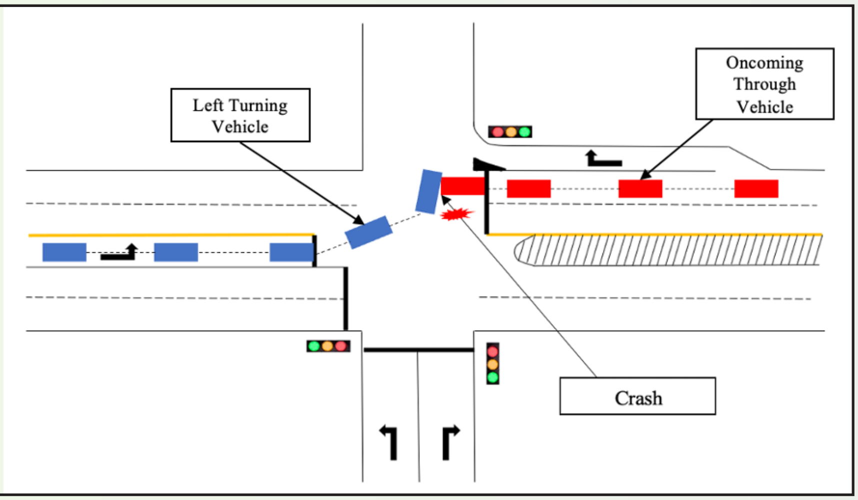
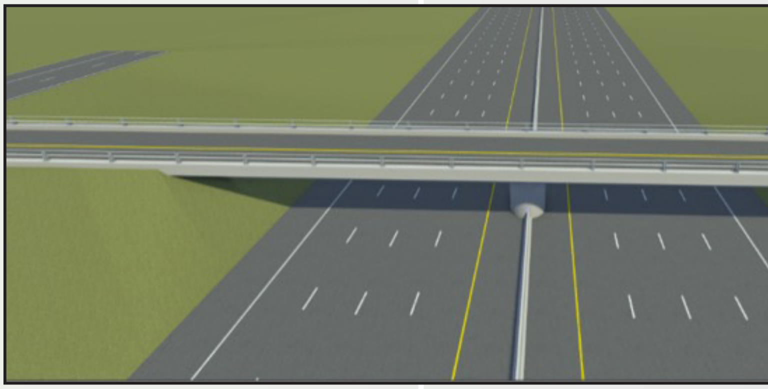
Exploratory Advanced Research Program

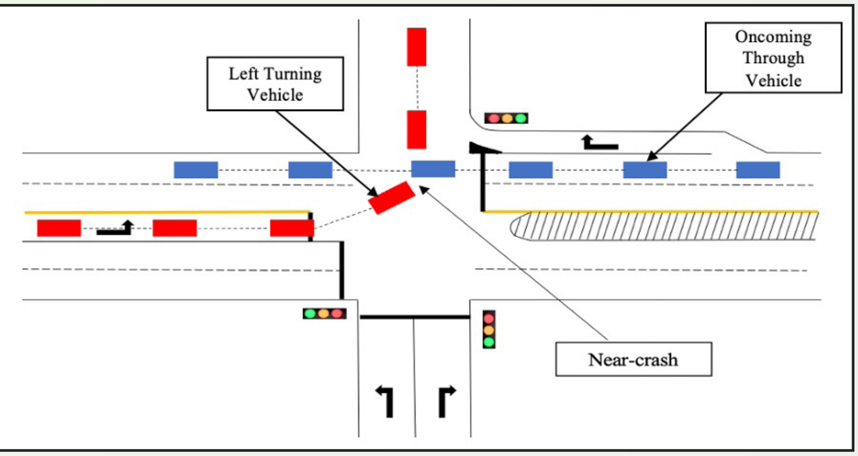
Multidisciplinary Initiative to Create and Integrate Realistic Artificial Datasets

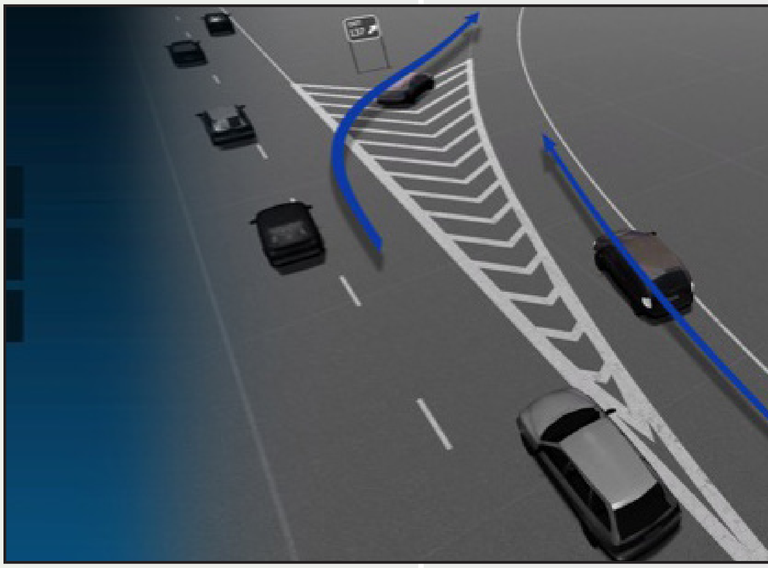
Research Summary Report

EXPLORATORY ADVANCED RESEARCH









U.S. Department of Transportation
Federal Highway Administration

Turner-Fairbank
Highway Research Center

Notice

Quality Assurance Statement

Cover Photo Credits:
All images source: FHWA.

