A Field-Crash-Validated Female Lower Extremity Injury Criterion

Matthew L. Brumbelow

Abstract Females are at elevated risk of lower extremity injury in frontal crashes. This study builds on previous research to develop a new female lower extremity injury criterion (FLEIC) calculated from maximum values recorded with the Hybrid III 50th percentile male dummy. For vehicles with good overall driver and structure ratings in the IIHS moderate overlap test, a 5-unit increase in FLEIC (approximately 1 standard deviation) was associated with increases of 50–60% in the odds of any female driver sustaining an AIS≥2 lower extremity injury (p=0.01) or an AIS≥2 right foot/ankle injury (p=0.01). FLEIC was a better predictor of female injury than toe pan or brake pedal intrusion measured in the crash test. FLEIC was not associated with a significant change in male driver injury risk. Analysis of data and footwell video from 100 recent crash tests showed that FLEIC increased with intrusion magnitude, but the highest values were often observed early in the intrusion phase rather than at maximum intrusion. Another common contributor to FLEIC was loading from a deploying knee airbag. FLEIC exhibits better repeatability than its component measures. The inclusion of FLEIC in existing rating programmes would help encourage vehicle designs with reduced lower extremity injury risk for female drivers.

Keywords Crashworthiness evaluations, females, frontal crashes, Hybrid III, lower extremities.

I. INTRODUCTION

Lower extremity injury risk in frontal crashes is greater for female drivers than for males [1–3]. While the injury risk for both sexes is lower in newer vehicles with improved crashworthiness ratings, the difference between female and male injury risk in these vehicles remains significant [4]. Additional vehicle design changes are necessary to specifically address the additional risk faced by female drivers.

Research is ongoing to evaluate the mechanisms responsible for the elevated lower extremity injury risk for females, but already there is evidence that existing crashworthiness evaluations could be adjusted to encourage improvements for females. Results from a prior study demonstrated lower extremity injury values measured with the Hybrid III 50th percentile male (HIII-50M) dummy in the IIHS moderate overlap test are significant predictors of real-world injury risk for female drivers [1]. This was the case even though the analysis was restricted to vehicles that had good ratings for structural intrusion and good overall ratings (test version 1.0, prior to addition of a rear-seated dummy in 2022). Principal component analysis (PCA) using maxima from all 18 recorded HIII-50M lower extremity measures found that the most variance among tested vehicles was captured by a linear combination in which all measures contributed additively. While controlling for other crash and driver factors, female lower extremity injury risk was greater in vehicles with higher values of this first principal component (PC1).

The relationship between the PC1 of HIII-50M measures in good-rated vehicles and real-world injury risk for female drivers suggests that the existing IIHS rating system could be improved. Currently, driver lower extremity injury ratings are assigned after determining sub-ratings using separate boundaries for femur and tibia axial forces, knee displacement, foot acceleration, and upper and lower tibia indices calculated from a combination of forces and moments [5]. Each of these measures and their representative boundaries were selected based on biomechanical research as described in [5]. The worst sub-rating for each of these measures is selected as the overall rating for the lower extremities. The results using PC1 suggest that a single score, representing a combination of all measurements, may provide a meaningful distinction among vehicles that currently have the same overall rating. In other words, while the current rating procedure involves categorising individual measures (into sub-ratings) prior to combining them in an overall rating, this could be improved by combining continuous measures prior to categorisation.

M. L. Brumbelow (e-mail: mbrumbelow@iihs.org; tel: +1-434-985-4600) is Principal Research Engineer at the Insurance Institute for Highway Safety in Virginia, USA.

There are a few reasons why PC1 is not necessarily the best candidate for a combined injury score despite its relationship with real-world female injury risk. First, PCA identifies the combinations of loading (weighting) values that describe the greatest variance among tested vehicles, but there is no reason to believe this combination best predicts injury risk. For example, loading values differ for the left and right legs, but weighting the same injury metric differently for each leg has no biomechanical basis. Secondly, PCA requires mean-centred data so that the first component will only reflect variance within the dataset and subsequent components will be orthogonal. However, centring is an unnecessary step for defining an injury criterion and complicates interpretation since resulting values can be negative or positive. Using uncentred data produces a metric with zero as an absolute minimum and also allows any weighting and/or scaling to be performed with a single value for each measure. Third, for reasons described in the previous study [1], PCA was performed using component-level rather than resultant data for tibia moments and foot accelerations. A crash test metric using resultants would be more consistent with biomechanical studies [6] and existing injury criteria and less sensitive to slight differences in dummy positioning, kinematics, or loading. Finally, PCA involves scaling values based on the standard deviation of test data for each metric, but these also vary by leg and may be disproportionately influenced by the range expected from normal test repeatability. Knee (tibia vs. femur) displacement is the most obvious example of the latter issue. Due to the high number of good-rated vehicles with knee displacement values of zero, the standard deviation used for scaling the data during PCA was just 3 mm. Selecting scaling values that are the same for the left and right leg and that are less likely to result in large repeatability differences may produce an injury criterion that is superior to PC1.

One possible obstacle to adoption of a new lower extremity injury criterion is uncertainty surrounding the influence of contributing factors, such as dummy kinematics and interactions with vehicle structure and restraints. While video footage has been used to corroborate dummy measurements for some body regions in IIHS crash tests, it has not been used for the lower extremities. However, since December 2022, IIHS has collected high-speed video footage in the driver footwell for almost every moderate overlap crash test. Observations from these videos can be used to better understand the sources of dummy loading relevant to the new injury criterion and to identify possible countermeasures.

The purpose of this study is to define a female lower extremity injury criterion (FLEIC), building on previous findings that the overall lower extremity response of the HIII-50M dummy is predictive of female driver injury in real-world crashes. The study describes how the proposed FLEIC addresses the biomechanical and usability challenges of PC1, described above, while still relating to female injury risk. Possible alternatives to FLEIC are also evaluated. Finally, the study relates observations from high-speed videos of 100 crash tests to resulting FLEIC scores, identifying contributing factors and possible countermeasures for reducing female lower extremity injury and quantifying the repeatability of the new metric.

II. METHODS

The methods are separated into three categories: defining requirements for the FLEIC; field crash analysis; and crash test analysis.

FLEIC Requirements

As described in the Introduction, FLEIC development began with the finding that female lower extremity risk was related to the first principal component (PC1) from all HIII-50M lower extremity injury measurements for 390 vehicles with good overall and structure ratings in the IIHS moderate overlap test (version 1.0) [1]. The process of defining FLEIC from PC1 was based on the following goals:

1. Weighting factors for component metrics should be eliminated or simplified.
2. Measurements on the left and right legs should contribute equally.
3. Resultant data should be used instead of individual components.
4. Non-centred data should be used to improve interpretability.
5. FLEIC should predict female driver crash injuries to a similar degree as PC1.
6. Scaling factors for each metric should be large enough to produce repeatable FLEIC scores.

