Can front crash rating programs using Hybrid III predict real-world thoracic injuries?

Matthew L. Brumbelow

Abstract In front crashes, belted drivers of vehicles with good crash test ratings sustain serious injuries to the thorax at much higher rates than to other non-extremity body regions. The full width US NCAP and moderate overlap IIHS evaluation programs both assess driver injury risk using the Hybrid III 50th percentile male dummy. The current study evaluated the ability of peak sternum deflection and other dummy and vehicle metrics to predict NASS-CDS thoracic injury outcomes for drivers restrained by a seat belt and airbag while controlling for delta-V and driver age. Delta-V estimates were based on adjustments derived from event data recorder output for specific front crash configurations. Results showed that two vehicle metrics, bumper-to-firewall distance and shoulder belt force, each were better able to predict outcomes across the range of driver ages and thoracic injury types than any single dummy metric. Some dummy metrics did improve injury prediction for drivers aged 50 and older, including sternum deflection, deflection rate, and the ratio of sternum deflection to thoracic acceleration. However, none of these findings was significant at the $p = 0.05$ level for both soft- and hard-tissue injury, and reductions in the predictive ability of most metrics were associated with procedural changes made to each test program.

Keywords Crashworthiness, Hybrid III, IIHS, NCAP, thoracic injury.

I. INTRODUCTION

In 2018, 57% of belted occupant fatalities in passenger vehicles less than five years old occurred in front non-rollover crashes [1]. While good performance in US front crash test ratings is associated with lower fatality risk [2-3], a 2019 study found that there is still substantial risk of serious injury to belted drivers in real-world crashes at severities similar to the crash tests [4]. This was especially true of thoracic injuries, and especially for older drivers, with those aged 60 years or older facing a 60% risk of sustaining a thoracic injury of Level 3 or greater on the Abbreviated Injury Scale (AIS) in a crash with a delta-V of 70 km/h. Further improvements in the front crashworthiness of the passenger vehicle fleet require reductions to the current level of thoracic injury risk.

The two longest running consumer front crashworthiness evaluation programs in the US, i.e. the full width New Car Assessment Program (NCAP) and the moderate overlap Insurance Institute for Highway Safety (IIHS) tests, both evaluate driver injury risk using the Hybrid III 50th percentile male dummy. While sternum deflection rate, viscous criterion and thoracic acceleration also have been used to assess thoracic injury risk, peak sternum deflection is the only common metric currently used in both tests as well as in other NCAPs worldwide. Peak deflection has been shown to correspond to cadaver rib deflections and to predict the number of fractures when the dummy and cadaver are loaded under identical conditions [5-6]. However, the range of real-world front crash conditions, restraint system designs and occupant ages, sizes, and positions all may contribute to variability between the injury risks predicted from crash test measurements and actual injury outcomes in the field. While these other factors may change the magnitude of the effect of a dummy metric on thoracic injury likelihood, as long as they do not negate or invert the effect, the metric still has value in assessing vehicle designs in crashworthiness evaluations. This has not been clearly demonstrated with field data, however, and an analysis of paired dummy and cadaver sled tests has suggested that the single-point Hybrid III deflection measurement may not be able to distinguish between the likelihood of different injury outcomes across a range of seat belt and airbag load-sharing conditions [7].

Brumbelow and Farmer [8] studied the effect of Hybrid III sternum deflection measured in the IIHS moderate overlap test on the probabilities of AIS≥3 thoracic and AIS≥3 non-extremity injury in the National Automotive Sampling System-Crashworthiness Data System (NASS-CDS). All vehicles had good test ratings in order to reduce the effect of vehicle pulse and intrusion differences. They found that the probability of both injury outcomes

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

increased in real-world moderate overlap, large overlap, and centre impact crashes with increasing test sternum deflection. However, in small overlap crashes the relationship was reversed, with greater (moderate overlap) test deflections associated with reduced risk of injury. Differences in the patterns of injuries suggested that those occurring in small overlaps were more likely associated with excursion and intrusion, while those in other crash modes were more consistent with restraint loading. The research also found that in Fatality Analysis Reporting System front-to-front crashes with two belted drivers and one death, the effect of test sternum deflection was not a significant predictor of which driver was killed. Two major limitations of the NASS-CDS analyses were the coarse control of crash severity based on vehicle damage and the small sample size. As small overlap crashes were included in the analyses, calculated delta-Vs were not used as a severity control, while the small number of cases in each crash group made it impossible to control for differences in driver age.

The goal of the current study was to address the limitations of the Brumbelow and Farmer analysis, while also assessing the predictive ability of other injury metrics from the moderate overlap and full width crash tests. The sample size was improved by the five additional years of NASS-CDS data that are now available. Regarding crash severity controls, a 2019 study [4] described a procedure for combining front crash configurations with event data recorder (EDR) output to determine the configuration-specific accuracy of damage-based delta-Vs and to adjust them where feasible. This study found that damage-based delta-Vs in small overlap crashes are almost unrelated to EDR-reported delta-Vs, and that these crashes should not be included in risk analyses with crashes that have more overlap. Even without this discrepancy, including small overlap crashes in the current analysis would be complicated by the relatively recent design changes made to improve crashworthiness in this mode. The IIHS small overlap evaluation began in 2012, and the proportion of models receiving good ratings has increased from 12% to 85% between then and 2020. As 2015 was the last calendar year included in NASS-CDS, it will likely be several more years before the replacement Crash Investigation Sampling System contains sufficient data to analyse modern vehicles with good small overlap ratings. Until then, the possibility that the observed reductions in intrusion and improvements in dummy engagement with the restraint system have addressed the discrepancy between sternum deflection and real-world small overlap injuries cannot be evaluated.

II. METHODS

The main components of the study methods were the acquisition of field crash data, selection and processing of matching crash test parameters, and statistical modeling of the relationship between the test parameters and thoracic injury outcomes. This process is illustrated in the Appendix (Fig. A1) and described in detail below.

Field Data

A previous analysis of NASS-CDS front crashes provided the field data for the current study [4]. NASS-CDS was a sample-weighted survey of police-reported crashes in the US conducted from 1979 to 2015. Front crashes involving a vehicle with a good rating in the IIHS moderate overlap test were included if the driver was age 18 or older and restrained by a three-point seat belt and deployed front airbag. For the 2019 study, delta-Vs from vehicles equipped with EDRs were used to calculate “EDR-equivalent” delta-Vs for vehicles without EDRs based on the WinSMASH delta-V and the front crash configuration. Real-world configurations other than moderate overlap, large overlap, or centre impact were determined to have delta-Vs that could not be reliably adjusted, therefore only front crashes with those configurations were included in the current study. These three configurations accounted for 61% of the weighted drivers sustaining an AIS≥3 thoracic injury.