Technical Report Documentation Page

1. Report No. FHWA-HRT-23-058	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Multidisciplinary Initiative to Create and Integrate Realistic Artificial Datasets		5. Report Date June 2023 6. Performing Organization Code
7. Author(s) P. Edara (ORCID: 0000-0003-2707-642X), C. Sun (ORCID: 0000-0002-8857-9648), H. Brown (ORCID: 0000-0003-1473-901X), P. Savolainen (ORCID: 0000-0001-5767-9104), V. Shankar (ORCID: 0000-0002-6671-2268), B. Balakrishnan (ORCID: 0000-0002-0994-0213), Y. Shang ORCID: (0000-0001-7771-4034), S. Chakraborty (ORCID: 0000-0003-2022-1735), Y. Adu-Gyamfi (ORCID: 0000-0002-1924-9792), C. Li (ORCID: 0000-0002-3237-1477), K. Aati (ORCID: 0000-0001-8834-7735), S. Lima, Y. Huang (ORCID: 0000-0002-7346-5293), A. Mussah (ORCID: 0000-0002-1084-5598), J. Hopfenblatt		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Missouri-Columbia E2509 Lafferre Hall Columbia, MO 65211 Schatz Publishing Group 11950 W. Highland Ave. Blackwell, OK 74631		10. Work Unit No. 11. Contract or Grant No. 693JJ31950023
12. Sponsoring Agency Name and Address Office of Corporate Research, Technology, and Innovation Management Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101		13. Type of Report and Period Covered Research Summary Report September 20, 2019–March 19, 2022 14. Sponsoring Agency Code HRTM-30
15. Supplementary Notes The Contracting Officer's representative was Yusuf Mohamedshah (HRSO-2; ORCID: 0000-0003-0105-5559).		

16. Abstract

Safety models traditionally focus on crash frequencies or rates and do not always reflect the underlying causes of the crashes. Even models that accurately estimate crash measures and consist of multiple causal factors may not explain all causal relationships. Yet understanding what causes crashes is crucial to developing countermeasures and ensuring safety. To expand traditional modeling practices and results, the Federal Highway Administration's Exploratory Advanced Research Program sponsored a project to develop a framework that would generate realistic artificial datasets (RADs) that mimic the known causal relationships between contributing factors and crashes. The researchers applied the framework to generate RADs for ramp terminals and speed change lane facilities at diamond interchanges. They also developed web-based software to provide easy access to the RADs, so other researchers could use them to test their models. The software contains 196 pregenerated datasets and the option to submit custom data requests. The framework is generic, so it can be used to generate RADs for other types of facilities. The researchers evaluated and compared the performance of different models using a model evaluation rubric. They also used RADs to evaluate new behavioral and roadway countermeasures by generating virtual reality simulation testbeds for crashes and near-crashes occurring at interchanges. They used safety-critical events recorded in the second Strategic Highway Research Program Naturalistic Driving Study to build the testbeds. In addition, the research team developed a graphical user interface to facilitate the testbeds for left-turn and speed change lane crashes. Virtual reality provides an engaging platform for evaluating countermeasures and educating the public about interchange crashes, which can help achieve the U.S. Department of Transportation's goal of zero roadway fatalities.

17. Key Words Crashes, realistic artificial data, machine-learning models, safety, simulation testbed, synthetic data, visualization		18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. https://www.ntis.gov	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 16	22. Price N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized.

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.755	cubic meters	m^3
		NOTE: volumes greater than 1,000 L shall be shown in m^3		
		MASS		
ounces		28.35	grams	g
pounds		0.454	kilograms	kg
short tons (2,000 lb)		0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
foot-candles		10.76	lux	lx
foot-Lamberts		3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
poundforce		4.45	newtons	N
poundforce per square inch		6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2,000 lbs)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	2.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbfin^2

*SI is the symbol for International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.(revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Crash prediction models are vital to transportation safety decisionmaking. Models provide the basis for developing countermeasures to reduce crashes. However, traditional safety models focus on crash frequencies and rates and not on the cause-and-effect relationships that lead to crashes.

The U.S. Department of Transportation (USDOT) recognizes the importance of using data-driven safety analysis (DDSA) models to improve transportation safety. USDOT's Strategic Plan 2022–2026 emphasizes the importance of using DDSA methods as part of the Safe System Approach, which includes roadway countermeasures, behavioral interventions, enforcement, vehicle safety features, and emergency medical care to prevent crashes and minimize injuries when crashes occur (USDOT 2022).

