The Effect of Usage on Instrumented Mouthguard Data Quality Indicators: A Preliminary Analysis.

Enora Le Flao, Samantha DeAngelo, Adam Culiver, Nick Shoaf, Durshil Doshi, Ryan Tracy, Jenna Lickovitch, Carly Smith, Jeff Pan, Jingzhen Yang, Sean Rose, James Onate, Zeynep Saygin, Jaclyn Caccese

Abstract Instrumented mouthguards are frequently used to study head impacts in sports. However, we hypothesise that repeated usage of the mouthguard may compromise the fit on the teeth and result in lower data quality. Sensor acceleration events were recorded from youth American football players. Events verifiable on video were included and subjected to the mouthguard manufacturer's classification algorithm. For each event, the device provides a quality class (high, moderate, or low), which we examined for the various data cleaning methods, and in relation to the duration of mouthguard usage. Out of all events (n=9,375), the proportions of high-, moderate-, or low-quality events were 8%, 28%, and 64%, respectively. Among those, 184 events were determined as true positives by video review and algorithm, with high-, moderate- and low-quality proportions of 81%, 14%, and 5%. There was no correlation between mouthguard usage duration and signal quality for video-verified events, however, there was a positive correlation when including all events (r=0.170, p=0.031), which is contrary to our hypothesis. Whether the kinematics from low-quality events are valid is unknown, and whether these should be included in analyses is questionable. Ongoing analysis of data from practices may provide further clarification on signal quality and mouthguard wear-and-tear.

Keywords Concussion, Head impacts, Head kinematics, Sports, Traumatic brain injury.

I. INTRODUCTION

Instrumented mouthguards (iMGs) designed to measure head impacts during sports participation have been used increasingly since 2014 [1]. The devices have allowed researchers to estimate the number and magnitude of head acceleration events that may present clinical consequences on brain health in various sports and with athletes of both sexes and various ages [2]. Beyond academia, many iMGs are available to the wider public, such that parents can monitor their children's head impacts, or a coach can assess their team's impact load. In 2023, iMGs also became mandated by World Rugby at the highest levels of the professional game to improve their head injury identification process [3]. However, one issue when working with iMGs relates to damage caused by inappropriate device care (e.g., excessive chewing, tearing of the mouthguard) resulting in loss of connectivity, trouble charging the iMGs, and need for replacement, ultimately leading to head impacts not being recorded [4-5].

It is also reasonable to expect, from field observations and published images [4], that damage to the iMG caused by excessive chewing can seriously affect the fit on the teeth and make the iMG loose [6]. The lack of proper coupling between a head impact sensor and the skull is a well-known challenge, as it leads to low-quality signals incorporating large noise and sharp spikes [7-9], resulting in inaccurate kinematics and larger numbers of events being recorded [10-12]. In consequence, the estimation of both the number and the magnitude of head impacts measured on the field may be inaccurate, ultimately limiting our understanding of the effects of head impacts on brain health. Therefore, it is important to determine whether and how damage and normal wear-and-tear throughout a study period affect the quality and accuracy of the data being collected.

Although assessing the association between iMGs usage and data accuracy is not possible at this point due to the absence of a reference measurement provided by a rigidly coupled device, analysing the quality of kinematic signals can provide valuable insights on data quality. Primarily, if a sensor acceleration event (SAE), i.e. any

E. Le Flao (enora.leflao@osumc.edu) and A. Culiver are postdoctoral scholars; S. DeAngelo, C. Smith are postgraduate students; N. Shoaf, D. Durshil, R. Tracy, J. Lickovitch are undergraduate students; J. Onate is an Associate Professor.; J. Caccese is an Assistant Professor - all in Health & Rehabilitation Sciences at The Ohio State University, USA. J. Pan is an Associate Professor in Biomedical Informatics; Z. Saygin is an Assistant Professor in Psychology - all at The Ohio State University, USA. J. Yang is a Professor of Paediatrics; S. Rose is an Associate Professor of Pediatrics, both at the Ohio State University and Nationwide Children's Hospital.

recording by the iMG, is not associated with a true impact to the iMG user, as determined via video review, then the spurious measurement is likely the result of noise or unwanted sensor motion. Secondly, the presence of noise or spikes in the time series data has been visually assessed as an indicator of poor coupling to the skull [7]. The recent consensus on head acceleration measurement practices (CHAMP) suggests reviewing waveforms to “identify spurious impacts—those with time-clipped impulses, ringing, multiple nonsensical impulses, and poor signal-to-noise ratios” [9]. Signal-to-noise ratio and other frequency metrics have also been assessed quantitatively [1][6]. Finally, some iMG manufacturers have incorporated a quality assessment into their post-processing pipelines [13].