Field Crash Analysis

Theoretically, it would be possible to fine-tune FLEIC weighting and scaling factors in order to produce a metric that best fits the field crash outcomes (Item 5 in “FLEIC Requirements” above). However, this would reintroduce the risk of overfitting, defeating the purpose of conducting PCA in the first place. Instead, the guidelines above were intended to produce an injury criterion that would rank vehicles similarly to PC1 while addressing its usability and biomechanical limitations. Only after defining a potential FLEIC were field crash data analysed to determine whether the test metric predicted injury risk similar to PC1.

Field crash inclusion criteria for this study were the same as in [1]. Files from the 2000–2015 National Automotive Sampling System-Crashworthiness Data System (NASS-CDS) were queried for frontal crashes involving a driver restrained by a seat belt and deployed airbag. Frontal crashes were defined broadly (general area of damage = “F” in the collision deformation classification). Only vehicle designs represented among the 390 with good overall and structure ratings in the IIHS moderate overlap test (version 1.0) were included. Real-world cases were filtered to exclude those with toe pan or instrument panel intrusion values that correspond to a structure rating in the crash test that would be below the top rating of good. Rollovers were excluded, as were crashes with large trucks, buses or other non-passenger-vehicle partners, as these may confound the relationship with crash test measurements.

Logistic regression was used to estimate the effect of PC1 or FLEIC on lower extremity injury while controlling for calculated delta-V, driver age and driver BMI, as well as for the interaction between delta-V and the version of the WinSMASH algorithm used for the calculation. The principal direction of force (PDOF) also was included in initial models but dropped since it was not a significant predictor of injury (p=0.25 for females, p=0.71 for males). Three different lower extremity injury outcomes – any injury, any non-ankle/foot injury and any right ankle/foot injury – were considered, all with a severity of 2 or greater on the Abbreviated Injury Scale (AIS≥2). There were insufficient left foot/ankle injuries to evaluate individually. Separate models were fit for female and male drivers. In addition to PC1 and FLEIC, maximum toe pan intrusion and brake pedal intrusion from static post-test measurements were also evaluated for their potential to predict injury. All models were fit using the “survey” package in R [7] to account for NASS-CDS case weights.

Crash Test Data Analysis

Including a new FLEIC score in crashworthiness rating programmes will result in meaningful vehicle design improvements only if the factors that relate to FLEIC are understood and if results are repeatable. Between December 2022 and January 2025, IIHS recorded high-speed footwell video in 100 moderate overlap crash tests. The videos were analysed to establish the timing of dummy kinematics, restraint interactions and structural intrusion. Results were compiled to identify characteristic overall patterns in relation to the timing of increases in FLEIC. Linear regression was used to evaluate relationships between overall FLEIC scores and specifics of the test vehicle. Among the 100 tests were 11 vehicle groups that involved two or three tests of the same model or of a structurally similar model. These multiple tests were conducted for reasons related to the rear-seated dummy that would not be expected to affect driver lower extremity measurements. FLEIC repeatability was evaluated by calculating the coefficient of variation (CV) for the FLEIC scores in each group of tests, addressing Item 6 under “FLEIC Requirements” above.

III. RESULTS

FLEIC Formulation

The first candidate formulation of FLEIC involved eliminating the loading values used for the first principal component (PC1) in [1] and defining scaling values for uncentred maxima that are identical for the left and right legs, are roughly based on the standard deviations, and are large enough to facilitate repeatable scores. The other major difference from PC1 is inclusion of the resultant tibia moments and foot acceleration values instead of individual components. These steps fulfil Items 1–4 under “FLEIC Requirements” in the Methods section. Table I shows PC1 loading, mean and standard deviation values, along with FLEIC scale values. PC1 is given by:

$$\mathrm{PC1} = L_1 \frac{(X_1 - \overline{X_1})}{\sigma_1} + L_2 \frac{(X_2 - \overline{X_2})}{\sigma_2} + \cdots \quad \text{(Eq. 1)}$$

where $X_1, X_2, \cdots$ are test-specific maximum values recorded for each measure, $L_1, L_2, \cdots$ are the loadings, $\overline{X_1}, \overline{X_2}, \cdots$ are the means, and $\sigma_1, \sigma_2, \cdots$ are the standard deviations for each measure, as shown in Table I. FLEIC is calculated as:

$$\text{FLEIC} = \frac{X_1}{s_1} + \frac{X_2}{s_2} + \cdots \quad \text{(Eq. 2)}$$

where $X_1, X_2, \cdots$ are test-specific values recorded for each measure, and $s_1, s_2, \cdots$ are the scales for each measure, as shown in Table I.

TABLE I
MEASURED PARAMETERS AND CONSTANTS FOR CALCULATING PC1 AND FLEIC

Measure	PC1 Location, direction	PC1 Loading	PC1 Mean	PC1 Std. dev.	FLEIC Location, direction	FLEIC Scale
Femur axial force (kN)	Left	0.115	2.08	1.63	Left	1.5
	Right	0.276	2.26	1.37	Right	
Knee displacement (mm)	Left	0.234	1.94	3.05	Left	5.0
	Right	0.252	2.21	3.08	Right	
Tibia axial force (kN)	Left	0.216	1.91	0.77	Left	1.0
	Right	0.283	2.82	1.12	Right	
Tibia moment (Nm)	Left upper, X	0.200	48.9	20.2	Left upper, resultant	40
	Left upper, Y	0.245	70.4	32.0		
	Right upper, X	0.195	68.7	30.3	Right upper, resultant	
	Right upper, Y	0.304	76.2	34.1		
	Left lower, X	0.088	48.5	29.3	Left lower, resultant	
	Left lower, Y	0.144	52.8	28.6		
	Right lower, X	0.206	91.2	48.7	Right lower, resultant	
	Right lower, Y	0.239	63.5	34.4		
Foot acceleration (g)	Left, X	0.311	56.0	20.9	Left, resultant	30
	Left, Z	0.284	48.3	16.3		
	Right, X	0.273	67.0	29.0	Right, resultant	
	Right, Z	0.236	51.6	21.0		

Figure 1 shows the relationship between PC1 and FLEIC for 390 moderate overlap tests with good overall and structure ratings. The strong correlation between PC1 and FLEIC results from the relative similarity of most of the loading values for PC1 and the use of the standard deviations as the basis for the FLEIC scales. For the same tests, Fig. 2 shows the relative contribution of each of the component metrics to the overall FLEIC score. Maximum tibia forces, upper and lower tibia moments, and foot accelerations contribute approximately similar proportions, with the right leg and foot contributing more than the left. Femur forces and knee displacements usually contribute less to FLEIC than the other metrics.