To address the need to incorporate causal relationships in crash prediction models, the Federal Highway Administration's (FHWA) Exploratory Advanced Research (EAR) Program supported a project titled "MIMIC: Multidisciplinary Initiative on Methods to Integrate and Create Artificial Realistic Data." The project aimed to develop a framework to generate realistic artificial datasets (RADs)—datasets containing computer-generated data rather than real-world information—for interchange facilities to test how well a crash prediction model reflects the underlying cause-and-effect relationships. A team of researchers from the University of Missouri, Iowa State University, and Texas Tech University applied statistical and machine-learning methods to generate RADs to mimic the relationships between contributing factors and crashes and test how well a given model performs.

Original photo: Imagery © 2020 Maxar Technologies. Map data © 2020 Google®. Modifications by FHWA (see Acknowledgments section). **Figure 1. Map. Example of a diamond interchange with two ramp terminals.**

Original photo: Imagery © 2020 Maxar Technologies. Map data © 2020 Google®. Modifications by FHWA (see Acknowledgments section). **Figure 2. Map. Real-world example of an entrance speed change lane.**

RADs can aid highway research in the following ways:

- Assess a new crash estimation method.
- Compare methods to analyze alternatives.
- Conduct human factors evaluation of behavioral and roadway countermeasures.

Advancing RADs also will enhance FHWA's efforts to encourage practitioners to apply DDSA methods to safety decisionmaking through programs such as Every Day Counts (EDC) by expanding the number of tools available for safety analysis (FHWA 2022a). EDC is a State-based model that identifies and rapidly deploys proven but underutilized innovations that make the Nation's transportation system adaptable, sustainable, equitable, and safer.

The researchers applied the RAD-generation framework to two types of crashes occurring at diamond interchanges: ramp terminal left-turn crashes (figure 1) and speed change lane (an uncontrolled terminal between a ramp and a freeway) crashes (figure 2). The framework is generic and can be used to generate RADs for work zones, innovative geometric designs, bicyclist/pedestrian lanes and crossings, and other facilities. The researchers developed web-based software to access the datasets.

The researchers developed a rubric to evaluate and compare the performance of the models. They applied the rubric using statistical and machine-learning methods to each of the developed models. The tests evaluated six criteria, including the ability of a model to explain the cause–effect relationship between an independent variable and the outcome.

In addition to establishing RADs in a traditional table format and then developing the underlying models, the research team developed a simulation testbed. They used safety-critical events recorded in the second Strategic Highway Research Program (SHRP2) Naturalistic Driving Study (NDS) to create the testbeds (Virginia Tech Transportation Institute 2020). Although not a true RAD, the simulation testbed allowed the researchers to create variables for a situation, run a virtual-reality simulation model for crashes and near-crashes at interchanges, and evaluate new behavioral and roadway countermeasures. Virtual reality provides an engaging visualization platform for examining human factors and countermeasures in crash scenarios and educating the public about interchange crashes.

Project Overview

Currently, researchers collect data and develop models to predict the likelihood of crashes or the relationship of crash occurrence to a geometric feature of the highway, such as traffic, number of lanes, or shoulder width. The researchers then test the model to determine how well it predicts crashes.

When testing the model's validity, researchers typically randomly select a part of the original data used to develop the model and perform a type of analysis, such as statistical analysis. Although this method is the best practice right now, by using the same data to develop and test the model, the results provide a very high relationship factor.

Modelers estimating different forms of statistical models can only compare them using overall goodness of fit measures, such as prediction accuracy and likelihood value. They cannot compare the models based on their ability to extract cause–effect relationships because the cause–effect relationships between independent and dependent variables are seldom known. RADs provide an effective tool for testing a model's predicted outcome because RADs are created using assumed cause–effect relationships.

The research team developed RADs in two ways:

1. Tabular form—The traditional way of looking at the dataset in a table and developing the underlying models.
2. Simulation testbed—Researchers use RADs to create the variables for their situation and run a simulation model.

The research team chose to work with urban interchanges as a representative facility type because accurate crash data is lacking for these roadway features, and they are overrepresented in real-world crash data, with more than 50 percent of fatal or injury crashes involving intersections (FHWA 2019). Specifically, they developed RADs for ramp terminal left-turn crashes and speed change lane crashes at diamond intersections.

The first step the researchers undertook was a literature review.

Literature Review

The research team reviewed the literature on developing and using synthetic, or realistic artificial, data and understanding the causes of crashes at interchanges. Studies documenting synthetic data development and use provided information on available data-generation methods. Literature on crash causation at interchanges helped provide information on key independent variables, their impact on crash frequency, and state-of-the-practice crash prediction models.

Although new to transportation, synthetic data has been used in other disciplines. The researchers reviewed studies that demonstrated the successful development, evaluation, and application of artificial datasets. For this project, the researchers developed rubrics using RADs to rank the different models based on their performance. The goal was to create a rubric grading system to rank different models based on their performance to help modelers revise and improve their models.

The researchers also reviewed the literature on ramp terminal crashes. They found that most studies are constrained by the lack of crash data based on different types of freeway interchanges, such as speed change lanes, entrance and exit ramp segments, ramp terminals (intersections), and freeway segments. As a result, several studies relied on simulations to analyze scenarios that could improve safety and traffic flow through these highway sections (Portera and Bassani 2021; Elefteriadou et al. 2005). The researchers in this study also turned to simulations to test safety countermeasures.