Given the frequent damage to iMGs from chewing and the known issues of poor mouthguard-to-skull coupling, there is a need to assess whether wear-and-tear affects the quality of kinematic signals from head impacts. Therefore, the objectives of this study were to (1) examine the quality of kinematic signals measured during sports participation, using metrics provided by the iMG manufacturer, and (2) investigate the association between usage and data quality throughout the study period. We hypothesise that repeated iMG usage leads to lower-quality data, which may have consequences on the estimation of the exposure to head impacts.

II. METHODS

Youth American tackle football players were recruited for this study approved by The Ohio State University Institutional Review Board. Head impacts were measured using instrumented mouthguards during the 2023 football season. The present analyses pertain to a preliminary subset of data recorded for 40 athletes over 13 games, or 162 athlete exposures.

Instrumentation

Athletes (8-12 years, mean age 10.3 ±1.2 years) were equipped with iMGs (V2, Prevent Biometrics [PB], Minneapolis, MN). All but five iMGs were custom-made from upper dentition scans obtained by a trained dentist or a trained member of the research team; the remaining five were of the boil-and-bite type to accommodate braces. Each iMG comprises high-g and low-g triaxial accelerometers and an angular rate sensor, all sampling at 3200 Hz, with ranges of ±200 g (high-g) and ±35 rad.s^{-1} on each axis, respectively. Sensor acceleration events (SAEs) were triggered when any axis of linear acceleration exceeded 8 g and captured 10 ms pre- and 40 ms post-trigger. Prevent Biometrics’ proprietary algorithm derives angular acceleration from angular velocity and transforms linear acceleration to the estimated head’s centre of gravity. Data are filtered using a 4th order, zero-lag low pass Butterworth filter, for which the cutoff frequency varies based on the quality of the signal. According to the manufacturer, a machine learning algorithm quantifies the signal-to-noise ratio for each event, attributes it a quality class of 0-high, 1-moderate, or 2-low quality, and filters respectively with 200, 100, or 50 Hz cut-off frequencies (Fig. 1). For moderate and low-quality events, the PB algorithm multiplied the resultant peak velocities and accelerations by a correction factor (~1.3 and ~2.0, respectively) established from laboratory tests and proposed to better approximate the true peaks for noisy events. Events with a resultant peak linear acceleration (PLA) comprised between 10 and 200 g were downloaded from the PB portal in November 2023 (‘SoftwareVersion: 2.1.24’, ‘DTAVersion: 2.1.24 scaled’).

Event Verification and Inclusion

To distinguish between true head impacts and false positives, SAEs were verified using two methods independently: (1) video review and (2) the manufacturer's proprietary algorithm. Method (1) for video review, games were filmed using one camera (Sony Cybershot RX100 V, 4K resolution, 24 fps) zoomed-in on the play and manually operated to follow the main group of players on the field. After starting the video recording, the operator showed the time from their phone to the camera. Video reviewers blinded to the sensor data used this time to match SAE timestamps to visible events and classify SAEs as true positive, false positive, or unverifiable (e.g., when the athlete could not be seen on video). Each event was verified by one reviewer. Events occurring before the start of the game, after the end, or during breaks (as noted by the camera operator) were automatically classified as false positives. Method (2) for the manufacturer classification, PB mouthguards incorporate a proximity sensor to determine whether the iMG is on the teeth at the moment of the event and classify SAEs as true or false positives. This algorithm is independent of the quality determination model described above, and algorithm specifics were unavailable to the research team. Events were included for analysis only if they were video-verified as true or false positives, and if the PB quality class was available.