Field Data Analysis

Logistic regression was used to evaluate the effects of different test metrics on the risk of three different lower extremity injury outcomes in frontal crashes. Separate models estimated the risk for female and male drivers while controlling for one test metric along with driver BMI, driver age, longitudinal delta-V and WinSMASH version. The evaluated test metrics were FLEIC, PC1, toe pan intrusion and brake pedal intrusion. The female and male models were based on 1,353 and 1,219 NASS-CDS cases, respectively, involving 177 matching vehicle designs out of the 390 tested with good overall and structure ratings. Distributions of delta-V and PDOF values by driver sex are shown in Fig. A1 in the Appendix. Table II shows the results for the models estimating the effect of FLEIC on the risk of any AIS≥2 lower extremity injury for female and male drivers, while Fig. 2 shows estimated effects of all test metrics for all three injury outcomes and both sexes. Results indicate that the effect of FLEIC on lower extremity injury risk for female drivers is at least as strong as the effect of PC1, fulfilling Item 5 under “FLEIC Requirements” in the Methods section. None of the crash test metrics was a statistically significant predictor of injury for male drivers, nor were there significant effects of intrusion for females. As previously mentioned, “intrusion” refers to measurements taken after the IIHS moderate overlap test, not crash-specific values from the NASS-CDS cases.

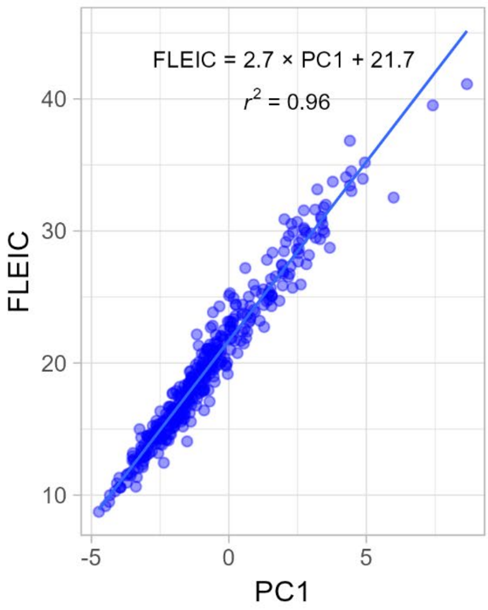

Fig. 1. PC1 vs. FLEIC for 390 moderate overlap crash tests with good overall driver and structure ratings.

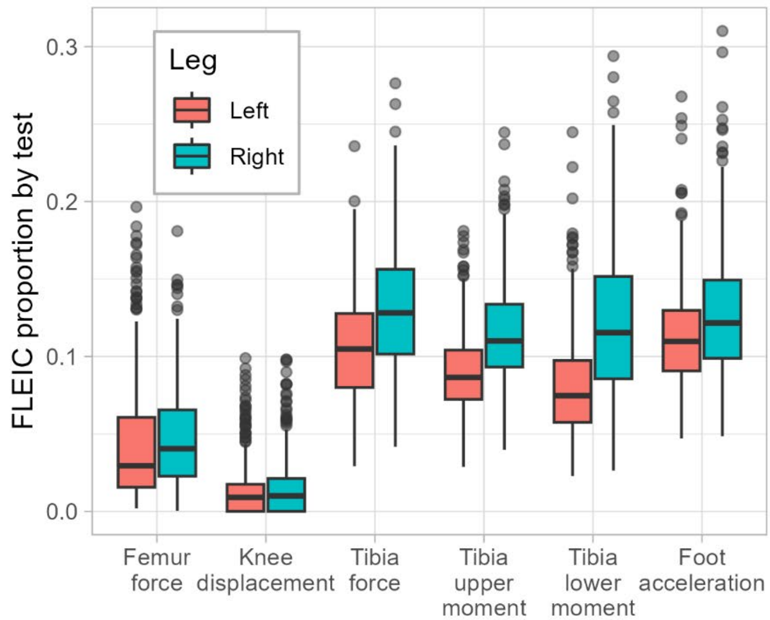

Fig. 2. Contribution of each metric to total FLEIC score in 390 moderate overlap crash tests with good overall driver and structure ratings.

TABLE II
RESULTS OF LOGISTIC REGRESSION MODELS USING FLEIC TO ESTIMATE AIS≥2 LOWER EXTREMITY INJURY RISK FOR FEMALE AND MALE DRIVERS

Driver Sex	**Term**	**Estimate**	**Standard Error**	***p*-value**
Female	Intercept	-11	1.5	NA
	Delta-V (+1 km/h)	**0.11**	**0.016**	**<0.001**
	Age (+1 year)	0.018	0.01	0.08
	BMI (+1 kg/m^2)	**0.089**	**0.032**	**0.005**
	FLEIC (+1)	**0.078**	**0.032**	**0.01**
	WinSMASH ≥ 2008: Delta-V interaction	**-0.039**	**0.013**	**0.002**
Male	Intercept	-6.6	1.7	NA
	Delta-V (+1 km/h)	**0.17**	**0.029**	**<0.001**
	Age (+1 year)	0.016	0.011	0.15
	BMI (+1 kg/m^2)	-0.071	0.085	0.41
	FLEIC (+1)	-0.043	0.036	0.24
	WinSMASH ≥ 2008: Delta-V interaction	**-0.053**	**0.018**	**0.003**

Note: estimates with $p \leq 0.05$ are shown in **bold**.