Once the literature review was complete, the researchers developed the data-generation framework. They developed RADs for crashes at ramp terminals and speed change lanes at diamond interchanges, which are highly prevalent on U.S. roadways. Then they applied statistical and machine-learning approaches to the RADs to model crash frequency and ascertain the cause–effect relationships. The researchers used the following three steps to develop the framework.

IDENTIFY CONTRIBUTING FACTORS AT THE SELECTED INTERCHANGE FACILITY

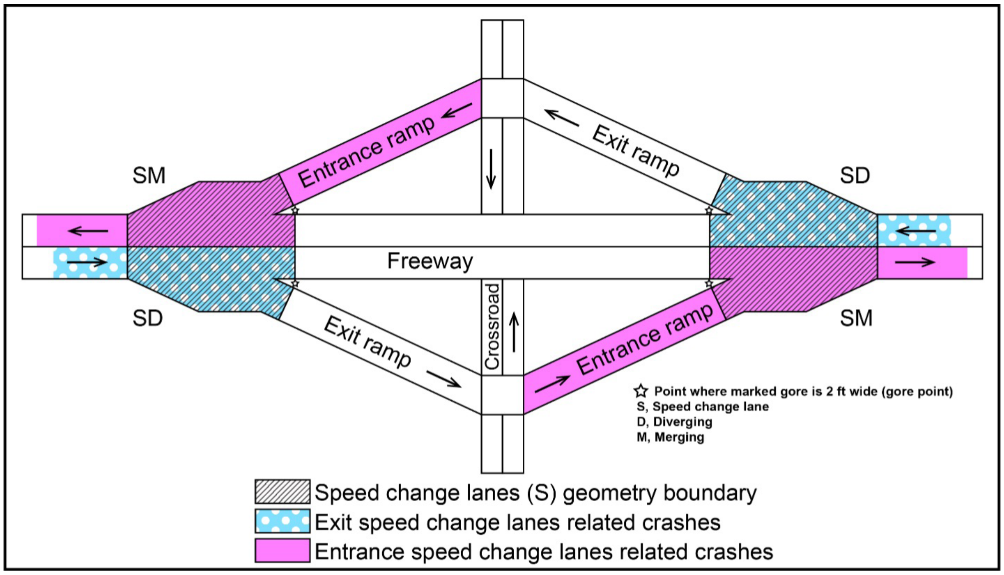
In step one, the researchers identified the contributing factors for each selected facility, including roadway, traffic, and driver characteristics. They generated sampling distributions for each factor using observed data from the States of Washington and Missouri. The Missouri data was acquired from the Missouri Department of Transportation's (DOT) Transportation Management Systems database as part of a recently completed *Highway Safety Manual* calibration project (Sun, Brown, Edara, et al. 2013). By repeatedly sampling the distributions for a given sample size (e.g., 500 sites), the researchers generated realistic artificial data for these factors. They also obtained several States' data from FHWA's Highway Safety Information System on interchange crashes for a 5-yr period. While the researchers reviewed data from California, Illinois, Maine, and Minnesota, they found the data from Washington were the most complete and recent (2013–2017) for developing the RADs.

While extracting left-turn crash data for ramp terminals was straightforward, extracting crashes for speed change lane facilities was more involved. Figure 3 shows the four speed change lane facilities at a diamond interchange, two related to the exit and two related to the entrance. Crash reports do not indicate if a crash was a speed change lane crash, so the researchers analyzed interchange schematics and the milepost where the crash occurred to determine if a crash occurred in the speed change lanes. The researchers analyzed 205 of the Washington State DOT's (WSDOT) diamond interchange drawings. The WSDOT's schematics (published online) also provided information on the location of the speed change lane, the direction of travel, and the location of crashes related to the speed change lane by direction and location (WSDOT 2016). They used a similar process to extract crash data from 75 diamond interchanges in Missouri.

The researchers located the potential crashes that occurred at the speed change lanes and used geographic information systems software and HSIS shapefile crash data to map the crash geolocations. Then they merged the potential crash data with the location data. Finally, they aggregated the crashes by location to get the number of crashes that occurred at each speed change lane site. This process distinguished the crashes that occurred on the speed change lanes from the mainline and ramp crashes.

ESTABLISH THE CAUSE–EFFECT RELATIONSHIPS

In step two, the researchers established the cause–effect relationship for each significant factor that influences crash frequency, synthesizing information from published literature, the *Highway Safety Manual*, and the Crash Modification Factor (CMF) Clearinghouse (American Association of State



Modifications by FHWA to show the gore point, speed change, diverging, and merging lanes.

Figure 3. Graphic. Components of speed change lanes at an intersection (Sun, Edara, Brown, et al. 2016).

Highway and Transportation Officials 2010; FHWA 2022b). When no reliable information was available for a particular variable, the researchers made assumptions based on observed data from Washington and Missouri.

GENERATE THE CRASH DATA

For the third and final step, the researchers combined the effect of all contributing factors on crash frequency and estimated a composite crash measure for a given site based on its roadway and traffic characteristics. In generating the composite score, they considered the individual effects of each factor and the interaction effects between two or more factors. The composite crash score was converted to realistic crash frequency (i.e., counts) using observed crash data and a hierarchical Poisson approach, with parameters optimized for each level of the hierarchy. The researchers then adjusted the generated crash data distribution parameters to match the individual sites' overall distributional shape and crash counts.