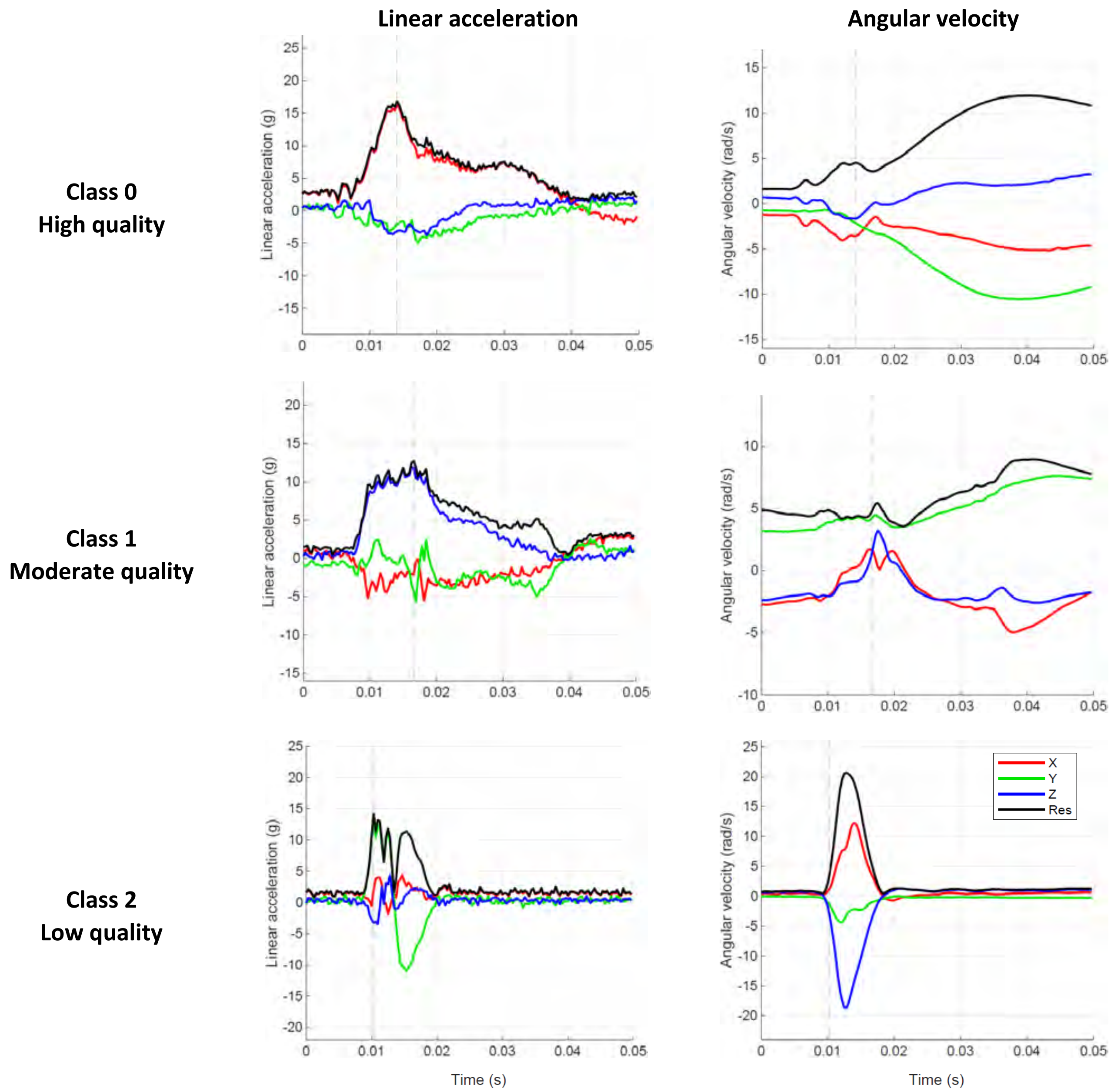

Fig. 1. Raw linear acceleration and angular velocity time series for examples of events classified as high (top), moderate (middle), or low-quality (bottom) by the Prevent Biometrics model. The dashed vertical lines highlight the time of the peak linear acceleration for each event.

Analyses

Throughout the study, of the 40 athletes, six needed their iMG replaced once and three needed it twice, primarily because of connectivity issues often due to damage from chewing. Therefore, a total of 52 iMGs were included in the analyses. The number of events determined as true or false positives were reported for both cleaning methods, as well as the performance of the PB algorithm, using video review as reference. The proportions of high-, moderate- and low-quality events, as determined by the PB model, were examined for all stages of data cleaning, for individual athlete exposures, and for each mouthguard. For each event, the number of days passed since the very first iMG use was calculated using the larger study dataset. The association between the number of days since first use and the proportion of high-quality events for each athlete exposure was explored with a linear regression. The hypothesis was that the longer the iMG had been in use, the fewer high-quality events would be recorded by the iMG.

