknowledges helmets with the greatest ability to reduce severity of impacts at those conditions. Variation between samples of the same helmet model was less than a tenth of the range of helmet performance variation in HARM averaged across test conditions and all helmet models ($4.0\% \pm 3.6\%$ compared to $41 \pm 7.9\%$). Coefficient of variation for HARM for samples A and B for individual test conditions averaged $3.3 \pm 1.0\%$, whereas HMV of HARM for all helmets tested averaged $10 \pm 1.6\%$ for individual test conditions. Though inter-lab variation was larger than intra-lab HARM variation compared at individual

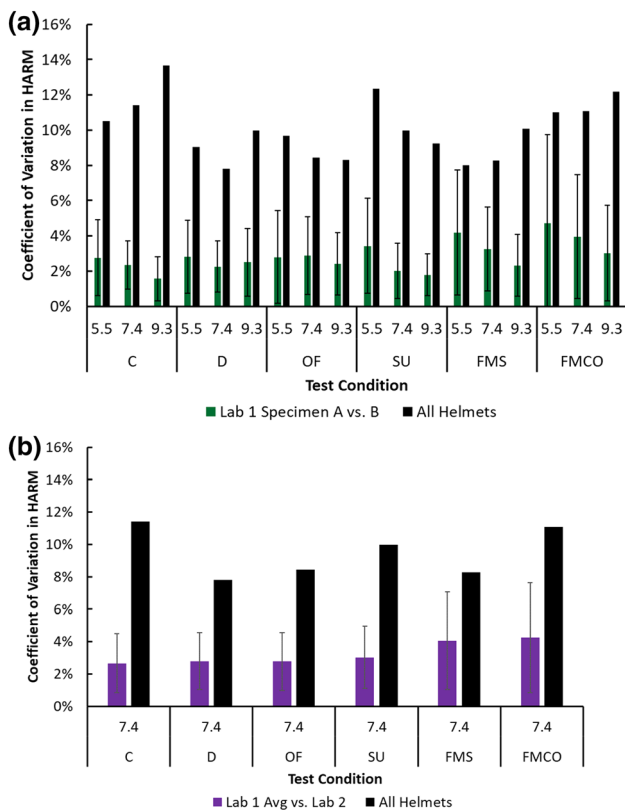


FIGURE 6. (a) Average (and standard deviation) of the coefficient of variation in HARM for intra-lab helmet tests by test condition compared to the coefficient of variation among all helmets tested at each test condition. (b) Average (and standard deviation) of the coefficient of variation in HARM for inter-lab helmet tests by test condition compared to the coefficient of variation among all helmets tested at each 7.4 m s^{-1} test condition. The standard deviation normalized by the average HARM for helmets tested is included as a reference for the inter-lab and intra-lab variation for each condition.

impact locations for the 7.4 m s^{-1} ($4.6 \pm 3.5\%$ compared to $3.9 \pm 3.2\%$), the regression model suggested there was no systematic bias in the relationship between test lab and HARM (Fig. 6). The larger inter-lab variation is likely attributable to a combination of small differences between the test fixture setup and human factors related to helmet positioning differences.

Variation in HPS for samples of the same helmet model was roughly 3.0% of the total range in HPS of helmets tested (33% difference between best and worst-performing helmet). It is important to note that while the overall range of helmet performance was large, the average difference in HPS for helmets ranked sequentially was only $0.8 \pm 0.7\%$. Given the small differences in performance that may exist between helmet models, interpretation of the data can benefit from the use of a statistical grouping analysis that accounts for the variation in performance of individual helmet models.

A limitation of this study is that no consideration was given to the effect of different faceguard styles or chin straps. An effort was made to identify a faceguard style common to all helmet models with similar patterns of vertical and horizontal bars in order to minimize the effect of variations in faceguard design. Previous studies have indicated the importance of including faceguards during the testing of football and other helmets for shell impacts, but to the authors' knowledge the effect of faceguard design on helmet performance has not been investigated and published.^{6,24} While some of the variation between samples A and B may be explained by manufacturing variation, much of the variation is likely attributable to differences in test-to-test variation including helmet positioning. While steps were taken to reduce variability in test fixture setup and positioning of the helmet, variation in reproducibility remains. In particular, larger variation was observed for impacts to the faceguard for which differences in helmet position can lead to the end cap interfacing with different bar structures on the mask and for which the chin strap tightness and interaction with the dummy chin may play a larger role.