Figure 1 shows the driver age, delta-V and thoracic injury status for all moderate overlap, large overlap and centre impact crashes. As the objective of the current study was to determine the relevance of dummy measures obtained in two crash test programs (with mean test delta-Vs of 64 and 70 km/h for the full width and moderate overlap evaluations, respectively), the full range of real-world crash speeds was not considered. While the premise of consumer evaluation crash testing is that the variation in vehicle performance observed at a specific impact speed and configuration will be meaningful in a range of real-world conditions, greater divergence from the test conditions inevitably will result in reduced test relevance. However, the choice of appropriate case selection filters also must account for the requirement for sufficient data to produce meaningful results from regression analyses. To balance these demands, NASS-CDS crashes with delta-Vs from 30 to 110 km/h were included. These accounted for 423 raw cases, 41 of which had an AIS≥3 thoracic injury. Limiting to this delta-V range included 91% and 84% of the raw and weighted cases with AIS≥3 thoracic injuries, respectively, and 45%

and 28% of the raw and weighted cases without serious thoracic injury, respectively. By comparison, cases with delta-Vs below 30 km/h and drivers aged 45 years or younger accounted for 44% of the weighted sample without a single occurrence of AIS≥3 thoracic injury.

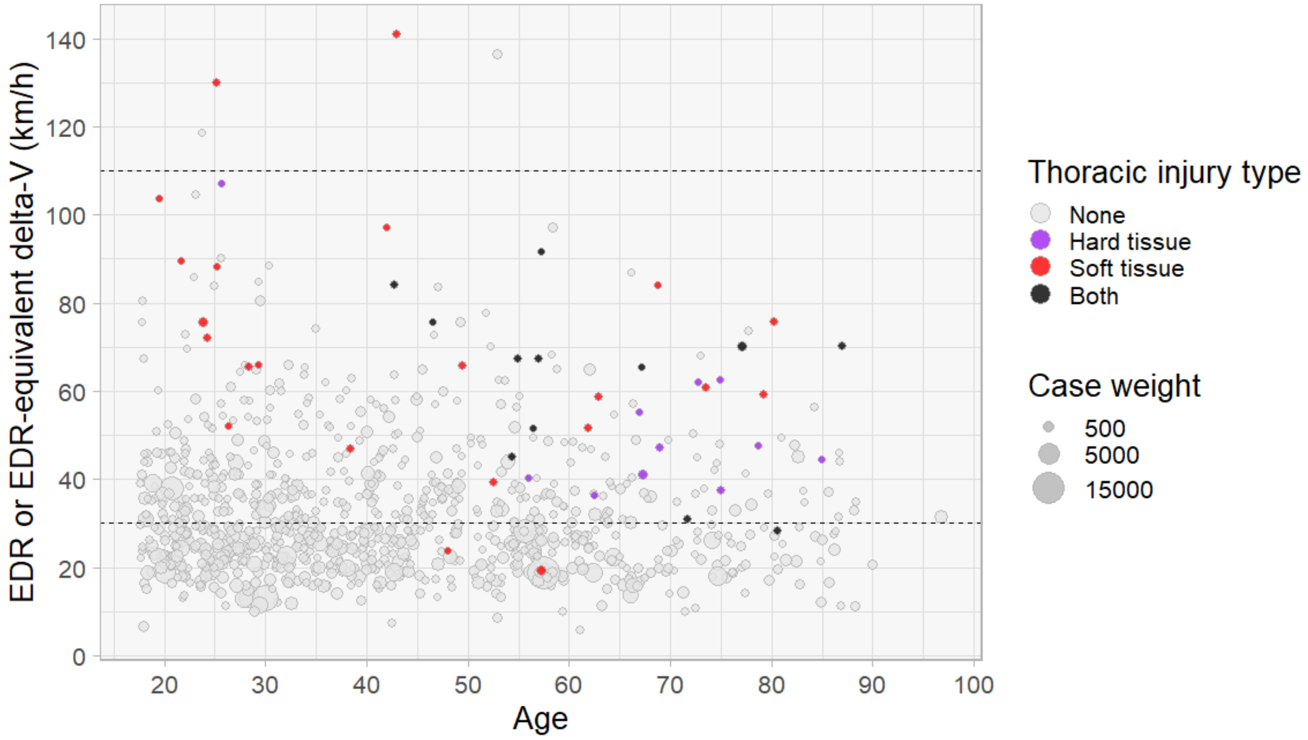


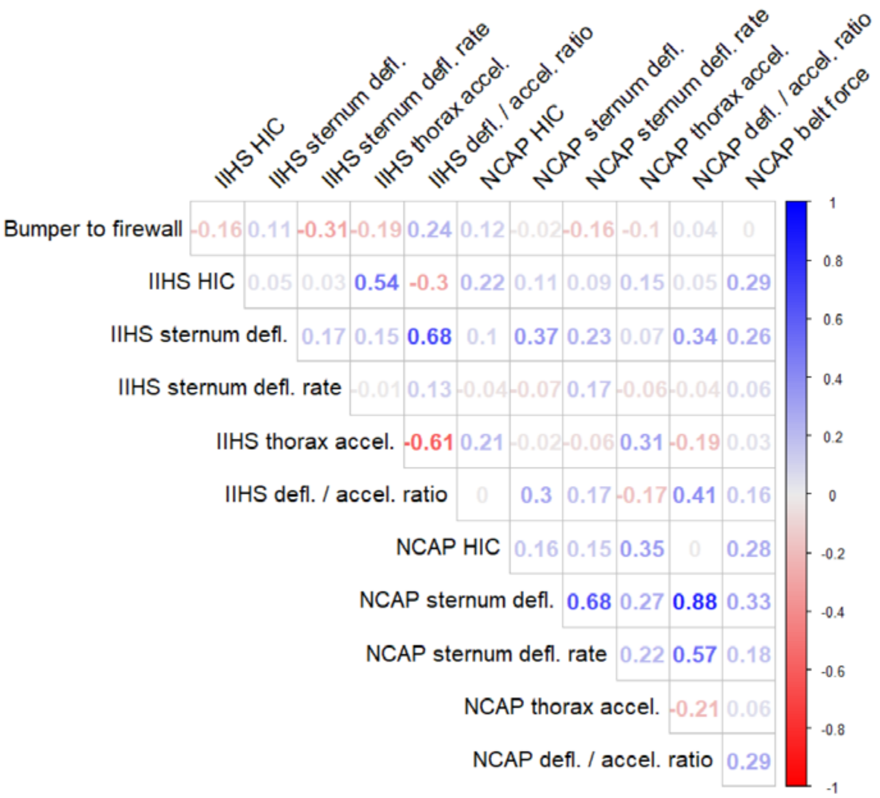
Fig. 1. Delta-V, driver age and AIS≥3 thoracic injury status for large overlap, moderate overlap and centre impacts. Cases with delta-Vs between 30 and 110 km/h were included in the current study.