III. RESULTS

Over the duration of the study, mouthguards were used for up to 95 subsequent days. The most uses for one mouthguard was 54 exposures, including 7 games and 47 practices. Some mouthguards showed evident mechanical wear, from slight bite marks to severe chewing, causing the non-sensor side to be flattened out.

A total of 9,375 sensor acceleration events (SAEs) measured from 162 athlete exposures were included for analysis (40 athletes for 52 mouthguards in total, over 13 games) (Fig. 2). Mouthguards recorded a median of 22 SAEs per athlete exposure (interquartile range [IQR]: 10–68, range: 1–404). Out of those events, medians of four video-verified and two PB-algorithm-classified events were available per athlete exposure (IQR: 1–8, range: 1–33, and IQR: 1–6, range: 1–39, respectively). For both the video review and the PB algorithm, true positives amounted to 5% of all included events. When considering video review as the reference, the performance of the PB algorithm to correctly identify true contact events resulted in a sensitivity of 37%, a specificity of 97%, a positive predictive value of 41%, and a negative predictive value of 96%.

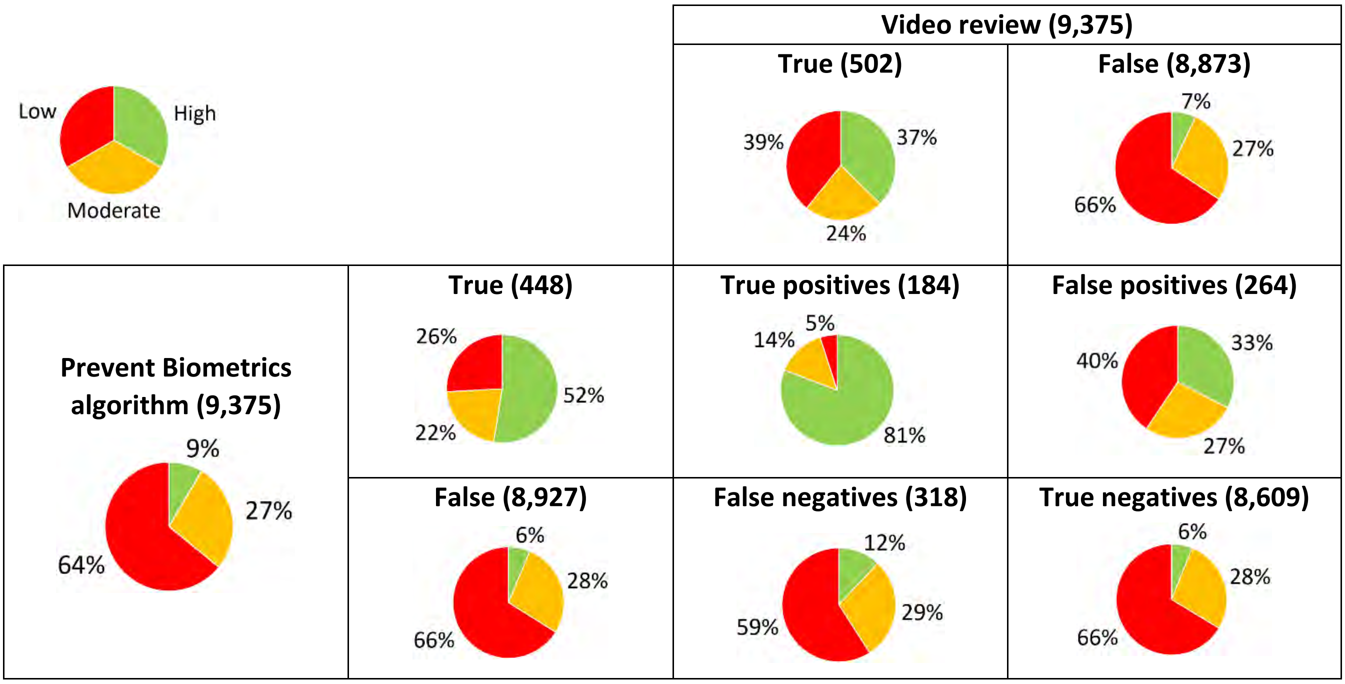

Fig. 2. Contingency table for the number of events determined as True or False positives by the Prevent Biometrics' algorithm and the video review. The pie charts represent the proportions of high (green), moderate (yellow), and low-quality (red) events.