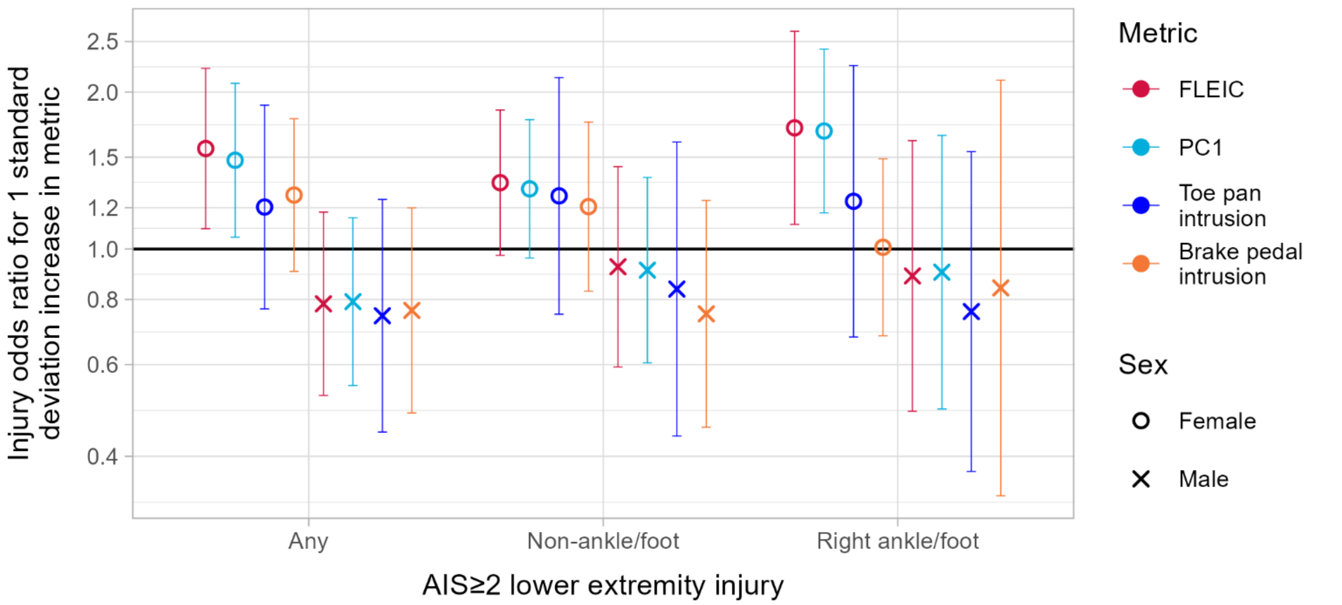

Fig. 3. Results of logistic regression models for AIS≥2 lower extremity injury risk by driver sex and crash test metric. Odds ratios are scaled to represent 1 standard deviation for each metric and are shown with 95% confidence intervals. Standard deviations: FLEIC = 5.7; PC1 = 2.1; toe pan intrusion = 4.3 cm; brake pedal intrusion = 4.2 cm.

Crash Test Analysis

Figure 4 shows results of high-speed footwell video analysis for 100 crash tests. The top portion of the Figure illustrates the observed timing of different events related to the dummy kinematics, structure and restraint system, while the bottom portion shows how loading in each 10 ms window contributed to a vehicle's FLEIC. The values for each 10 ms window were calculated by constructing a time-based cumulative FLEIC score for each test, determining how much this score increased during each time window, and then dividing this value by the final FLEIC for the test. Based on limitations of the video coverage, not every event could be observed in every test. Some events never occurred in certain tests, such as visible dynamic intrusion or dummy contact with the instrument panel.

Overall dummy kinematics and loading patterns were similar for most vehicles and involved an initial leg/foot excursion phase, a ride-down phase with minimal intrusion, and a third phase in which intrusion and forward excursion of the entire dummy contributed to lower extremity loading. One major factor in the timing of the associated FLEIC scores was related to the presence of a knee airbag. Vehicles with knee airbags (n=80) often had a substantial proportion of their overall score associated with initial knee airbag contact, especially when this contact occurred while the knee airbag was still deploying. In contrast, vehicles without knee airbags usually had a greater proportion of overall FLEIC scores during the later intrusion and dummy-excursion phase. The time period with visible dynamic toe pan intrusion lasted for an average of 37 ms for all vehicles and accounted for mean proportions of 36% and 64% of the total FLEIC for vehicles with and without knee airbags, respectively. These proportions do not necessarily represent the direct contribution of intrusion to FLEIC but rather the contribution of all dummy loading during the time when there also was dynamic intrusion.

All comparisons drawn from Fig. 4 are based on values normalised to the total FLEIC for each test; they do not allow an evaluation of the factors that might contribute to higher absolute FLEIC scores. A linear regression model was constructed to facilitate these comparisons; results are shown in Table III. Increasing toe pan intrusion (measured statically) and resultant vehicle acceleration were associated with increasing FLEIC scores. The presence of a knee airbag was associated with a non-significant FLEIC increase of 0.9. The model's r-squared value was 0.34, indicating that there were still large relative differences in FLEIC among vehicles with similar intrusion and vehicle acceleration.

FLEIC repeatability was evaluated among the 11 vehicle groups with at least two tests. The coefficient of variation (CV) for FLEIC ranged from 0.05% to 8.6% (Fig. 5). The mean FLEIC CV was 4.6%, which was lower than CV values for any of the dummy measures that contribute to FLEIC. (Knee displacement CVs are not shown in Fig. 5 given the large number of cases without any displacement.)