Crash Test Data

Each vehicle in the NASS-CDS sample was matched with corresponding crash test data from the IIHS moderate overlap and NCAP full width tests based on make, model and model year. Hybrid III metrics included in the analyses were sternum deflection, sternum deflection rate, resultant thoracic acceleration (3 ms clip), head injury criterion (HIC15), and the ratio between peak sternum deflection and peak thoracic acceleration. Each of these metrics was available for both test conditions. In addition, shoulder belt force was available for most full width tests. The various types of belt force limiters produce a wide range of force time histories, with large relative differences in the timing of the peak force. For this reason, and to avoid the influence of temporary peaks, the metric selected for analysis was the maximum force level sustained for at least 20 ms over the entire pulse. Bumper-to-firewall distance was included as a second vehicle-based metric. Various vehicle acceleration metrics in the two tests were considered as well, but bumper-to-firewall distance was used in the final models since it was available for all vehicles, was correlated with most acceleration metrics, and eliminated the need to establish inclusion criteria based on the large number of full width tests with questionable data. Some other pairs of selected parameters also had high correlations, but they were retained to provide a more complete comparison between the two test modes. Correlations between all included metrics are shown in Fig. 2.

Previous research has shown that shoulder belt placement can affect peak Hybrid III sternum deflection [9-11]. In addition to the reported sternum deflection, the effect of adjusted deflections on injury outcome was also estimated. The adjustments were made based on the location of the shoulder belt as measured from pretest photographs. The adjustment process is described in a separate study [9].

NCAP data channels were downloaded from the National Highway Traffic Safety Administration's (NHTSA's) online test database [12] and then filtered according to SAE J211 [13]. All test data processing was performed in DIAdem [14]. Vehicle bumper-to-firewall distances were extracted from NCAP report files using the "pdftools" package [15] in the R programming language [16].



	IIHS HIC	IIHS sternum defl.	IIHS sternum defl. rate	IIHS thorax accel.	IIHS defl. / accel. ratio	NCAP HIC	NCAP sternum defl.	NCAP sternum defl. rate	NCAP thorax accel.	NCAP defl. / accel. ratio	NCAP belt force
Bumper to firewall	-0.16	0.11	-0.31	-0.19	0.24	0.12	-0.02	-0.16	-0.1	0.04	0
IIHS HIC		0.05	0.03	0.54	-0.3	0.22	0.11	0.09	0.15	0.05	0.29
IIHS sternum defl.			0.17	0.15	0.68	0.1	0.37	0.23	0.07	0.34	0.26
IIHS sternum defl. rate				-0.01	0.13	-0.04	-0.07	0.17	-0.06	-0.04	0.06
IIHS thorax accel.					-0.61	0.21	-0.02	-0.06	0.31	-0.19	0.03
IIHS defl. / accel. ratio						0	0.3	0.17	-0.17	0.41	0.16
NCAP HIC							0.16	0.15	0.35	0	0.28
NCAP sternum defl.								0.68	0.27	0.88	0.33
NCAP sternum defl. rate									0.22	0.57	0.18
NCAP thorax accel.										-0.21	0.06
NCAP defl. / accel. ratio											0.29

Fig. 2. Pearson correlation coefficients for the dummy and vehicle metrics selected as predictors of thoracic injury.

Regression Models

The effect of each dummy and vehicle metric was estimated for three different real-world thoracic injury outcomes: the risk of any serious (AIS≥3) thoracic injury for drivers aged 18–55 years (outcome A), the risk of a serious soft-tissue injury for drivers older than 50 years (outcome B), and the risk of a serious hard-tissue (skeletal) injury for drivers older than 50 years (outcome C). These outcomes were selected to account for the previously reported finding that the effect of increasing age on thoracic injury risk is not consistent across the range of driver ages [4] and the observation that the distribution of the type of thoracic injury sustained changes with age (e.g. Fig. 1 shows that very few younger drivers in the NASS-CDS sample sustained serious skeletal injury). Additional discussion of the age group boundary selection is given in the Appendix.

For each of the three injury outcomes, logistic regression was used to estimate the effect of the individual metrics while controlling for EDR or EDR-equivalent delta-V. As the effect of driver age on soft-tissue injury risk was not significant within each age group, it was not included in the final models estimating the risk of injury outcomes A or B. However, driver age was included as a covariate for the models estimating the risk of injury outcome C (serious skeletal injury for drivers aged 50 years or older). All regression models were calculated using the “survey” package [17] in R to account for sampling weights. As described in Brumbelow’s 2019 study [4], sampling weights were adjusted downward for cases that were mis-stratified or where a driver was hospitalised without a police-reported incapacitating injury. Adjusting these weights prevents the injury sample from becoming dominated by a relatively small number of cases that did not meet the design intentions of NASS-CDS. In the current study, cases meeting the criteria for weight adjustment accounted for 12% of those in the overall sample.

Multiple Imputation

As a good moderate overlap rating was required for case inclusion, every vehicle had a matching moderate overlap test, but a few did not have corresponding full width test data. In addition, some data were missing for various reasons even when a test was conducted. The IIHS verification program discussed in the Introduction was one common reason for missing moderate overlap test data. While all submissions include peak sternum deflection and HIC15, the other injury metrics analysed in the current study are shared only when requested by

IIHS at the time of rating assignment. Manufacturers were requested to submit missing data for the current research, but some were unable to do so.

Multiple imputation was used to enable comparisons of the predictive ability of all the dummy metrics for the same cases even when some data were missing. The multiple imputation process fills in missing data using a combination of predictions modeled using non-missing variables and a random component. By imputing the data multiple times and comparing results, standard errors can be calculated [18].

In addition to accounting for completely missing test data, the multiple imputation process also was utilised to investigate the possibility that injury metrics captured prior to certain IIHS and NCAP procedural changes could have a different relationship to real-world injury outcomes. Secondary models were constructed based on limited test datasets in which replacement values were imputed for any dummy measure either submitted by vehicle manufacturers as part of the IIHS verification program or recorded after the NCAP rating procedure change in 2011. Additional details of these changes are given in the Appendix. Table I shows the number of NASS-CDS cases available for analysis for each driver age group as well as the number of cases for which each metric was imputed for the full and limited sets of test data. The "mice" [19] and "mitools" [20] packages in R were used to perform the multiple imputation, with each model based on 40 separate imputations of the missing data. The number of imputations aligns with the recommendation of Bodner based on the proportion of missing data [21].