Quality indicator. Over the 9,375 events included, 9% were classified by the PB algorithm as high-quality, 27% as moderate-quality, and 64% as low-quality (Fig. 2). The proportions of high-, moderate-, and low-quality events varied substantially across data cleaning methods, and the combination of video verification and PB algorithm provided the most favourable quality distribution with 81% of high-quality events.

For individual athlete exposures, the proportion of high-quality events ranged from 0 to 73% (median [IQR]: 6.4% [0.0–16.7]) for all SAEs recorded (Fig. 3), and from 0 to 100% (25.0% [0.0–62.5]) for video-verified events (Fig. 2). Six mouthguards showed only high-quality events, while 12 showed no high-quality events (26% of all athlete exposures resulted in no high-quality events being recorded). There was variability in the proportion of high-quality events across mouthguards and within athlete exposures.

Association between repeated usage and data quality. Over 162 athlete exposures and for all recorded SAEs, there was a significant but weak linear correlation showing that the longer an iMG was in use, the larger the proportion of high-quality events ($r = 0.170$, $p = 0.031$, Fig. 4). However, there was no correlation for SAEs determined as true impacts through video review ($r = 0.120$, $p = 0.246$) or by the PB algorithm ($r = 0.152$, p = 0.154). There were insufficient true positive events determined by both methods for analysis.

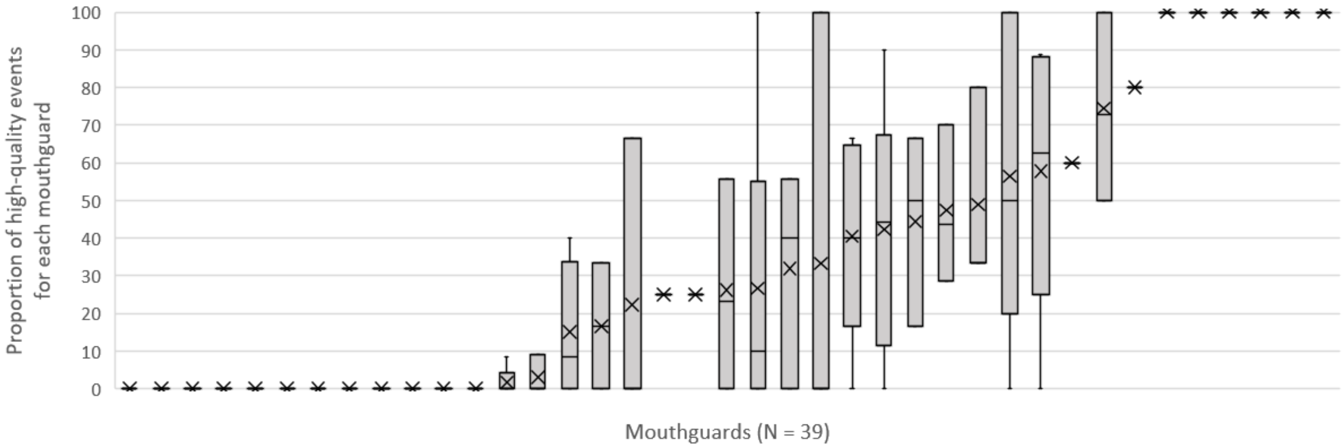

Fig. 3. Distribution of the proportion of high-quality video-verified events for individual mouthguards over the available athlete exposures (39 out of 52 mouthguards recorded video-verified events). Mouthguards on the x-axis are sorted by the median value in ascending order. Each mouthguard recorded for one to seven athlete exposures, and each exposure had 1 to 33 video-verified events, independent of the Prevent Biometrics algorithm classification. For each mouthguard, the dash represents the median and the X represents the mean.