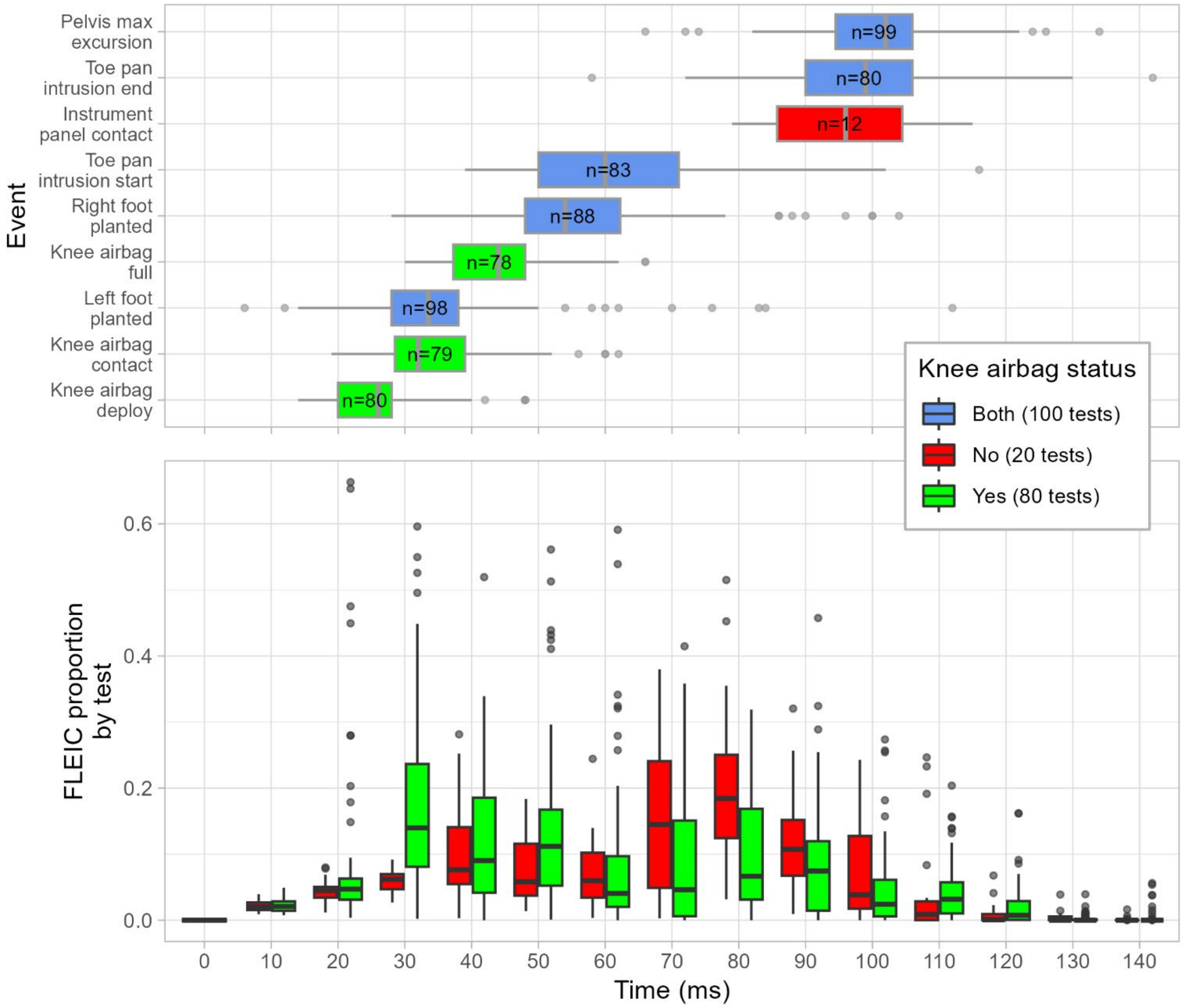

Fig. 4. Timing of crash test events (top) and the proportion of the total FLEIC score for each vehicle produced by loads during 10 ms intervals (bottom).

TABLE III
RESULTS OF LINEAR REGRESSION MODEL PREDICTING CRASH TEST FLEIC

Term	Estimate	SE	*p*-value
Intercept	9.6	1.6	NA
Max. toe pan intrusion (+1 cm)	**0.45**	**0.073**	**<0.001**
Max. vehicle resultant acceleration (+1 g)	**0.061**	**0.031**	**0.05**
Knee airbag (vs. none)	0.87	0.84	0.3

Note: model r^2 = 0.34. SE = standard error. Estimates with $p \leq 0.05$ are shown in **bold**.

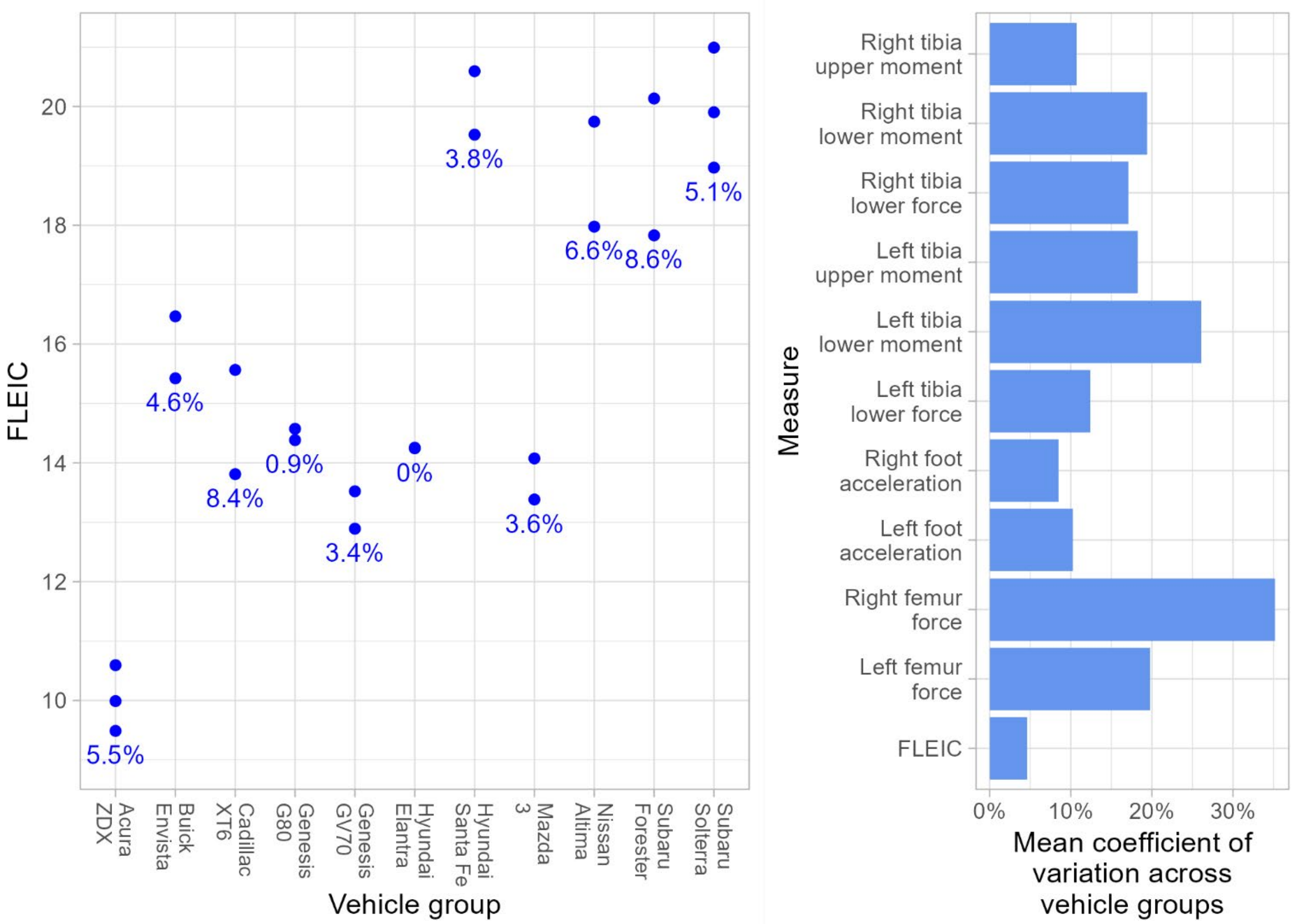

Fig. 5. FLEIC coefficient of variation for 11 vehicle groups with repeat tests (left) and mean coefficient of variation across the groups for FLEIC and all contributing measures (right).

IV. DISCUSSION

The proposed female lower extremity injury criterion (FLEIC) is a combination of lower extremity injury measures recorded with the HIII-50M dummy. In the IIHS moderate overlap test, it is strongly correlated with the first principal component of the dummy measures (Fig. 1) and is a significant predictor of lower extremity injury for female drivers in front crashes (Fig. 3). A FLEIC increase of around 5 units, just under 1 standard deviation, was associated with 50–60% greater odds of a female driver sustaining any AIS≥2 lower extremity injury or a right ankle/foot injury. This relationship exists for vehicles with good overall driver and structure ratings in real-world crashes with limited intrusion, suggesting that incorporating FLEIC into front crash test programmes could encourage countermeasures that reduce the risk of female lower extremity injury.