TABLE I

MISSING VALUES FOR TEST METRICS REPLACED WITH MULTIPLE IMPUTATION

	Age group	18 to 55-year-olds				50+ year-olds			
	NASS-CDS cases	321				133			
	Test data	All		Limited		All		Limited	
		Missing	Percent	Missing	Percent	Missing	Percent	Missing	Percent
NCAP full width	Sternum defl.	12	4%	78	24%	7	5%	42	32%
	Sternum defl. rate	12	4%	78	24%	7	5%	42	32%
	Thoracic accel.	9	3%	75	23%	7	5%	42	32%
	Defl. / accel. ratio	12	4%	78	24%	7	5%	42	32%
	HIC15	9	3%	75	23%	7	5%	42	32%
	Shoulder belt force	44	14%	100	31%	19	14%	50	38%
IIHS moderate overlap	Sternum defl.	1	0%	80	25%	0	0%	36	27%
	Sternum defl. rate	54	17%	79	25%	29	22%	36	27%
	Thoracic accel.	54	17%	79	25%	29	22%	36	27%
	Defl. / accel. ratio	55	17%	80	25%	29	22%	36	27%
	HIC15	0	0%	79	25%	0	0%	36	27%
Bumper-to-firewall dist.		0	0%	0	0%	0	0%	0	0%

Note: "Limited" test datasets exclude full-width data from 2011 and later NCAP tests and moderate overlap data submitted by manufacturers.

Model Comparison

To compare the predictive ability of different injury metrics, the area under the receiver operator characteristic curve (AUROC) was calculated for each model [22]. The AUROC for a model can be interpreted as the probability that the estimated risk for a randomly chosen case with injury will be greater than that for a randomly chosen case without injury. AUROC frequently has been used to assess the ability of regression models to predict binary outcomes, including thoracic injury outcomes from vehicle restraint systems [7]. An AUROC of 1 would indicate that the model's risk estimate for every case with an injury was greater than the risk estimated for every case without an injury, while a value of 0.5 would indicate that the model had no predictive value. The AUROC of each model using a dummy or vehicle metric as a predictor was compared with the AUROC for a baseline regression model that included only delta-V (outcomes A and B) or delta-V and driver age (outcome C) as predictors. The injury classification ability of each model relative to the baseline model was calculated as:

$$\left(AUROC_{model} - AUROC_{base}\right) \Big/ \left(1 - AUROC_{base}\right) \tag{1}$$

The greater the AUROC of the base model, the lower the possibility that a model using a dummy or vehicle metric will be able to better classify the real-world injury outcomes. For example, if the minimum delta-V of all 18- to 55-year-olds with an AIS≥3 thoracic injury was greater than the maximum delta-V of drivers without an injury, the AUROC of the baseline model would be 1 and models including additional covariates could not improve the injury classification. Each AUROC was calculated using the case weights, and those for models based on multiply imputed data were averaged across the 40 separate imputations to produce a single value.

III. RESULTS

Regression model odds ratios and p-values are presented in Table II, with relative model injury classification ability (Eq. 1) in Table III. The effect of each metric was estimated on the risk of three different injury outcomes: any type of AIS≥3 thoracic injury for drivers aged 18–55 years (outcome A), an AIS≥3 soft-tissue injury for drivers aged 50 years or older (outcome B), or an AIS≥3 skeletal injury for drivers aged 50 years or older (outcome C). No single metric had statistically significant effects on the risk of all three injury outcomes at the $p = 0.05$ level. HIC15 in the full width test was the only metric with statistically significant effects at the $p = 0.1$ level for all three outcomes, but the effects were directionally opposite for the two different age groups. Increasing NCAP HIC15 was estimated to reduce the overall risk of outcome A while increasing the risk of outcomes B and C. The model including HIC15 did not classify injury outcome B more accurately than the baseline model using delta-V alone.

Two metrics had estimated effects on the risk of outcomes B and C that were statistically significant at the $p = 0.1$ level. Increases in sternum deflection rate in the full width test and the ratio of sternum deflection to thoracic acceleration in the moderate overlap test were both estimated to increase the risk of soft- and hard-tissue injury for drivers aged 50 years and older, and models including these terms improved injury classification compared with the baseline models. The significance of both these effects was somewhat stronger when excluding post-2011 NCAP test data and moderate overlap data submitted by manufacturers. In general, across all metrics, this more restricted set of test data was associated with stronger estimated effects on injury outcomes B and C and greater improvements in injury classification.

In addition to reduced NCAP HIC15, significant injury risk increases for drivers in the younger age group were associated with increased shoulder belt force ($p = 0.01$) and reduced bumper-to-firewall distance ($p = 0.003$). While the effects of these metrics were not statistically significant for older drivers, the direction was consistent for drivers in both age groups, and injury classification was improved for one (shoulder belt) or both (bumper-to-firewall distance) older driver outcomes. Bumper-to-firewall distance was the only metric that improved injury classification for all three outcomes with a directionally consistent effect estimate. Like HIC15, most other metrics had opposing estimated effects for the two driver age groups. While often not statistically significant, increases in full-width sternum deflection, deflection rate, deflection-to-acceleration ratio, and moderate-overlap deflection-to-acceleration ratio were all estimated to increase thoracic injury risk for older drivers while reducing the risk for younger drivers. Increasing moderate overlap HIC15 was estimated to reduce injury risk for older drivers, especially of skeletal injury, but to increase risk for younger drivers.

Delta-V was a significant predictor of injury in all models at $p \leq 0.001$. The median odds ratios associated with a 1 km/h increase in delta-V were 1.12 for outcome A, 1.15 for outcome B, and 1.10 for outcome C. Age was often a significant predictor of outcome C at $p = 0.05$, with a median p-value of 0.055 and a median odds ratio of 1.05 associated with a one-year age increase. The results presented for sternum deflection in Tables II and III represent estimates using the values recorded directly from the dummy. All models using sternum deflection values adjusted for shoulder belt placement produced weaker estimated effects with larger standard errors than those produced by the unadjusted deflection.

IV. DISCUSSION

Hybrid III Risk Prediction

The goal of the current study was to evaluate the utility of the Hybrid III dummy in assessing real-world thoracic injury risk for drivers restrained by a seat belt and airbag. However, when considering the overall population of drivers and injury outcomes, the most consistent metrics were the two not recorded with the dummy. Vehicle bumper-to-firewall distance was the only metric that had a directionally consistent effect and improved the injury classification ability of all baseline models when added as a predictor. Since baseline models already controlled for crash delta-V, the risk reductions associated with longer available front crush distance suggest that the vehicle acceleration pulse is an important injury risk factor in addition to the overall delta-V. Shoulder belt force also had a directionally consistent effect across age groups and injury outcomes, and it improved on the baseline models' injury classification ability for outcomes A and B. As shoulder belt force had one of the highest rates of missing data for older drivers (Table I), it is possible that its true effect on injury outcome may be even stronger.