Another limitation of this study relates to issues with the biofidelity of the Hybrid III for helmet testing. Because of these known issues associated with the Hybrid III neck for axial loading,^{4,18} impacts to the top of the helmet have not been included in this test protocol and, for other target locations, alternate impact vectors were chosen to reduce the axial component of neck loading. The Hybrid III headform, unlike the NOCSAE headform, lacks a realistic nape of the neck and chin.⁸ Despite this issue, the Hybrid III head was chosen for this work because of a more anthropometrically correct connection to the Hybrid III neck.⁸ Future work should focus on an improved design of a headform and flexible neck for helmet testing. The biofidelity of the kinematics produced by the test fixture is also affected by the limitation of the attachment to the base of the neck that allowed only one DOF of translation in the direction of the impact. Despite this limitation, head kinematics produced by the test fixture were found to be reasonable compared to those from reconstructions of concussive impacts with similar impact velocities (Figs. 7 and 8).

The test method developed in this study simulates helmet-to-helmet impacts, which account for 36% of NFL game concussions during the 2015-2016 and 2016-2017 seasons.¹⁷ Though impact locations relevant to other impact sources such as ground impacts have been partially addressed in the choice of helmet impact locations, future work should focus upon methods for simulating the mechanical response of helmet-to-body or helmet-to-ground impacts. Previous studies have indicated that while resulting helmet kinematics from

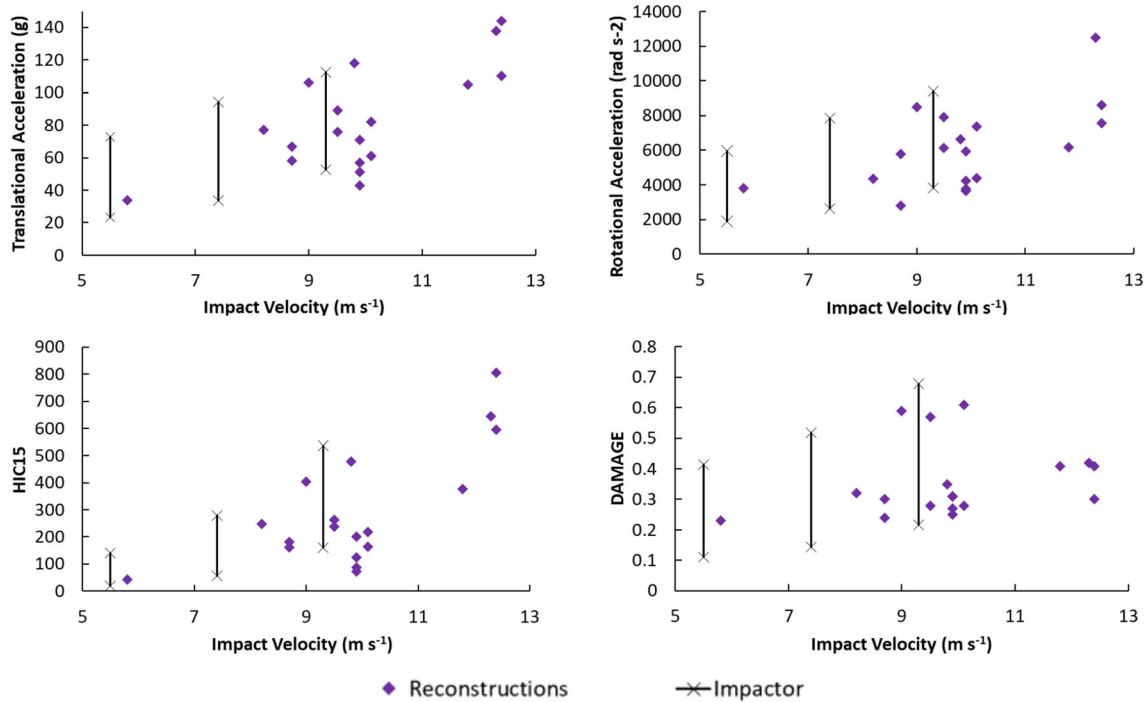


FIGURE 7. Peak head kinematics by impact velocity for laboratory reconstruction of concussive impacts and the range of responses for helmets tested using the linear impactor test conditions.