Existing IIHS lower extremity ratings using the HIII-50M are assigned based on comparing several different measures with individual rating boundaries and then assigning the worst overall rating category to that leg. This is fundamentally different from the proposed FLEIC, for which rating assignment would only occur after combining metrics. While the lowest FLEIC scores will necessarily correspond to good ratings under the existing protocol, the opposite is not true. Figure 6 shows the current lower extremity ratings for all vehicles with good overall driver and structure ratings along with the associated FLEIC. While there is a general relationship between current ratings and FLEIC, there also is a wide range of FLEIC scores for vehicles with the same existing rating. This is further evidence that using FLEIC could drive improvements beyond those currently required for good ratings.

FLEIC Alternatives and Potential Countermeasures

A possible alternative to FLEIC would be adjustment of one or more of the existing dummy lower extremity rating boundaries. It is possible that the predictive value of FLEIC in real-world crashes is driven by a subset of all the component metrics. However, as the individual metrics are covariant, determining which ones best predict field outcomes would require a much larger dataset than currently available. This is illustrated by Fig. 2, which shows

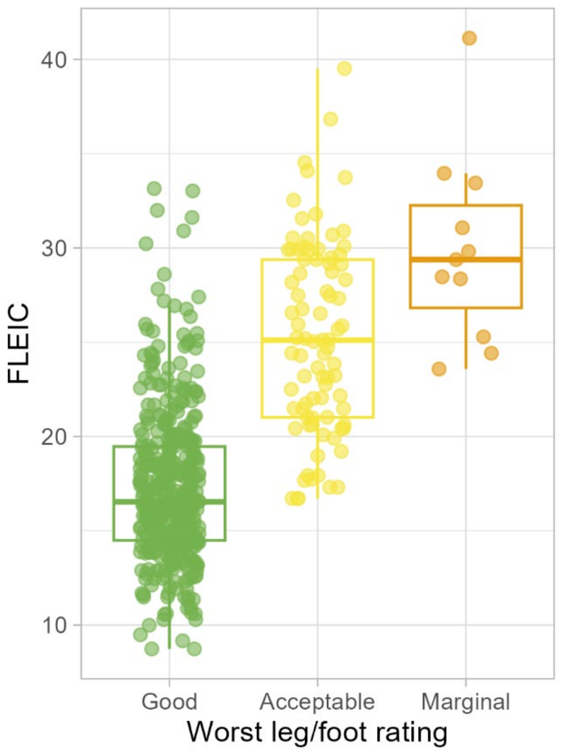

Fig. 6. FLEIC vs. existing lower extremity rating for vehicles with good overall driver and structure ratings.

that the contribution of most metrics to the overall FLEIC is similar and that the outliers still contribute relatively small proportions; even the highest values are unlikely to become disproportionate contributors since other metrics generally increase as well. This covariance means that not only would choosing which metrics to rate more strictly be difficult to support analytically, it also would be unlikely to produce different countermeasures than a score based on the combination of all metrics. Furthermore, FLEIC's lower coefficient of variation makes it a better suited alternative rating criterion than reduced thresholds for any of the component metrics (Fig. 5).

Reducing the amount of intrusion allowed for a good rating is another change that could benefit female drivers. As with FLEIC, higher toe pan intrusion was associated with higher injury risk for female drivers but not for males (Fig. 3). However, the effect magnitudes were lower for intrusion than for FLEIC and were not statistically significant ($p \geq 0.17$). Greater benefits would likely be derived from prioritising FLEIC reductions rather than intrusion reductions, although for some vehicle designs these will coincide. In the 100 recent tests analysed in detail, the intrusion phase accounted for a large proportion of the total FLEIC, especially for vehicles without knee airbags (Fig. 4), and the magnitude of intrusion was a significant predictor of FLEIC for all vehicles (Table III). However, only around one-third of the total variation in FLEIC was explained by intrusion magnitude. Video analysis showed that loading within 10 ms of the beginning of visible intrusion contributed to higher FLEIC scores than loading within 10 ms of the end of intrusion. This suggests that dynamic measures of intrusion, such as velocity or acceleration, may be more relevant to FLEIC in this test configuration than intrusion magnitude. While these quantities are difficult to measure directly in a physical crash test, reducing intrusion velocity and acceleration are likely important means of reducing FLEIC and the associated female injury risk.

Another change that would be needed to reduce FLEIC for some vehicle designs is a reduction of the loads created by the deploying knee airbag. In most equipped vehicles, knee airbag contact with the dummy legs occurred prior to maximum deployment. This usually appeared minor and generated relatively low tibia moments. However, in a few cases the most significant loading during the entire test was generated by the deploying knee airbag. Vehicle manufacturers may include knee airbags to manage overall occupant loading and kinematics rather than as a means to reduce lower extremity injury risk, and real-world analyses of knee airbag effectiveness have been mixed [1][8-10]. Knee airbag benefits for lower extremities could become more evident in the future if they were designed to minimise loading during the deployment phase.

Incorporating FLEIC into frontal test programmes also may encourage faster development and implementation of novel countermeasures. Prior studies have explored the potential benefits of placing an energy-absorbing structure [11] or an airbag [12-13] between the feet and toe pan. Some vehicles are already equipped with pelvic restraint cushions that could reduce lower extremity injury risk [14]. While restraint suppliers and vehicle manufacturers may continue developing these technologies voluntarily, updated crashworthiness evaluation programmes could encourage more widespread adoption.

Dummy Considerations

Introducing a new injury metric using the HIII-50M may seem counterintuitive given the additional lower extremity measurement capabilities of the THOR 50th percentile male (THOR-50M) dummy as well as its more human-like geometry and mass properties [15]. Indeed, many studies have found THOR-50M is more biofidelic when evaluated in most – though not all – component-level tests (e.g. [16-17]). However, whole-body comparisons between the dummies have often been limited by unrealistic lower extremity restraint conditions, such as rigid zero-clearance knee bolsters. In sled tests with a driver airbag and seat belt, [18] reported that HIII-

50M had better overall lower extremity biofidelity than THOR-50M. These tests were still limited by the lack of field crash factors, such as intrusion or vehicle rotation, and by the lack of active braking and/or other bracing effects likely exhibited by the majority of female drivers who sustain lower extremity injury [1]. In reality, determining which dummy is more biofidelic in the specific crash conditions most likely to produce injury in the real world is not possible. But the validity of FLEIC does not directly depend on the ability of HIII-50M to precisely reproduce the specific loading durations and magnitudes experienced by female drivers. Rather, FLEIC is a more general characterisation of the lower extremity injury risks represented by a specific vehicle design.