TABLE II

ODDS RATIOS AND P-VALUES FOR LOGISTIC REGRESSION MODELS ESTIMATING RISK OF THORACIC INJURY

	Thoracic injury outcome:		A: AIS≥3 18 to 55-year-olds				B: AIS≥3 soft tissue 50+ year-olds				C: AIS≥3 hard tissue 50+ year-olds			
		Test data:	All		Limited		All		Limited		All		Limited	
	Metric	Increase	OR	p-val.	OR	p-val.	OR	p-val.	OR	p-val.	OR	p-val.	OR	p-val.
NCAP full width	Sternum defl.	7 mm	0.86	0.60	0.66	0.41	1.92	0.33	2.62	0.14	2.47	0.32	3.71	0.05
	Sternum defl. rate	0.5 m/s	0.98	0.95	0.75	0.55	4.80	0.08	5.94	0.08	4.54	0.12	7.40	0.03
	Thoracic accel.	7 g	1.13	0.84	1.09	0.89	2.84	0.14	3.16	0.11	1.52	0.57	1.89	0.31
	Defl. / accel. ratio	0.2 mm/g	0.77	0.60	0.56	0.46	1.49	0.64	2.32	0.41	2.25	0.48	4.03	0.20
	HIC15	120	0.08	0.01	0.10	0.05	2.59	0.02	2.04	0.09	1.59	0.37	2.14	0.08
	Shoulder belt force	1100 N	2.51	0.01	2.36	0.06	1.81	0.34	2.31	0.12	1.15	0.84	1.46	0.55
IIHS moderate overlap	Sternum defl.	7 mm	1.69	0.37	1.22	0.81	1.81	0.24	2.97	0.08	1.67	0.48	2.81	0.19
	Sternum defl. rate	0.5 m/s	1.49	0.44	1.43	0.55	1.50	0.45	1.39	0.46	1.55	0.17	1.48	0.30
	Thoracic accel.	7 g	1.49	0.50	1.16	0.80	0.93	0.90	1.14	0.82	0.47	0.05	0.58	0.26
	Defl. / accel. ratio	0.2 mm/g	0.91	0.91	1.01	0.99	2.16	0.18	3.14	0.07	2.91	0.08	3.29	0.07
	HIC15	120	1.58	0.14	1.51	0.28	0.91	0.85	0.82	0.73	0.36	0.13	0.39	0.21
	Bumper-to-firewall distance	20 cm	0.23	0.003	N/A		0.56	0.32	N/A		0.48	0.25	N/A	

p-values 0.10 0.05 0.01 0.001

Note: Odds ratios represent an increase in each metric approximately equal to its interquartile range over all crash tests. For "limited" test datasets, multiple imputation replaced full width data from 2011 and later NCAP tests and moderate overlap data submitted by manufacturers. For consistency with other metrics, increasing sternum deflection and deflection rate are inverted (compression positive).

Hybrid III sternum deflection is the primary metric used to assess thoracic injury risk in front crashes. While peak deflection is significantly influenced by shoulder belt force [9], its ability to predict thoracic injury outcomes in this set of real-world crashes was more limited. For 18- to 55-year-old drivers, increasing sternum deflection in the full width test had a weak association with reduced thoracic injury risk (outcome A), while the model including moderate overlap deflection and delta-V as predictors of injury had lower injury classification ability than one based on delta-V alone. For drivers aged 50 years or older, increasing sternum deflections in both tests were associated with increased risk of soft-tissue (outcome B) and hard-tissue (outcome C) injury, but the significance and classification ability of these effects varied. The effect of NCAP deflection on the risk of skeletal injury was the only effect with a p-value ≤ 0.05, and then only when based on tests conducted before 2011. Similarly, IIHS deflection had the strongest effect on soft-tissue injury outcome (p = 0.08), but only when excluding results submitted by manufacturers as part of the test verification program. (Further discussion of the test changes and age-related directionally opposite effects is found in the Appendix and “Differences by Age” section, respectively.)

TABLE III

MODEL INJURY CLASSIFICATION ABILITY RELATIVE TO THE BASE MODEL FOR EACH INJURY OUTCOME

	Thoracic injury outcome:	A: AIS ≥ 3 18-55 year-olds		B: AIS ≥ 3 soft tissue 50+ year-olds		C: AIS ≥ 3 hard tissue 50+ year-olds	
	Base model AUROC:	0.947		0.867		0.816	
	Test data:	All	Limited	All	Limited	All	Limited
NCAP full width	Sternum defl.	7%	15%	32%	31%	1%	35%
	Sternum defl. rate	1%	8%	39%	33%	14%	28%
	Thoracic accel.	-2%	-1%	0%	4%	-2%	-1%
	Defl / accel ratio	4%	10%	25%	29%	1%	26%
	HIC15	15%	17%	0%	-4%	-3%	12%
	Shoulder belt	14%	13%	19%	18%	2%	-5%
IIHS moderate overlap	Sternum defl.	-6%	-4%	23%	37%	-7%	4%
	Sternum defl. rate	-9%	-8%	11%	10%	3%	-1%
	Thoracic accel.	-9%	0%	-1%	-1%	3%	-6%
	Defl / accel ratio	-2%	2%	27%	29%	15%	21%
	HIC15	-11%	-15%	7%	10%	13%	8%
Bumper-to-firewall distance		13%	N/A	14%	N/A	20%	N/A

Note: Percentages reflect the extent of the base model’s misclassification (1 – AUROC) that is removed by models using each test metric (Eq. 1). The base models for outcomes A and B include delta-V as the only predictor. The base model for outcome C includes delta-V and driver age as predictors.

Notwithstanding these limitations, certain models based on sternum deflection produced some of the largest improvements in injury classification relative to baseline models. In addition, the estimated injury risks for median test values were fairly consistent for the older group of drivers. For a 30 mm sternum deflection in a crash with a 65 km/h delta-V, the models predicted AIS≥3 soft-tissue injury risks of 22–35% for drivers aged 50 years and older and skeletal injury risks of 17–23% for a 65-year-old, with the lower estimates associated with the moderate overlap test deflections. These predictions can be compared with those obtained from previously published risk curves for overall AIS≥3 thoracic injury for the same 30 mm peak deflection. The ranges predicted by the models in the current study are higher than the age-independent predictions of 10% for belt loading [23] and <0.1% for distributed loading [24], but similar to the age-dependent 22% for belt loading of a 65-year-old [25].

As shown in Fig. 2, the correlation between NCAP sternum deflection and sternum deflection rate was stronger than the correlation between any other independent pair of metrics in the study. Yet the variation in NCAP deflection rate was slightly better at predicting injury outcomes B and C. An investigation of the timing of the peak rates indicated an approximately equal split between those occurring during belt pretensioning and those occurring during dummy forward excursion. Separate analyses of deflection rates calculated only during pretensioning and those calculated during excursion indicated that neither had an effect as strong as the overall maximum deflection rate. The median NCAP deflection rate was 1.3 m/s and the maximum was 2.1 m/s, both of which are correlated to less than 1% risk of AIS≥3 heart or lung injuries for all ages, according to the risk curve published by Mertz *et al*. based on porcine airbag deployment tests [24]. In contrast, the regression results from the current study using the full NCAP dataset produce estimated AIS≥3 soft-tissue injury risks of 25% and 81%, respectively, for the median and maximum deflection rates, for a crash with a 65 km/h delta-V. While these risks are based on effect estimates with borderline statistical significance, the ability of the deflection rate models to improve real-world injury classification still suggests that the published risk curve [24] underestimates serious thoracic injury risk for older drivers who experience combined seat belt and airbag loading.