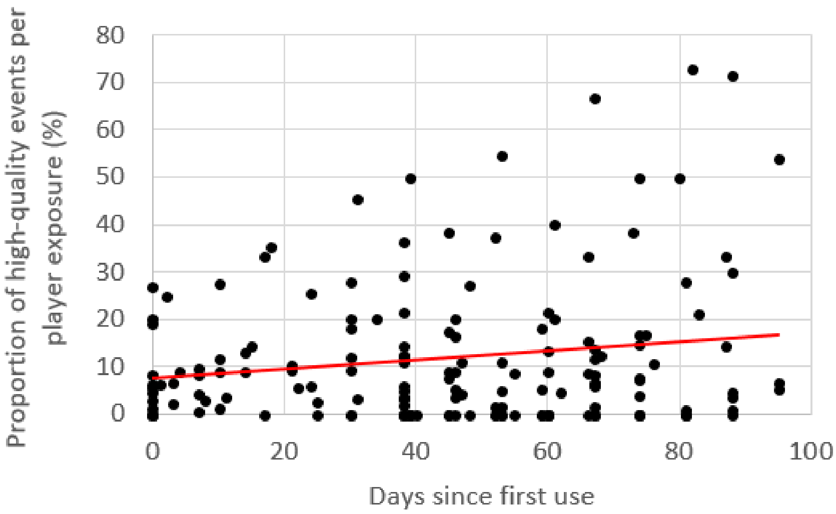

Fig. 4. Proportion of high-quality sensor acceleration events for each athlete exposure relative to the number of days between the exposure and the first use of the mouthguard (N = 162 athlete exposures, including 9,375 events).

IV. DISCUSSION

This study analysed a manufacturer-provided indicator of signal quality for instrumented mouthguard events resulting from youth American tackle football. Results showed that the proportion of high-quality events was low in the full dataset (9%), but substantially increased for events that were confirmed by video review (37%) or the manufacturer's algorithm (52%), or both (81%). The proportion of high-quality events also varied between and within participants, and overall seemed to maintain or even increase throughout the study period, which goes against our hypothesis that event quality would decrease with iMG usage over time.

Data cleaning with video review or the manufacturer's algorithm led to more favourable distributions of high-, moderate-, and low-quality events. Of the events that were triggered when no verified contact happened, and events where the proximity sensor determined the mouthguard was not on the teeth, the majority were of low quality (66% for both methods), confirming a link between spurious events and noise in the kinematic signals. Our findings also suggest that if an SAE verified by the PB algorithm is of low quality, it is likely not associated with a contact event and could be discarded. The combination of both cleaning methods resulted in large proportions of high- and moderate-quality events, and most video-verified low-quality signals were discarded. However, whether to include or exclude video-verified low-quality events is unclear. We suggest that while the kinematics are likely unreliable, such events should be included in the analysis of impact counts after thorough video verification. The improvement of quality-classification algorithms and methods to process noisy events is necessary and on-going [14-15], although improving the fit of the iMG should be the priority.

There was no correlation between the number of days since the first iMG use and the proportion of high-quality events for video-verified or algorithm-confirmed SAEs. Furthermore, there was a marginal increase in quality for all recorded SAEs over iMG usage time, which is contrary to our hypothesis. Our primary explanation is that of survivorship bias, where the iMGs that fit best stayed in use for longer, while poorly fitted iMGs (mostly from mechanical damage) were replaced more quickly. Additionally, from communication with the manufacturer, we understand that the algorithm that uses the proximity sensor information may “learn” over time by refining the individual thresholds used to determine whether the iMG is on the teeth or not, becoming more accurate as data are recorded and processed. It might therefore affect the number of events that are recorded and explain some of the inter- and intra-individual variations. Future analyses of the detection thresholds on our complete dataset may shed some light on this hypothesis and verify that this association is not mainly driven by noise. With our complete dataset including more athletes and exposures, we plan to conduct athlete-wise analyses to account for the inter-individual variations, and further investigate the potential association between iMG usage and quality.

As a preliminary analysis, some of this work’s limitations pertain to the limited availability of data while video review for additional players and all practice sessions is in progress. Including SAEs from practices will allow for a more complete record of iMG usage, as athletes trained at least three times more often than they played in games. Therefore, rather than counting the number of days since first use as the main usage metric, we will count the number of sessions the iMG was worn by a player. While this more detailed usage metric may explain some of the intra-individual variability, we also expect that visible damage to the mouthguard influences the quality of the data (some players chewed on their mouthguards so badly that no teeth indentation could be seen after just a few practices). The subjective analysis of the visual wear-and-tear of the mouthguards is underway. We also hypothesise that custom-fitted mouthguards would record less low-quality events by providing a better fit than boil-and-bite iMGs [16]. We have insufficient data for boil-and-bite iMGs in this preliminary dataset, but initial analyses indicate that there were more high-quality events for boil-and-bite iMGs than for custom iMGs, which would go against our hypothesis. Finally, we reported on one of many metrics that could be used for assessing signal quality. While our visual appraisal of the raw time traces matched well with the PB quality class, these classes are determined through a “black box” algorithm: we do not know what features were used or how it was developed, nor do we know when the algorithm has been or will be updated. Other metrics or replicable methods to assess signal quality may prove informative.