It is possible that a FLEIC-like score developed using THOR-50M test data would provide an even more robust prediction of real-world injury risk than one using HIII-50M, but there are insufficient data to make this determination. The NASS-CDS cases analysed for this study were matched to 177 different vehicle crash tests conducted over many years with HIII-50M. While similar comparisons to field crash outcomes cannot yet be made with THOR-50M tests, it is possible to make a more limited comparison of the lower extremity response of the two dummies in the moderate overlap test. IIHS conducted paired tests of 39 vehicles with both a HIII-50M and THOR-50M for a previous study evaluating thoracic injury risk [19]. PCA of the lower extremity measurements was possible using 27 of these pairs. The first principal component (PC1) scores for each dummy were positively correlated, but the association was weak (r^2=0.33) [20]. While such a small sample cannot be used to draw definitive conclusions, the existing data suggest the two dummies would often rank the same vehicle differently within the fleet, making it challenging for a composite THOR-based score to predict injury risk as well as FLEIC. However, this does not address the possibility that some individual THOR metrics better predict injury.

Another seeming incongruity surrounding FLEIC is the concept that a score calculated from a dummy designed to represent a midsize male, whether HIII-50M or THOR-50M, is a valid measure of female injury risk. However, the available field crash data suggest precisely this; FLEIC scores are correlated with injury outcome for female drivers but not for males (Fig. 3). While reasons for the difference remain unknown, there is no evidence that it is explained by sex-specific differences in the field crashes relative to the test; if anything, crashes with male drivers more closely match the test conditions (Fig. A1). Regarding the dummy, HIII-50M may represent female drivers better than would be initially assumed. As previously reported [1], the real-world injury risk for females diverges from that for males as BMI increases. The HIII-50M mass is equal to the median mass (78 kg) of the female drivers who sustained a lower extremity injury in this study and is less than the mean mass (86 kg). Additionally, for a female with the median mass and BMI of those sustaining an injury, the anthropometry models created by the University of Michigan Transportation Research Institute suggest a typical seated buttock-knee length of 596 mm [21], which is extremely close to the HIII-50M specification of 592 mm [22]. (For comparison, dimensions for the THOR-50M and HIII 5th percentile female dummies and for injured male drivers are shown in Table AI in the Appendix.) While the dummy does not represent obese female drivers in other important ways, including lap-belt fit [23], from a dimensional standpoint it may be a reasonable surrogate for evaluating loading from the knee bolster and toe pan.

Limitations

Despite its ability to identify levels of female lower extremity risk beyond methods used in existing crash test protocols, on its own FLEIC is unlikely to be a sufficient means of addressing the discrepancy between female and male lower extremity injury risk. It functions as a general assessment of vehicle crashworthiness without identifying the precise vehicle-specific injury mechanisms to which females are most susceptible. Using the same example cited above [23], incorporating FLEIC into crash ratings is unlikely to promote the changes in lap-belt geometry that may be needed specifically for obese female drivers. In addition to belt fit, sex-based differences in foot placement, leg posture, and braking and bracing behaviour all may contribute to differences in injury risk, and the ongoing research in these areas is essential.

V. CONCLUSION

This study proposes a new female lower extremity injury criterion (FLEIC) to be used in front crash tests. FLEIC is a combined score based on maximum femur, knee, tibia and foot measurements taken with a HIII-50M dummy. For vehicles that currently have good overall driver injury and structure ratings in the IIHS moderate overlap test, FLEIC is a significant predictor of lower extremity injury for female drivers in front crashes. Video analysis of 100 recent crash tests demonstrates that loading from the toe pan, usually prior to peak intrusion, and from knee airbag deployment commonly contributes to FLEIC. FLEIC exhibits good repeatability, with CV values lower than its individual components. While female-specific lower extremity injury risk factors need further study, including FLEIC in existing rating protocols represents a relatively quick means of encouraging more protective vehicle designs.

VI. ACKNOWLEDGEMENTS

I am grateful for the team of technicians and test coordinators at IIHS, whose consistency and attention to detail enabled the quality of test data and video necessary for this study. I would also like to acknowledge the staff of the Crash Investigation Division at NHTSA. Without their work it would not be possible to establish the real-world relevance of IIHS test programs.

VII. REFERENCES

[1] Brumbelow, M. L. (2023) Female driver lower extremity injury: Contributing factors and crash test relevance. *Proceedings of IRCOBI Conference*, 2023, Cambridge, United Kingdom.

[2] Forman, J., Poplin, G. S., *et al.* (2019) Automobile injury trends in the contemporary fleet: Belted occupants in frontal collisions. *Traffic Injury Prevention,* **20**(6): pp. 607–612.

[3] Craig, M. J., Liu, C., Zhang, F., Enriquez, J. (2024) Sex-based differences in odds of motor vehicle crash injury outcomes. *Accident Analysis & Prevention*, **195**: 107100.

[4] Brumbelow, M. L., Jermakian, J. S. (2022) Injury risks and crashworthiness benefits for females and males: Which differences are physiological? *Traffic Injury Prevention*, **23**(1): pp. 11–16.

[5] IIHS (2024) Moderate overlap crashworthiness evaluation 2.0 rating guidelines. Available: https://www.iihs.org/ratings/about-our-tests/test-protocols-and-technical-information.

[6] Crandall, J. R., Funk, J. R., Rudd, R. W., Tourret, L. J. (2000) The tibia index: A step in the right direction. In: Kajzer, J., Tanaka, E., Yamada, H. (eds), *Human Biomechanics and Injury Prevention*, pp. 29–40. Springer: Japan.

[7] Lumley, T. (2004) Analysis of complex survey samples. *Journal of Statistical Software*, **9**(1): pp. 1–19.

[8] McMurry, T. L., Forman, J. L., Shaw, G., Crandall, J. R. (2020) Evaluating the influence of knee airbags on lower limb and whole-body injury. *Traffic Injury Prevention*, **21**(1): pp. 72–77.

[9] Weaver, A. A., Loftis, K. L., Stitzel, J. D. (2013) Investigation of the safety effects of knee bolster air bag deployment in similar real-world crash comparisons. *Traffic Injury Prevention*, **14**(2): pp. 168–180.

[10] Monfort, S. S., Mueller, B. C. (2019) *Effectiveness of knee airbags across two crash paradigms*. IIHS: Ruckersville, VA, USA.

[11] Rudd, R. W., Kitagawa, Y., Crandall, J. R., Poteau, F. C. (2004) Evaluation of energy-absorbing materials as a means to reduce foot/ankle axial load injury risk. *Proceedings of the Institution of Mechanical Engineers*, *Part D: Journal of Automobile Engineering*, **218**(3): pp. 279–293.

[12] Håland, Y., Hjerpe, E., Lövsund, P. (1998) An inflatable carpet to reduce the loading of the lower extremities-evaluation by a new sled test method with toepan intrusion. *Proceedings of 16th ESV Conference*, 1998, Windsor, Ontario, Canada, pp. 292–301.

[13] Freisler, W., Fischer, C., Tamme, S., Elija, D., Kramer, S. (2024) Interior of the future and new seating positions require new airbag technology to improve lower leg protection of occupants. *Proceedings of the Airbag 2024 Conference*, Mannheim, Germany.