The moderate overlap test metric with the best ability to predict outcomes B and C was the ratio of sternum deflection to thoracic acceleration. The utility of this metric may reflect the limitation associated with the deflection measurement capability of Hybrid III. Vehicle restraint systems have changed substantially since the dummy was developed, and the single-point deflection design is incapable of distinguishing between the localised loading from a seat belt and the more distributed loading from a front airbag. While all field and laboratory crashes in the current study involved belted drivers with front airbag deployments, cadaver results would suggest that differences in the degree of load sharing between the belt and airbag still contribute to real-world risk [5]. While not ideal, the ratio between deflection and acceleration could be one way of measuring this load sharing, with higher values indicating a greater localisation of restraint forces on the sternum from the belt restraint. Similarly, the estimated reduction in the risk of outcome C associated with increasing thoracic acceleration may indicate this issue. As IIHS thoracic acceleration has a much stronger correlation to HIC than to sternum deflection (Fig. 2), greater thoracic acceleration in the test could reflect greater airbag loading (the main contributor to HIC in this configuration).

In addition to the limited single-point measurement, there are other potential factors that could weaken the relationship between Hybrid III sternum deflection and real-world injury outcomes. First, it is possible that in this limited dataset, other occupant and crash factors predominate. A major component of the current research effort was establishing well-controlled crash severity estimates based on photographic review of damaged structures and delta-Vs directly obtained from, or adjusted for, EDR output. Limiting the analysis to good-rated vehicles further removes variability due to vehicle structure differences. However, the downside of such controls is the reduced number of cases, increasing the likelihood that other injury factors could be correlated with the variables of interest. But in another sense, the controls are still wider than ideal as they allow a range of delta-V values that have a high degree of injury classification ability on their own. This especially was true for drivers in the 18 to 55-year-old age group, where the baseline delta-V model had an AUROC of 0.947 (Table III), leaving very little room for any single test parameter to improve injury outcome classification. In addition, real-world variation in driver seat position, preimpact movement, restraint position and restraint misuse all likely influence the relevance of individual test parameters, but they could not be controlled in this study.

Secondly, it is possible that the sternum deflection measurement itself is subject to large variability. The sensitivity to shoulder belt placement [9–11] could help explain the differences between NCAP tests pre- and post-2011 and the differences between moderate overlap data measured by IIHS and submitted by vehicle manufacturers. However, adjusting deflections for belt placement [9] only reduced the models' predictive ability in the current study, suggesting that additional factors may be important. Others [26] have highlighted the fact that the existing chest deflection calibration procedure effectively controls rib cage stiffness only at deflection values (64–72 mm) that are never reached in consumer information tests. The maximum Hybrid III sternum deflection of any vehicle in the current study was 43 mm, with median values of 27 mm and 31 mm in the NCAP and IIHS tests, respectively. The reproducibility of the Hybrid III design at values less than half of the calibration corridor is unknown.

Differences by Age

Several results indicate that opposing thoracic protection requirements exist between the two driver age groups in the current study. As evident in Fig. 1, thoracic injuries tend to occur in more severe crashes for younger drivers than for older drivers. At higher delta-Vs, higher restraint forces have the advantage of reducing the likelihood of contact with intruding vehicle components. Prior research has established that the relationship between restraint forces and injury risk differs by age [27], even without vehicle intrusion. In combination, these factors could explain the opposing injury odds ratios observed for several metrics in the current study. Most of these were associated with full width test metrics, possibly because the higher deceleration pulse of that test better represents those of the higher delta-V impacts experienced by drivers in the younger age group. However, it should be noted that opposing effects were not observed for shoulder belt force. As this metric most directly measures the degree of restraint system load sharing, it indicates that some restraint system changes could benefit drivers of all ages, but that Hybrid III may not be able to measure these changes.

V. CONCLUSIONS

Vehicle bumper-to-firewall distance and shoulder belt force were better able to predict thoracic injury outcomes for all drivers restrained by a seat belt and airbag than measurements taken with the Hybrid III dummy. Some dummy metrics were able to improve predictions for drivers aged 50 years and older. Sternum deflection and deflection rate measured in the full width test improved prediction of soft- and hard-tissue injuries. Moderate overlap test measures of sternum deflection improved prediction of soft-tissue injuries, while the ratio of deflection to thoracic acceleration improved prediction of both injury types. For all of these metrics, there was evidence that data from NCAP tests conducted since 2011 and moderate overlap tests conducted by manufacturers as part of IIHS rating verification were less predictive of real-world injury outcomes.

VI. ACKNOWLEDGEMENTS

IIHS would like to thank Ford, Honda, Mazda, Mercedes-Benz, Mitsubishi, Nissan and Volkswagen for providing supplementary verification test data.

VII. REFERENCES

[1] National Highway Traffic Safety Administration. Fatality Analysis Reporting System, 2019.

[2] Farmer, C. M. (2005) Relationships of frontal offset crash test results to real-world driver fatality rates. *Traffic Injury Prevention*, **6**(1): pp.31–37.

[3] Kahane, C. J., Hackney, J. R., Berkowitz, A. M. (1994) Correlation of vehicle performance in the New Car Assessment Program with fatality risk in actual head-on collisions. Paper No. 94-S8-O-11. *Proceedings of the 14th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, pp.1388–1404.

[4] Brumbelow, M. L. (2019) Front crash injury risks for restrained drivers in good-rated vehicles by age, impact configuration, and EDR-based delta V. *Proceedings of IRCOBI Conference*, 2019, Florence, Italy.

[5] Kent, R., Bolton, J., *et al*. (2001) Restrained Hybrid III dummy-based criteria for thoracic hard-tissue injury prediction. *Proceedings of IRCOBI Conference*, 2001, Isle of Man, United Kingdom.

[6] Prasad, P. (1999) Biomechanical basis for injury criteria used in crashworthiness regulations. *Proceedings of IRCOBI Conference*, 1999, Sitges, Spain.

[7] Kent, R., Patrie, J., Benson, N. (2003) The Hybrid III dummy as a discriminator of injurious and non-injurious restraint loading. *Proceedings of the 47th Annual Conference of the Association for the Advancement of Automotive Medicine*, pp.51–75.

[8] Brumbelow, M. L., Farmer, C. F. (2013) Real-world injury patterns associated with Hybrid III sternal deflections in frontal crash tests. *Traffic Injury Prevention*, **14**(9): pp.807–815.

[9] Brumbelow, M. L. (2020) Adjusting for the effect of shoulder belt placement on Hybrid III sternum deflection. *Proceedings of IRCOBI Conference*, 2020, in press.