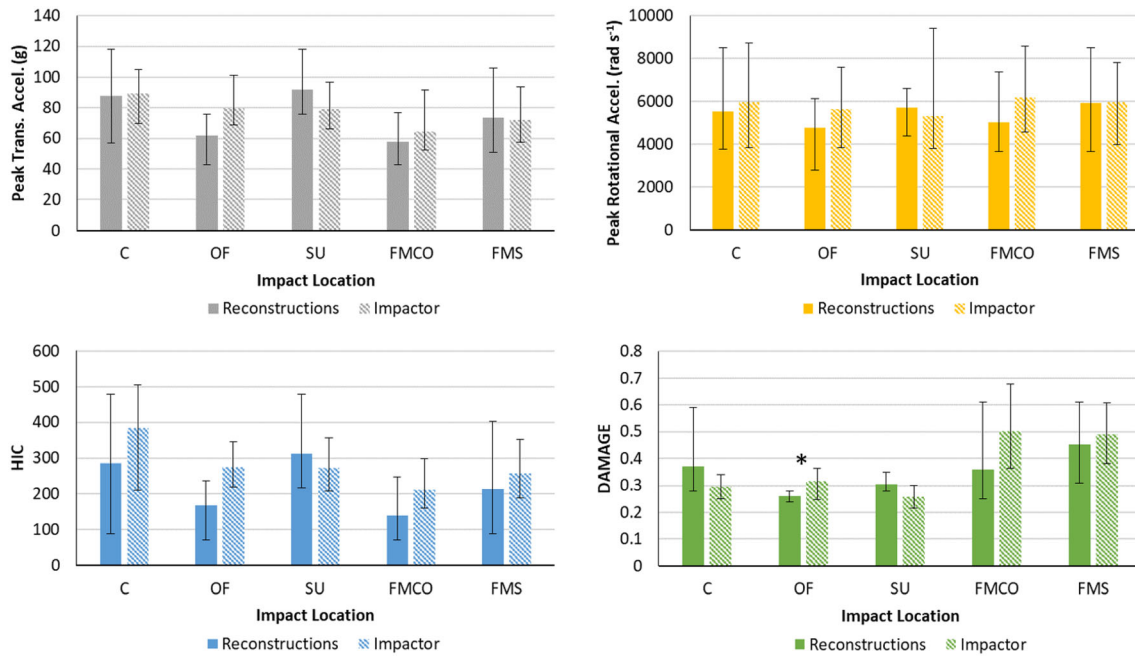


FIGURE 8. Peak head kinematics and metrics from laboratory reconstructions of concussive impacts¹³ and linear impactor tests at 9.3 m s⁻¹. Reconstruction results were grouped by impact location and limited to those with impact velocities ranging from 7.4 to 11.2 m s⁻¹. Each bar indicates the average for the group of impacts with error bars indicating the range of responses. Note that the D location was not included since there were no laboratory reconstructions of rear impacts. Asterisks denote statistically significant differences ($p < 0.05$) to a significance level of 0.05.

helmet-to-shoulder impacts were similar to those from helmet impacts,³ helmet-to-ground impacts often involve larger linear velocity changes and initial rotational velocities.¹⁶ These test methods represent impact conditions relevant for professional football and should not be translated to other levels of play or other sports. Further, these methods should be reassessed as helmet technology changes and more information about the injury mechanisms associated with concussion become available.

Building upon recent studies used to characterize the professional football impact environment and the use of new metrics for assessing head impact severity, this study presents a test methodology for assessing the performance of football helmets, with particular focus on severe and concussive impacts occurring the NFL. The ability of the test methodology to mimic the on-field environment and to generate repeatable and reproducible helmet responses was demonstrated. Lastly, testing of 36 helmet models with the new methodology provides an overview of the range of performance of modern football helmets with a 33% difference in HPS between the best and worst performing models. These results have been presented in the form of a poster to provide information to NFL players and team medical and equipment staff.¹⁹

ELECTRONIC SUPPLEMENTARY MATERIAL

The online version of this article (doi:<https://doi.org/10.1007/s10439-020-02626-6>) contains supplementary material, which is available to authorized users.