[14] Shaw, G., Lessley, D., *et al.* (2018) Pelvic restraint cushion sled test evaluation of pelvic forward motion. *Traffic Injury Prevention*, **19**(3): pp. 250–255.

[15] Shams, T., Beach, D., *et al*. (1999) Development and design of THOR-Lx: The THOR lower extremity. *Proceedings of the 43rd Stapp Car Crash Conference*, 1999, San Diego, CA, USA.

[16] Petit, P., Trosseille, X. (1999) Comparison of the THOR, Hybrid III and cadaver lower leg dynamic responses

in dorsiflexion. *Proceedings of the 43rd Stapp Car Crash Conference*, 1999, San Diego, CA, USA.

[17] Parent, D., Craig, M., Moorhouse, K. (2017) Biofidelity evaluation of the THOR and Hybrid III 50 percentile male frontal impact anthropomorphic test devices. *Stapp Car Crash Journal,* **61**: pp. 227–276.

[18] Albert, D. L., Beeman, S. M., Kemper, A. R. (2023) Comparative biofidelity of the Hybrid III and THOR 50th male ATDs under three restraint conditions during frontal sled tests. *Traffic Injury Prevention*, **24**(sup1): pp. S41–S46.

[19] Brumbelow, M. L., Jermakian, J. S., Arbelaez, R. A. (2022) Predicting real-world thoracic injury using THOR and Hybrid III crash tests. *Proceedings of IRCOBI Conference*, 2022, Porto, Portugal.

[20] Brumbelow, M. L. (2023) Comment to NHTSA Notice of Proposed Rulemaking: Anthropomorphic Test Devices; THOR 50th percentile adult male test dummy; Incorporation by Reference. Available: https://www.regulations.gov/comment/NHTSA-2023-0031-0015.

[21] UMTRI (2020) HumanShape: Realistic human body shape modeler based on real measurement data. Ann Arbor, MI, USA. Available: https://humanshape.org.

[22] Humanetics (2017) *Harmonized Hybrid III 50th male user manual [Rev. A]*. Farmington Hills, MI, USA. Available: https://www.humaneticsgroup.com/sites/default/files/2020-11/um-h-h3-50m_rev_a_0.pdf.

[23] Jones, M. L. H., Ebert, S. M., *et al*. (2021) Effect of Class I–III obesity on driver seat belt fit. *Traffic Injury Prevention*, **22**(7): pp. 547–552.

[24] NHTSA (2023) Anthropomorphic Test Devices; THOR 50th Percentile Adult Male Test Dummy; Incorporation by Reference. Available: https://www.federalregister.gov/documents/2023/09/07/2023-19008/anthropomorphic-test-devices-thor-50th-percentile-adult-male-test-dummy-incorporation-by-reference.

[25] Humanetics (2020) Hybrid III 5th female user manual [Rev G]. Farmington Hills, MI, USA. Available: https://www.humaneticsgroup.com/sites/default/files/2020-11/880105-9900-h3-5th-user-manual.pdf.

VIII. APPENDIX

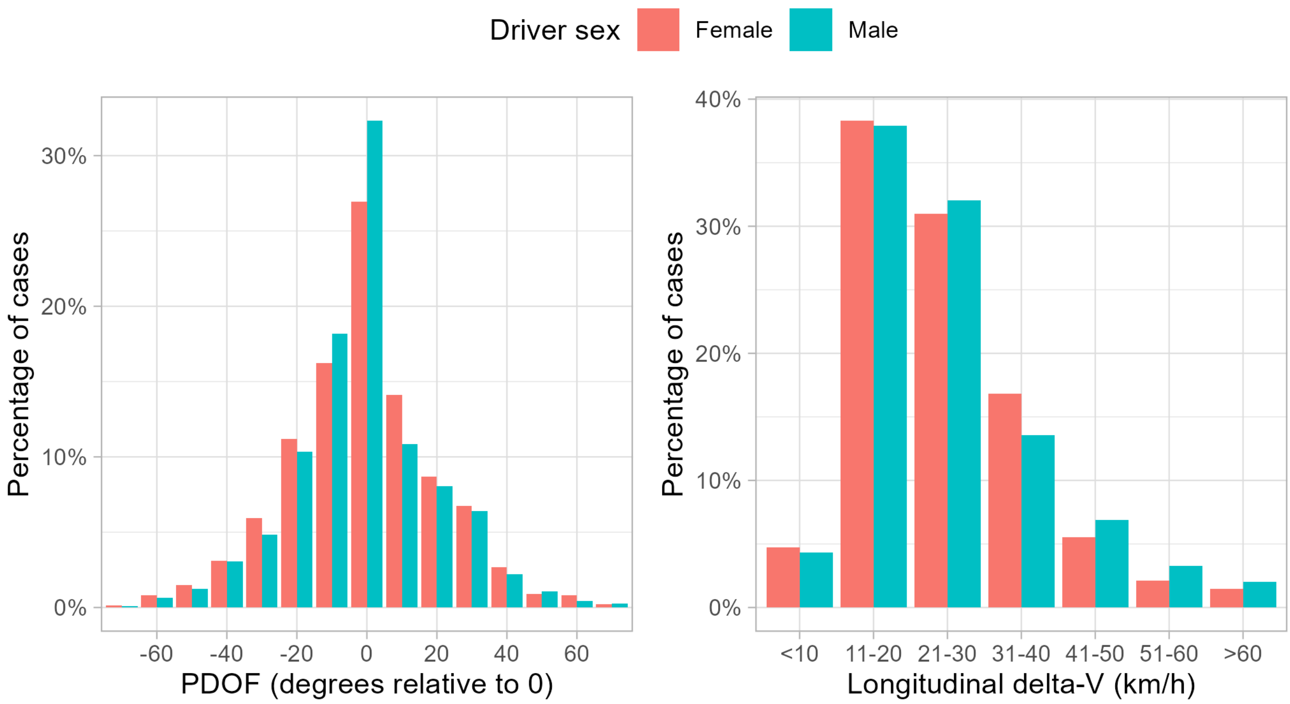

Fig. A1. Distributions of PDOF and delta-V for NASS-CDS cases by driver sex.

TABLE AI

MEDIAN ANTHROPOMETRY FOR DRIVERS WITH AIS≥2 LOWER EXTREMITY INJURY COMPARED WITH DUMMY DIMENSIONS

Occupant	Mass (kg)	Buttock-knee length (mm)	Sitting height (mm)
Injured female drivers (median)	78	596*	843*
Injured male drivers (median)	83	623*	935*
HIII-50M [22]	78	592	884
THOR-50M [24]	77	620	1018
HIII-5F [25]	49	533	787

*estimated using UMTRI models [19] for subjects with the median age, mass and BMI of injured drivers