[10] Eggers, A., Eickhoff, B., Dobberstein, J., Zellmer, H., Adolph, T. (2014) Effects of variations in belt geometry, double pretensioning and adaptive load limiting on advanced chest measurements of THOR and Hybrid III. *Proceedings of IRCOBI Conference*, 2014, Berlin, Germany.

[11] Digges, K., Dalmotas, D., Prasad, P., Mueller, B. (2017) The need to control belt routing for silver NCAP ratings. *Proceedings of the 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. Paper No. 17-0403.

[12] NHTSA. “Databases and software.” Internet: https://www.nhtsa.gov/research-data/databases-and-software. [Accessed March 2020].

[13] SAE. (2014) Surface vehicle recommended practice J211/1; Instrumentation for impact test, Part 1: Electronic instrumentation.

[14] National Instruments. DIAdem 2017.

[15] Ooms, J. (2019) Pdftools: Text extraction, rendering, and converting of PDF documents. R package version 2.2.

[16] R Core Team. (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing.

[17] Lumley, T. (2019) Survey: Analysis of complex survey samples. R package version 3.35-1.

[18] Rubin, D.B. (1987) Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, Inc., New York, USA.

[19] van Buuren, S., Groothuis-Oudshoorn, K. (2011) Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, **45**(3): pp.1–67.

[20] Lumley, T. (2019) Mitools: Tools for multiple imputation of missing data. R package version 2.4.

[21] Bodner, T. E. (2008) What improves with increased missing data imputations? *Structural Equation Modeling*, **15**(4): pp.651–675.

[22] Metz, C.E. (1978) Basic principles of ROC analysis. *Seminars in Nuclear Medicine,* **8**(4):pp.283–298.

[23] Mertz, H. J., Horsch, J. D., Horn, G., Lowne, R. W. (1991) Hybrid III sternal deflection associated with thoracic injury severities of occupants restrained with force-limiting shoulder belts. SAE Technical Paper 910812.

[24] Mertz, H. J., Prasad, P., Irwin, A. L. (1997) Injury risk curves for children and adults in frontal and rear collisions. SAE Technical Paper 973318.

[25] Laituri, T. R., Prasad, P., Sullivan, K., Frankstein, M., Thomas, R. S. (2005) Derivation and evaluation of a provisional, age-dependent, AIS3+ thoracic injury risk curve for belted adults in frontal impacts. SAE Technical Paper 2005-01-0297.

[26] Honda Motor Company. (2007) Comment in response to NHTSA's request for comments on NCAP; suggested approaches for future enhancement. Docket document No. NHTSA-2006-26555-0076.

[27] Kent, R., Patrie, J., Poteau, F., Mutsuoka, F., Mullen, C. (2003) Development of an age-dependent thoracic injury criterion for frontal impact restraint loading. *Proceedings of the 18th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, Paper No. 72.

[28] IIHS. "Verification." Internet: https://www.iihs.org/ratings/about-our-tests#verification. [Accessed March 2020].

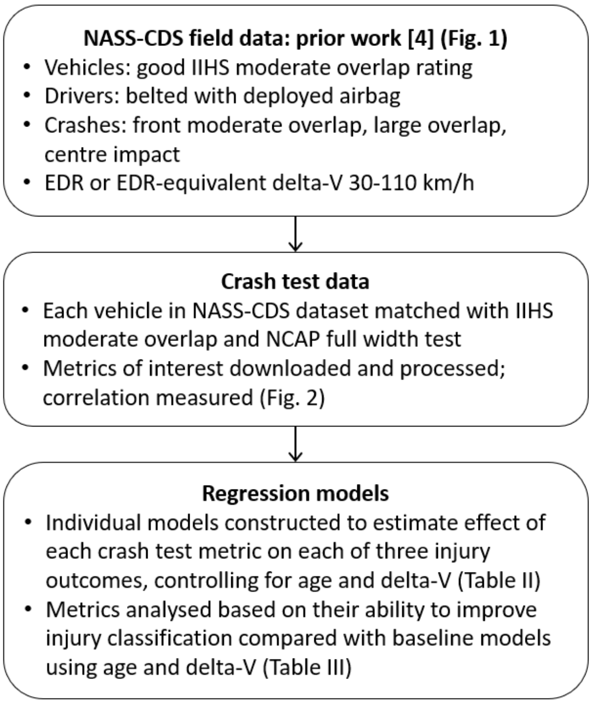
[29] Meyerson, S. L., Zuby, D. S., Lund, A. K. (1998) Repeatability of frontal offset crash tests. *Proceedings of the 16th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. Paper No. 98-S11-O-02.

[30] Mueller, B. C., Sherwood, C. P., Nolan, J. P., Zuby, D. S. (2013) Repeatability of IIHS small overlap frontal crash tests. *Proceedings of the 23rd International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, Paper No. 13-0079.

VIII. APPENDIX

Overall Methodology

A flow chart summarizing the main components of the study method is shown in Fig. A1.



NASS-CDS field data: prior work [4] (Fig. 1)

- Vehicles: good IIHS moderate overlap rating
- Drivers: belted with deployed airbag
- Crashes: front moderate overlap, large overlap, centre impact
- EDR or EDR-equivalent delta-V 30-110 km/h

Crash test data

- Each vehicle in NASS-CDS dataset matched with IIHS moderate overlap and NCAP full width test
- Metrics of interest downloaded and processed; correlation measured (Fig. 2)

Regression models

- Individual models constructed to estimate effect of each crash test metric on each of three injury outcomes, controlling for age and delta-V (Table II)
- Metrics analysed based on their ability to improve injury classification compared with baseline models using age and delta-V (Table III)

Fig. A1. Major components of study method

Age Group Boundary Selection

In order to evaluate the effect of crash test metrics on real-world crash outcomes, driver age had to be adequately controlled for two reasons: avoiding potential bias from different driver demographics among the study vehicles and identifying metrics that might be relevant to the unique risks faced by one portion of the study population. At a basic level, Fig. 1 suggests that increasing age corresponds to a change in both the degree as well as the type of thoracic injury risk. Older drivers generally sustained serious injury at lower delta-Vs and had a higher proportion of injuries involving hard tissue. To account for the possibility that these differences could obscure meaningful effects of crash test metrics, three separate real-world thoracic injury outcomes were considered: the risk of any serious (AIS≥3) thoracic injury for drivers aged 18–55 years, the risk of a serious soft-tissue injury for drivers older than 50 years, and the risk of a serious hard-tissue (skeletal) injury for drivers older than 50 years. The selection of these groups is described in detail below.

As hard-tissue injuries were almost exclusively experienced by drivers older than 50 years, models focused on this injury outcome did not include younger drivers. Including younger drivers in these analyses would have produced a single estimated age effect for the entire population with less relevance to hard-tissue injury, possibly obscuring the effects of crash test metrics.