ACKNOWLEDGMENTS

The research presented in this paper was made possible by a grant from Football Research, Inc. (FRI). FRI is a nonprofit corporation that receives funding from sources including the NFL and is dedicated to the research and development of novel methods to prevent, mitigate, and treat traumatic head injury. The views expressed are solely those of the authors and do not represent those of FRI or any of its affiliates or funding sources. The authors would also like to acknowledge Biokinetics and Associates Ltd., Next Generation Stats, and Kineticorp.

REFERENCES

¹Alem, N. M. Helmet impact test system development. Highway Safety Research Institute. Final Report UM-HSRI-80-72-1, 1980.

- ²Bailey, A., J. Funk, D. Lessley, C. Sherwood, J. Crandall, W. Neale, and N. Rose. Validation of a videogrammetry technique for analysing American football helmet kinematics. *Sports Biomech.* 19(5):678–700, 2018.
- ³Bailey, A. M., C. Sherwood, J. R. Funk, J. R. Crandall, N. Carter, D. Hessel, S. Beier, and W. Neale. Characterization of concussive events in professional American football using videogrammetry. *Ann. Biomed. Eng.* 2020. <https://doi.org/10.1007/s10439-020-02637-3>.
- ⁴Bartsch, A., E. Benzel, V. Miele, D. Morr, and V. Prakash. Hybrid III anthropomorphic test device (ATD) response to head impacts and potential implications for athletic headgear testing. *Accid. Anal. Prev.* 48:285–291, 2012.
- ⁵Biomechanics Consulting and Research, LLC. Helmet test protocol. 2019. at www.biocorellc.com/resources.
- ⁶Breedlove, K. M., E. L. Breedlove, T. G. Bowman, E. M. Arruda, and E. A. Nauman. The effect of football helmet facemasks on impact behavior during linear drop tests. *J. Biomech.* 79:227–231, 2018.
- ⁷Campolettano, E. T., D. W. Sproule, M. T. Begonia, and S. Rowson. Youth football helmet STAR methodology. 2019.
- ⁸Cobb, 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. *Proc. Inst. Mech. Eng. Part P J. Sports Eng. Technol.* 229:39–46, 2015.
- ⁹Cobb, B., Tyson, A. M., Rowson, S. Head acceleration measurement techniques: reliability of angular rate sensor data in helmeted impact testing. *Proc IMechE Part P: J Sports Eng. and Tech.* 232(2): 176–181.
- ¹⁰DiMasi, F. P. Transformation of nine-accelerometer-package (NAP) data for replicating headpart kinematics and dynamic loading. United States. National Highway Traffic Safety Administration, 1995. DiMasi, F. P. Transformation of nine-accelerometer-package (NAP) data for replicating headpart kinematics and dynamic loading. United States. National Highway Traffic Safety Administration, 1995.
- ¹¹Elkin, B. S., L. F. Gabler, M. B. Panzer, and G. P. Siegmund. Brain tissue strains vary with head impact location: a possible explanation for increased concussion risk in struck versus striking football players. *Clin. Biomech.* 64:49–57, 2019.
- ¹²Ellenbogen, R. G., H. Batjer, J. Cardenas, M. Berger, J. Bailes, E. Pieroth, and R. Heyer. National football league head, neck and spine committee's concussion diagnosis and management protocol: 2017-18 season. *Br. J. Sports Med.* 52:894–902, 2018.
- ¹³Funk, J. R., R. Jadischke, A. Bailey, J. R. Crandall, J. McCarthy, K. B. Arbogast, and B. S. Myers. Laboratory reconstructions of concussive helmet-to-helmet impacts in the National Football League. *Ann. Biomed. Eng.*, in review.
- ¹⁴Gabler, L. F., A. M. Bailey, J. R. Funk, J. R. Crandall, K. B. Arbogast, and B. S. Myers. Development of a metric for ranking the performance off football helmets. *Proceedings of the 2019 Biomedical Engineering Society Conference*, Philadelphia, PA, 2019.
- ¹⁵Gabler, L. F., J. R. Crandall, and M. B. Panzer. Development of a metric for predicting brain strain responses using head kinematics. *Ann. Biomed. Eng.* 46:972–985, 2018.