Once the crash counts were generated, the researchers categorized them as fatal, injury-causing, or property damage-only crashes. In addition to severity, they also generated characteristics about the driver (e.g., level of distraction, age, gender), vehicle, and roadway (e.g., road condition) for each crash.

APPLYING RAD TO CRASH PREDICTION MODELS

The RADs generated for left-turn and speed change lane facilities were used to test crash prediction models. The researchers used the developed datasets with state-of-the-art statistical and machine-learning approaches to model crash frequency and ascertain the cause–effect relationships. The researchers developed a rubric to evaluate the performance of models developed using RADs. The rubric had a scoring system of 0–100. Six criteria were evaluated to capture different complexities in modeling, including:

- Data descriptive analysis.
- Model selection.
- Data training and testing.
- Prediction accuracy.
- Model inference.

To allow for unbiased evaluation of the cause–effect relationships, three teams whose members were unaware of the RAD-generation procedures estimated the models. Two teams worked with statistical models, and one team developed a series of machine-learning models. The evaluation rubric was applied to each of the developed models.

The results showed the two statistical models performed similarly, while the machine-learning models outperformed the statistical models, especially in the model inference criterion. The CMFs in the machine-learning models were closer to the true CMFs, thus explaining the assumed cause–effect relationships. The machine-learning models performed better on the speed change lane RADs than on the ramp terminal dataset. One possible reason the machine-learning models performed better is their ability to capture the nonlinear relationships between crash frequency and the independent variables. Table 1 shows the evaluation of the statistical and machine-learning models developed using the RADs.

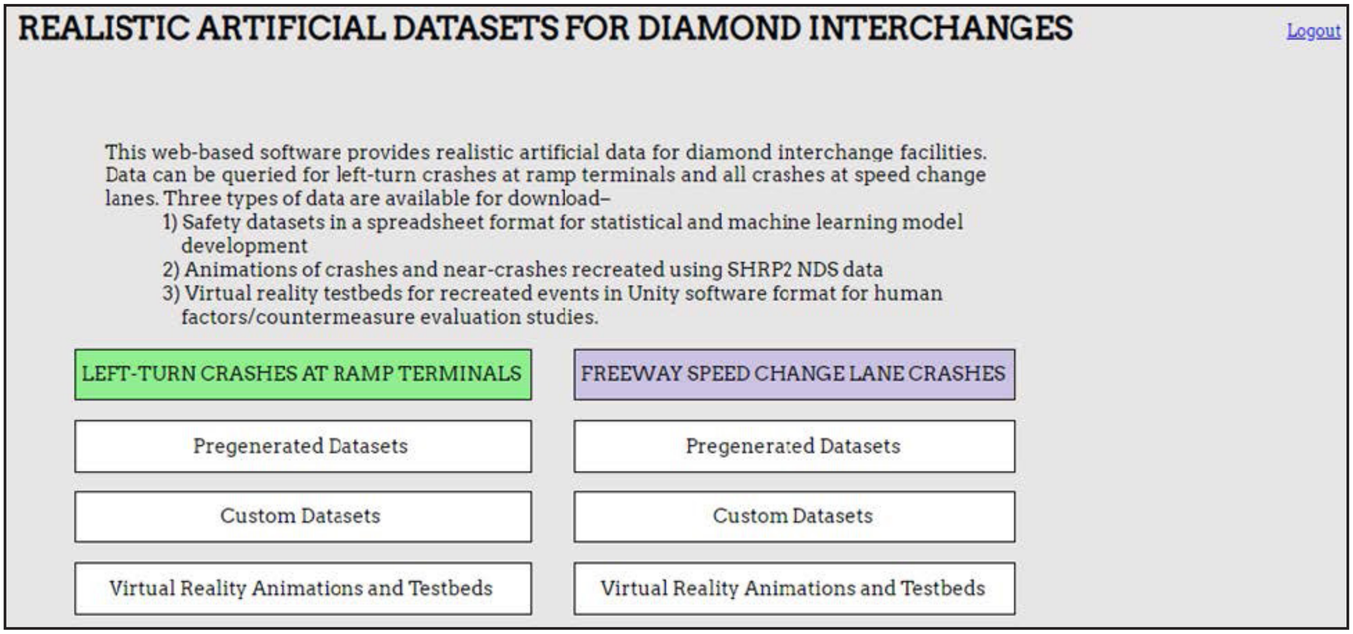
One other use of a RAD is to allow researchers to compare the performance of different models estimated using different observed data. Consider, for example, two studies developing crash frequency models for speed change lanes—Study A is using observed data from State A, and Study B is using observed data from State B. How can researchers accurately assess the CMFs generated from the two studies? How can they determine which CMFs explain the true cause–effect relationships between crash frequency and the corresponding input variables (e.g., freeway annual average daily traffic, speed change length)? RAD can help answer these questions. The research teams conducting the two studies apply their modeling approaches to the same RAD. The resulting models can be compared by performing statistical tests, such as goodness of fit, marginal effects, variable variance, etc. Alternatively, the comparison may also be made by only using the model inference criterion, that is, comparing the CMFs generated using the RAD to the known CMFs (i.e., those used to develop RAD). The modeling approach that performs the best on the RAD is more likely to explain the true cause–effect relationship in observed data. Thus, in the example, if the performance of the model developed in Study A on the RAD is better than the performance of the model developed in Study B, Study A is likely to also produce more reliable CMFs when using observed data. This type of testing was not conducted in this project due to the lack of readily available models (and CMFs) for interchange ramp terminals and speed change lanes already estimated using data from different datasets.

