The following Technical Brief is supplemental to *NCHRP Research Report 1103: The Effect of Vehicle on Crash Frequency and Crash Severity* (NCHRP Project 22-49). The full report can be found by searching on the report title on the National Academies Press website (nap.nationalacademies.org).

Technical Brief

The current version of the *Highway Safety Manual* (HSM) does not accommodate vehicle mix information while estimating total crashes, crashes by type, and crashes by severity. However, recent research efforts have shown that heavy vehicle traffic and vehicle mix have a substantial impact on crash frequency and severity. Further, these studies indicate that the consideration of vehicle mix would improve predictive methods for crash frequency and severity. To examine how vehicle mix data can influence the model infrastructure in HSM, this project, titled NCHRP Project 22-49, "The Effect of Vehicle Mix on Crash Frequency and Crash Severity," has two objectives: a) develop and validate a statistically valid predictive methodology to quantify the effect of vehicle mix on crash frequency and severity for various facility types, and b) develop a spreadsheet tool for practitioners to quantify the effect of vehicle mix on safety performance across the range of highway activities including planning, design, operations, and safety management.

Toward accomplishing the objectives, the research team proposed two alternative approaches to incorporate vehicle mix effect in crash frequency and severity analysis: (a) negative binomial-ordered probit fractional split (NB-OPFS) method to predict crash frequency by severity levels, and (b) multivariate Poisson-lognormal (MVPLN) model to develop separate frequency models by crash severity level. The research team also calibrated the models for existing HSM methods to serve as a benchmark for the proposed alternative approaches. These three approaches were employed at facility levels (segment and intersection) where the facilities were selected based on the total and heavy vehicle crashes. The team collected crash and facility-level data for 7 states (4 HSIS states including California, Illinois, Minnesota, and Washington, and 3 non-HSIS states including Connecticut, Florida, and Texas). Within each state, the research team considered the facilities that are covered in the first edition of the HSM including rural two-lane two-way roadways, rural multilane highways, urban/suburban arterials, freeway segments, and intersections. Each facility is further categorized into multiple categories based on its attributes such as the number of lanes and presence of a median (for segments), and the presence of traffic control devices (for intersections). Finally, the team identified 17 segments and 7 intersections as facilities (24 in total) for developing crash prediction models. It is to be noted that this project included two new facilities (Rural arterial 3-lane and Rural arterial 5-lane, not included in the current HSM) into the analysis as these facilities exhibited a significant number of crashes across the states.

For developing the models, the team considered KABCO severity scale (where K = fatal injury crash count, A = incapacitating injury crash count, B = non-incapacitating injury crash count, C = possible injury crash count, and O = no injury/PDO crash count) as the dependent variables. In terms of independent variables, the team explored three broad categories of variables: roadway characteristics (covered in HSM such as lane width, shoulder width, median width), traffic characteristics (covered in HSM aggregate level AADT), and vehicle mix information (new variables, not covered in HSM). In terms of vehicle mix data, the team explored the availability of the vehicle mix data for each state in the project. For states with vehicle mix data available, the team used the corresponding vehicle mix variable for model estimation. On the other hand, if vehicle mix data was not available, the research team adopted the Quasi-induced exposure (QIE) technique for generating the vehicle mix data across each facility type within the state and then used the generated vehicle mix data in crash frequency and severity models for the corresponding facility. Considering the data availability, the team decided to use coarser level vehicular mix data (passenger vehicle and truck) and the single unit truck data at the finer level in the model estimation process for the project. Based on the finalized datasets, model estimation and model comparison were conducted. The team employed several predictive metrics to compare the model performance for estimation and validation datasets and selected the better model for each facility. At the end, the research team in consultation with the project panel selected a single recommended model system for each facility group. The key findings of the project work are as follows:

- For urban and rural limited access facilities except the Urban limited access 4-lane divided segment facility, the performance of the HSM models was consistently better than the new alternative approaches.
- The NB-OPFS models performed better for both urban and rural arterial facilities including Urban 2-lane undivided, Urban 3-lane, Urban 5-lane, Urban 4-lane undivided, Urban 4-lane divided, Rural 2-lane undivided, Rural 3-lane, Rural 5-lane, Rural 4-lane undivided, and Rural 4-lane divided.
- For urban intersections, the results showed that the MVPLN models provided improved performance in terms of predictive performance measures than HSM and NB-OPFS models.
- For rural intersections, the NB-OPFS models showed improved predictions relative to the other two approaches.
- Across most of the facilities, it is observed that the vehicle mix information is found to significantly influence crash frequency and severity. The vehicle mix variables found to influence crash frequency and severity include truck traffic percentage (for segments), single unit truck traffic percentage (for segments), high truck zone (for segments), major and minor road truck traffic percentage (for intersections).

To aid the practitioners in implementing the new models developed in this project, the research team developed a suite of Excel spreadsheet tools including a) 22-49 Spreadsheet Tool without Calibration, b) 22-49 Spreadsheet Tool with Calibration, and c) 22-49 Data Input and Prediction Tool. The Spreadsheet Tool without Calibration provides crash frequency and severity predictions for the user-provided data directly without considering calibration while the Spreadsheet Tool with Calibration provides the predictions considering calibration. The Data Input and Prediction Tool provides practitioners with a tool to undertake crash frequency and severity analysis at a facility resolution (i.e., a specific segment or intersection). A step-by-step guidance document showing detailed instructions on how to use the three developed tools was also included.

The research team invested significant resources to develop the model systems outlined above. However, the research is not without limitations. From an empirical perspective, this research incorporated truck traffic percentage as coarser resolution and only single unit truck traffic percentage as finer resolution vehicle mix variables as the data is not available for all states. However, it might be beneficial to develop models with detailed, finer-resolution data and compare their performance to the models developed in this study. From a methodological perspective, the current research effort focused on aggregate-level vehicular mix information for crash frequency and severity analysis. However, crashes analyzed in frequency models are aggregated from disaggregate-level crash records (i.e., crash severity counts are estimated from crash severity records). Yet the model frameworks do not account for any disaggregate level independent variables (such as weather, heavy vehicle involvement) in modeling crash frequencies and severities. Hence, in future work, it might be beneficial to develop an integrated model system that allows for the influence of observed and unobserved variables at the disaggregate resolution from the severity model to influence crash frequency modeling process.