- ¹⁶Kent, R., J. L. Forman, A. B. Good, C. Sherwood, and J. R. Crandall. The biomechanics of concussive helmet-to-ground impacts in the National Football League. *J. Biomech.* 99, 2020.
- ¹⁷Lessley, D., R. Kent, C. Sherwood, J. Cormier, J. Crandall, K. B. Arbogast, and B. S. Myers. Video analysis of reported concussion events in the National Football League During the 2015-2016 and 2016-2017 Seasons. *Am. J. Sports Med.* 46(14):3502–3510, 2018.
- ¹⁸Mertz, H. J. Biofidelity of the Hybrid III head. *SAE Transactions.* SAE 851245:97–105, 1985.
- ¹⁹National Football League, and National Football League Players Association. NFL NFLPA 2020 helmet laboratory testing performance results, 2020. at <https://www.playsmartplaysafe.com/resource/helmet-laboratory-testing-performance-results/>.
- ²⁰National Operating Committee on Standards for Athletic Equipment. Standard linear impactor test method and equipment used in evaluating the performance characteristics of protective headgear and face guards (ND 081-18am19a), 2019.
- ²¹Pellman, 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.
- ²²Pellman, E. J., D. C. Viano, C. Withnall, N. Shewchenko, C. A. Bir, and P. D. Halstead. Concussion in professional football: helmet testing to assess impact performance—part 11. *Neurosurgery* 58:78–95, 2006.
- ²³Post, A., A. Oeur, B. Hoshizaki, and M. D. Gilchrist. An examination of American football helmets using brain deformation metrics associated with concussion. *Mater. Des.* 45:653–662, 2013.
- ²⁴Rowson, B., E. J. Terrell, and S. Rowson. Quantifying the effect of the facemask on helmet performance. *Proc. Inst. Mech. Eng. Part P J. Sports Eng. Technol.* 232:94–101, 2018.
- ²⁵Rowson, S., and S. M. Duma. Development of the STAR evaluation system for football helmets: integrating player head impact exposure and risk of concussion. *Ann. Biomed. Eng.* 39:2130–2140, 2011.
- ²⁶Rowson, S., and S. M. Duma. Brain injury prediction: assessing the combined probability of concussion using linear and rotational head acceleration. *Ann. Biomed. Eng.* 41:873–882, 2013.
- ²⁷Rowson, S., S. M. Duma, J. G. Beckwith, J. J. Chu, R. M. Greenwald, J. J. Crisco, P. G. Brolinson, A.-C. Duhaime, T. W. McAllister, and A. C. Maerlender. Rotational head kinematics in football impacts: an injury risk function for concussion. *Ann. Biomed. Eng.* 40:1–13, 2012.
- ²⁸Sanchez, E. J., L. F. Gabler, A. B. Good, J. R. Funk, J. R. Crandall, and M. B. Panzer. A reanalysis of football impact reconstructions for head kinematics and finite element modeling. *Clin. Biomech.* 64:82–89, 2018.
- ²⁹Society of Automotive Engineers. J211-1 instrumentation for impact test—Part 1—electronic instrumentation. *SAE Int.* 211/1201403, 2014.
- ³⁰Takhounts, E., R. Eppinger, R. Tannous, J. Q. Campbell, E. Power, L. Shook, and V. Hasija. Analysis of 3D rigid body motion using the nine accelerometer array system. *Injury Biomechanics research Proc. of the Thirty-First Int. Workshop*, 2003.
- ³¹Takhounts, 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, 2013.
- ³²Tyson, A. M., and S. Rowson. Varsity football helmet STAR protocol, 2018.
- ³³Versace, J. A review of the severity index. *Proc. 15th Stapp Car Crash Conference*, 771–796, 1971.
- ³⁴Viano, D. C., and D. Halstead. Change in size and impact performance of football helmets from the 1970s to 2010. *Ann. Biomed. Eng.* 40:175–184, 2012.
- ³⁵Viano, D. C., C. Withnall, and D. Halstead. Impact performance of modern football helmets. *Ann. Biomed. Eng.* 40:160–174, 2012.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.