Table 1. Overall model performance scores of the statistical and machine-learning models.

Criteria	Maximum Score	Team 1 LT	Team 2 LT	Team 3 LT	Team 1 SCL	Team 2 SCL	Team 3 SCL
Descriptive statistics	10	10	10	10	10	10	10
Model selection	10	8	8	10	8	8	10
Training and testing data	10	10	10	10	10	10	10
Overall model performance	20	14	16	16	14	14	16
Model inference	50	30	30	35	35	35	45
Total score	100	72	74	81	77	77	91

LT = left turn; SCL = speed change lane.

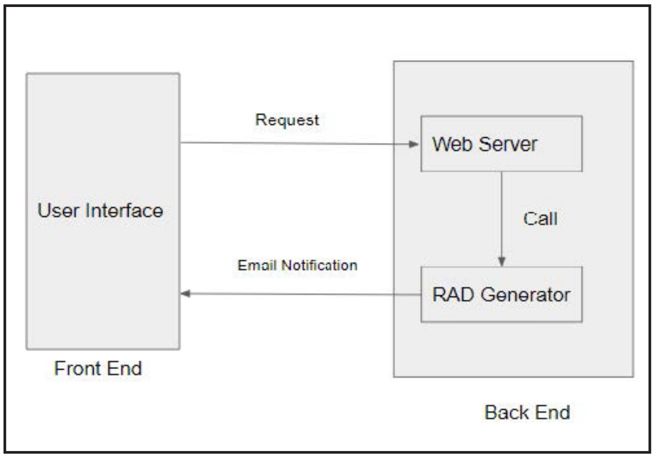
The researchers developed web-based software to provide access to the RADs and facilitate their use. The software includes a landing page (figure 4) for entering real-time user requests to the server and a RAD generator. The software consists of 196 pregenerated datasets in a spreadsheet format similar to the safety datasets provided by State DOTs. The software also allows the user to submit custom data requests.



Source: FHWA. **Figure 4. Screenshot. Landing page with main menu options in the RAD software.**

The pregenerated RADs are available for 16 combinations, based on the number of sites and years. For each combination, five different pregenerated datasets are available. To download a pregenerated RAD, users select the number of sites and years they want and then click Download. The system randomly provides one dataset from the five pregenerated datasets for the chosen combination.

Figure 5 shows the process when a user submits a custom RAD request. First, a user enters the number of sites, the number of years of the RAD, and an email address to receive the dataset once it is generated. After the RAD request is submitted, the RAD generator creates the dataset. When the request has finished processing, the user receives an email notification that the RAD is ready and can be downloaded through a provided link.



Source: FHWA.
Figure 5. Diagram. RAD software request process.

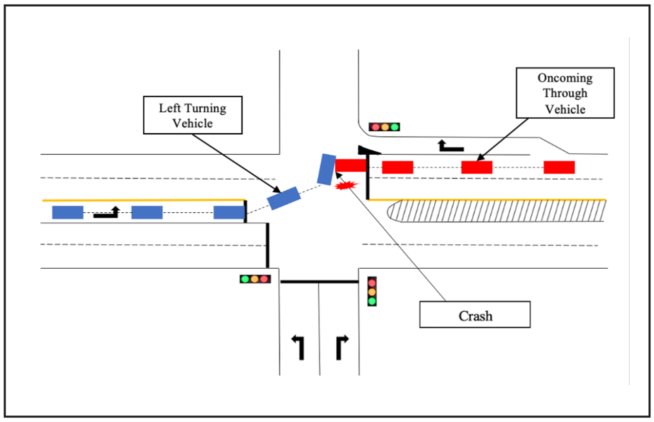
The researchers extended the use of the RADs to generate realistic simulation testbeds for crashes and near-crashes occurring at interchanges to evaluate new behavioral and roadway countermeasures. They created variables for a situation and ran the virtual-reality simulation models. To develop the testbeds, the researchers used the following four-step process:

- Obtain and analyze videos of safety-critical events. The researchers obtained de-identified data from the SHRP2 NDS database for 41,479 crash or near-crash events. By performing a series of data reductions, they identified 114 crash and near-crash events involving left-turning vehicles and 310 speed change lane events. Of the 310 speed change lane events, 179 occurred on entrance speed change lanes and 131 on exit speed change lanes. The event information included time of day, weather, vehicle status, vehicle trajectory, surrounding traffic, road conditions, signage, and pavement markings.