A different approach was required for soft-tissue injuries, given their occurrence across the range of driver ages studied. A series of logistic regression models for different age groups was used to estimate the effect of driver age on serious soft-tissue injury while controlling for delta-V. The results of these models are shown in Table A1. The age groups were constructed using both the minimum and maximum driver ages as boundaries (18 and 97, respectively) while increasing the age range in the opposite direction. When considering the widest driver age ranges (bottom rows in Table A1), the models indicated that there were statistically significant effects of age on the risk of serious soft-tissue injury. However, these effects appeared to be driven mainly by the contrast between the very oldest drivers (over 70 years) and those younger than 50 years. When considering drivers in smaller age ranges, the results either were based on small sample sizes (aged 80–97 years, n = 16) or else had p-values of at least 0.19. Based on these observations, the regression models used to evaluate the effect of crash test metrics were based on cases stratified into younger (aged 18–55 years) and older (aged 50–97 years) age groups. The main disadvantage of this approach was the reduced number of observed injury cases in each group. This was partially mitigated for the younger drivers by setting the upper age boundary at 55 years instead of 50 years and by modeling overall thoracic injury risk (though only one driver sustained a hard-tissue injury without a concurrent soft-tissue injury).

TABLE A1

RESULTS OF PRELIMINARY MODELS FOR DRIVERS ESTIMATING THE EFFECT OF DRIVER AGE ON THE RISK OF DRIVERS IN SPECIFIC AGE GROUPS SUSTAINING A THORACIC AIS≥3 SOFT-TISSUE INJURY WHILE CONTROLLING FOR DELTA-V

Age range	Raw cases	Age OR	p-value	Age range	Raw cases	Age OR	p-value
18–40	218	1.04	0.67	80–97	16	5.50	<0.001
18–50	292	1.00	0.98	70–97	39	0.84	0.23
18–60	346	1.04	0.26	60–97	79	1.05	0.38
18–70	384	1.04	0.19	50–97	133	1.03	0.30
18–80	408	1.06	0.01	40–97	211	1.07	<0.001

Note: OR=odds ratio.

Procedural Changes to Crash Test Programs

Procedural changes were made to both the IIHS moderate overlap and the NCAP full width evaluation programs during the calendar years under study. While these changes did not affect the test configuration (overlap and impact speed), it is possible that they could affect the relationship between dummy metrics and real-world injury risk. This possibility was accounted for in the current study, as discussed in the “Multiple Imputation” section in the Methods. Details of the crash test program changes are outlined below.

Beginning in 2004, IIHS has assigned an increasing number of moderate overlap ratings through a verification program in which vehicle manufacturers submit data for tests conducted in their own facilities [28]. As part of verification, IIHS periodically conducts audit testing to ensure that assigned ratings are consistent with ratings that would be obtained from its own tests. As of 2019, no moderate overlap audit test has produced an overall vehicle rating different than the rating assigned based on data submitted by manufacturers. However, as shown in Figs A2 and A3, there can still be relatively large differences in specific injury metrics when the same vehicle is tested by IIHS and the vehicle manufacturer. Many factors may contribute to these differences, but as they generally are larger than results from repeated moderate overlap [29] and small overlap [30] tests conducted at IIHS, they suggest the possibility of some variability associated with the test facility. This is consistent with the findings of an analysis of factors influencing sternum deflection in NCAP tests conducted at different labs [9].

In 2011, the NCAP rating procedure was changed to incorporate new tests as well as new metrics in existing tests. Before this time, Hybrid III sternum deflection was not used in the NCAP rating procedure. Since the NCAP procedure allows manufacturers to specify the seat belt upper anchorage location for the test, and to indirectly specify the location of the midtrack seat position through vehicle design, the shoulder belt can be aligned with a position on the dummy that will reduce the deflection at the measurement point. Research has shown that, on average, tests conducted since 2011 have had shoulder belt placements farther from the sternum potentiometer and that peak deflection measurements have been affected [9].

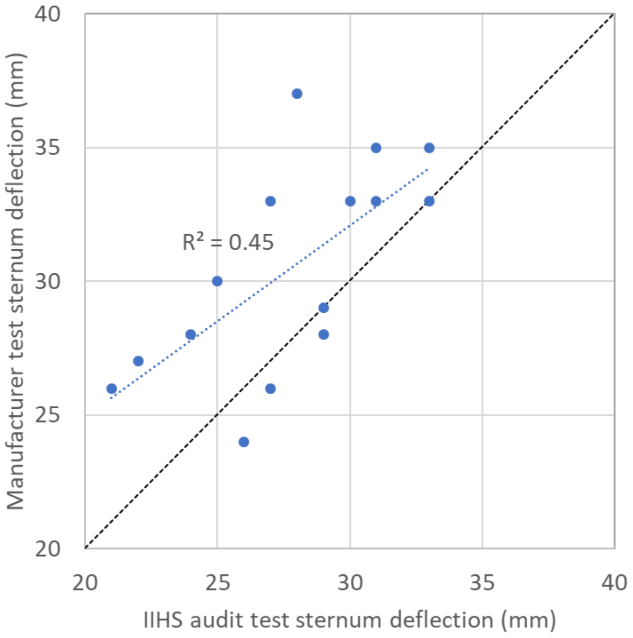


Fig. A2. Comparison of peak Hybrid III sternum deflection values submitted by manufacturers with those measured by IIHS in audit tests of the same vehicle.

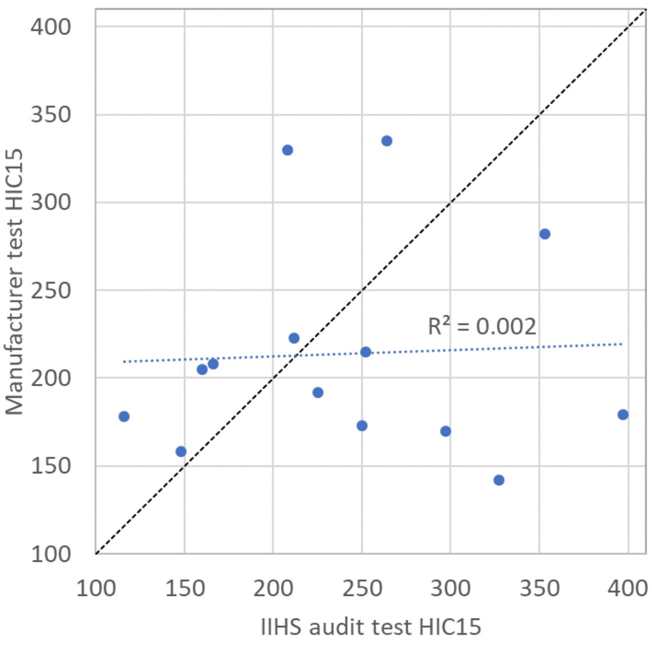


Fig. A3. Comparison of peak Hybrid III HIC15 values submitted by manufacturers with those measured by IIHS in audit tests of the same vehicle.