V. CONCLUSION

This preliminary study explored the relatively new quality indicator provided by an instrumented mouthguard manufacturer. We analysed the distribution of high-, moderate-, and low-quality events in our youth American tackle football preliminary dataset of 13 games. While the number of low-quality events in the raw dataset was large, the application of common data cleaning methods – video review to eliminate events not associated with a collision and proprietary “black box” algorithm to discard events recorded while the mouthguard was off the teeth – eliminated many such events. Our findings also highlighted that many events verified by video were of low quality, which raises the question of including or excluding those from analysis. There was no significant association between mouthguard usage duration and signal quality, which goes against our hypothesis that repeated usage leads to lower-quality data. This preliminary dataset was limited to games only, but future analyses will include practice sessions and a more accurate exposure quantification. Overall, the results of this work will allow for a better understanding of instrumented mouthguard data, which will help refine rigorous practices to improve data quality, and thereby generate more accurate datasets for the study of brain health.

VI. REFERENCES

[1] Bartsch, A., Samorezov, S., Benzel, E., Miele, V., and Brett, D. Validation of an "Intelligent Mouthguard" single event head impact dosimeter. *Stapp Car Crash J.*, 2014. 58: p. 1-27
[2] Le Flao, E., Siegmund, G.P., and Borotkanics, R. Head impact research using inertial sensors in sport: A systematic review of methods, demographics, and factors contributing to exposure. *Sports Med.*, 2021(52): p. 481–504
[3] Rugby, W. World Rugby integrates smart mouthguard technology to the Head Injury Assessment as part of new phase of global player welfare measures. 2023, World Rugby: www.world.rugby.
[4] Nauman, E., Talavage, T.M., and Auerbach, P.S. Mitigating the consequences of subconcussive head injuries. *Annu Rev Biomed Eng*, 2020. 22(1): p. 387-407
[5] Gabler, L.F., Huddleston, S.H., et al. On-field performance of an instrumented mouthguard for detecting head impacts in American football. *Ann Biomed Eng*, 2020. 48(11): p. 2599-2612
[6] Luke, D., Kenny, R., Bondi, D., Clansey, A.C., and Wu, L.C. On-field instrumented mouthguard coupling. *J Biomech*, 2024. 162: p. 111889
[7] Rooks, T.F., Dargie, A.S., and Chancey, V.C. Machine learning classification of head impact sensor data, in *ASME International Mechanical Engineering Congress and Exposition, Proceedings*. 2019.
[8] Tiernan, S., Meagher, A., et al. Concussion and the severity of head impacts in mixed martial arts, in *Proceedings of the Institution of Mechanical Engineers*. 2020.
[9] Rowson, S., Mihalik, J., et al. Consensus Head Acceleration Measurement Practices (CHAMP): Study design and statistical analysis. *Ann Biomed Eng*, 2022
[10] Wu, L.C., Nangia, V., et al. In vivo evaluation of wearable head impact sensors. *Ann Biomed Eng*, 2016. 44(4): p. 1234-45
[11] Kuo, C., Wu, L.C., et al. Effect of the mandible on mouthguard measurements of head kinematics. *J Biomech*, 2016. 49(9): p. 1845-53
[12] Le Flao, E., Lenetsky, S., Siegmund, G.P., and Borotkanics, R. Capturing head impacts in boxing: A video-based comparison of three wearable sensors. *Ann Biomed Eng*, 2023
[13] Tooby, J., Weaving, D., Al-Dawoud, M., and Tierney, G. Quantification of head acceleration events in rugby league: An instrumented mouthguard and video analysis pilot study. *Sensors*, 2022. 22(2): p. 584
[14] Zhan, X., Liu, Y., et al. Denoising instrumented mouthguard measurements of head impact kinematics with a convolutional neural network. *arXiv preprint arXiv:2212.09832*, 2022
[15] Tierney, G., Rowson, S., et al. Frequency content and filtering of head sensor kinematics: A method to enable field-based inter-study comparisons. *arXiv preprint arXiv:2303.03043*, 2023
[16] Liu, Y., Domel, A.G., et al. Validation and comparison of instrumented mouthguards for measuring head kinematics and assessing brain deformation in football impacts. *Ann Biomed Eng*, 2020. 48(11): p. 2580-2598