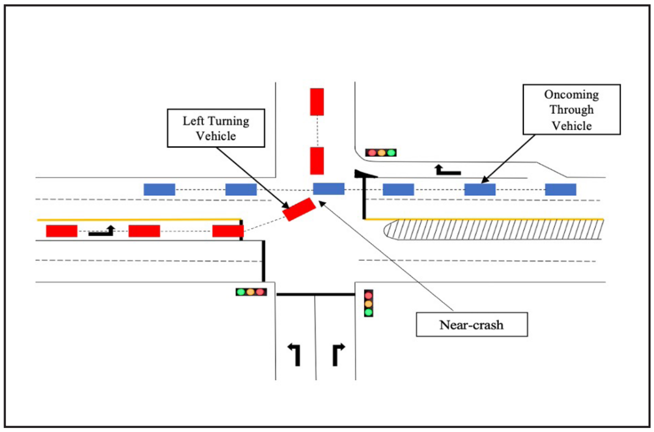
- Diagram crashes. The researchers then diagrammed crash events, drawing trajectories of vehicles involved in each crash (figure 6 and figure 7). They also generated road signs for the simulator based on signs observed in the videos.

- Create the roadway and surrounding environment. The researchers used three-dimensional (3D) modeling tools to create the roadway and surrounding environment, including shoulders, travel lanes, medians, terrain, pavement markings, and overpasses, as well as lighting and foliage (figure 8).

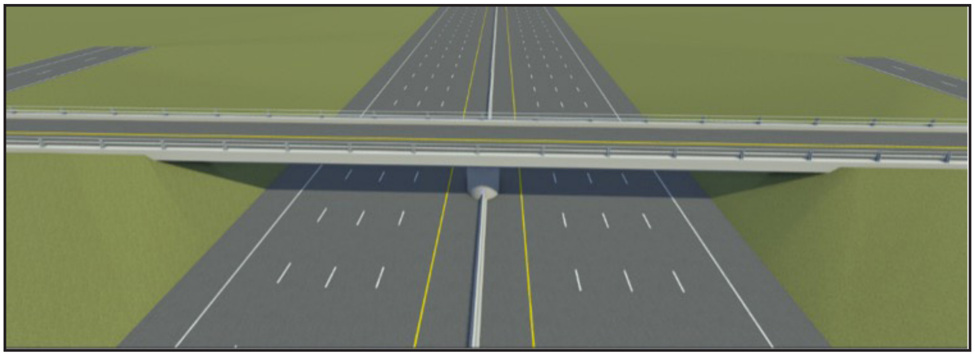
- Create the crash simulation. In the final step, the researchers reconstructed crashes by overlaying the vehicle information and trajectories on top of the roadway and environment elements.



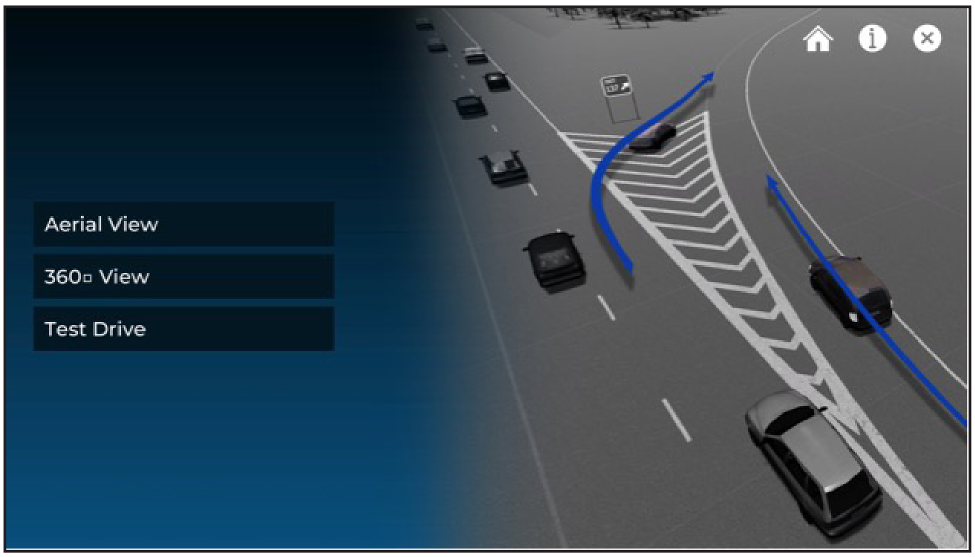
Source: FHWA.
Figure 6. Diagram. Left-turn crash event.



Source: FHWA.
Figure 7. Diagram. Left-turn near-crash event.



Source: FHWA.
Figure 8. Graphic. Example of a highway and an overpass structure.



Source: FHWA.
Figure 9. Screenshot. User menu showing three visualization options.

To make the simulator testbeds easy to use, the researchers created a graphical user interface that lets users choose from three different views (figure 9). The aerial view shows a recreated animation of a crash. The 360-degree view puts the user in the driver's seat of the subject vehicle during the crash. The test-drive view also puts the user in the driver's seat of the subject vehicle, but the user can actively control the vehicle. The aerial and 360-degree view options are noninteractive, but the test-drive view allows the user to control the scenario and react to conditions.

In addition to testing the models and safety countermeasures, the immersive experience provided by virtual-reality simulations of crashes and near-crashes has two clear benefits. First, simulations can be used for driver education. For example, teen drivers can use simulations to practice safe driving practices. Second, human factors researchers can use the testbeds to evaluate safety countermeasures, such as in-vehicle information systems, collision avoidance systems, and dynamic message signs. This purpose is well aligned with USDOT's Safe Systems Approach.

Conclusions

The researchers for this project developed two types of RADs to model the causal relationships of crashes at diamond interchanges. The first RADs used data compiled into tables, similar to what modelers use to conduct safety evaluation studies. Statistical and machine-learning models can be developed using these tabular RADs and their relative performance compared using the researchers' rubric. While the rubric standardizes the evaluation process and accommodates the differences in performance measures (e.g., goodness of fit) across different modeling methods, it might still be challenging to compare methods using different measures of effectiveness.

The RADs can be introduced in graduate courses at universities to encourage the application of new statistical and machine-learning approaches or variations of existing approaches. Since these datasets are ready to use, researchers do not have to expend effort obtaining, processing, and preparing the data for model development. A student competition organized by the Transportation Research Board or other professional transportation societies could further encourage using RADs to discover new modeling approaches.

Improving interchange safety involves implementing effective behavioral and engineering countermeasures. The virtual-reality testbeds developed in the study from NDS crash videos offer a realistic and engaging way to support driver education and countermeasure evaluation studies. The testbeds also provide human factors researchers with a fully developed ramp terminal or a speed change lane section with signage and traffic. The modeler can easily add interventions to the testbed and conduct human factors evaluations. The developed testbeds are not dependent on a particular hardware device or system. They can be used across different visualization options, such as head-mounted devices, driving simulators, and 3D projection systems.

While this project demonstrated the proposed RAD framework for interchange facilities, the same framework can be applied to generate artificial data for other roadway facilities. Of particular interest would be those facilities for which it is difficult to obtain accurate and complete real crash data, such as work zones, alternative intersections (e.g., diverging diamond, J-turns), and bicycle facilities. The RAD software developed in this project can easily be extended to include data for additional facilities. In the future, virtual-reality testbeds could also be used to evaluate some behavioral and roadway countermeasures. For example, a driving simulator experiment can be set up using a testbed developed for speed change lanes, and the effect of different driver alert systems (e.g., in-vehicle, dynamic message signs) can be evaluated using study participants.

ACKNOWLEDGMENTS

The original photograph in figure 1 (Imagery © 2020 Maxar Technologies, map data © 2020 Google®, which can be accessed at https://www.google.com/maps/@38.9441707,-91.9411792,583m/data=!3m1!1e3) was modified by adding arrows to show turning movements.

The original photograph in figure 2 (Imagery © 2020 Google, map data © 2021 Google®, which can be accessed at https://www.google.com/maps/@38.944116,-91.9364293,145m/data=!3m1!1e3) was modified by adding a rectangle and labels to show the speed change lane.

REFERENCES

American Association of State Highway and Transportation Officials. 2010. *Highway Safety Manual*. 1st ed. Washington, DC: American Association of State Highway and Transportation Officials.

Elefteriadou, L., C. Fang, R. Roess, and E. Prassas. 2005. "Methodology for Evaluating the Operational Performance of Interchange Ramp Terminals." *Transportation Research Record* 1920, no. 1: 13–24.

FHWA. 2019. "Intersection Safety" (web page). https://highways.dot.gov/research/research-programs/safety/intersection-safety, last accessed April 27, 2020.

FHWA. 2022a. "Center for Accelerating Innovation: Every Day Counts" (web page). https://www.fhwa.dot.gov/innovation/everydaycounts/, last accessed October 19, 2022.

FHWA. 2022b. "Crash Modification Factors Clearinghouse" (website). https://www.cmfclearinghouse.org/, last accessed October 30, 2022.

Google®. 2022. *Google® Maps™*, Mountain View, CA, obtained from: https://www.google.com/maps/@38.9662619,-77.2979777,15z, last accessed October 25, 2022.

Portera, A., and M. Bassani. 2021. "Experimental Investigation into Driver Behavior along Curved and Parallel Diverging Terminals of Exit Interchange Ramps." *Transportation Research Record* 2675, no. 8: 254–267. https://doi.org/10.1177/0361198121997420, last accessed October 26, 2022.

Sun, C., H. Brown, P. Edara, B. Carlos, and K. Nam. 2013. *Calibration of the Highway Safety Manual for Missouri*. Report No. 25-1121-0003-177. Washington, DC: Research and Innovative Technology Administration.

Sun, C., P. Edara, H. Brown, and C. Nemmers. 2016. *Crash Location Correction for Freeway Interchange Modeling: Final Report*. Report No. cmr 16-010. Jefferson City, MO: Missouri Department of Transportation.

USDOT. 2022. *U.S. Department of Transportation Strategic Plan FY 2022–2026*. Washington, DC: U.S. Department of Transportation. https://www.transportation.gov/sites/dot.gov/files/2022-04/US_DOT_FY2022-26_Strategic_Plan.pdf, last accessed September 21, 2022.

Virginia Tech Transportation Institute. 2020. "InSight Data Access Website: SHRP2 Naturalistic Driving Study" (website). https://insight.shrp2nds.us/, last accessed October 19, 2022.

WSDOT. 2016. "Interchange Viewer" (website). https://www.wsdot.wa.gov/mapsdata/tools/InterchangeViewer/default.htm, last accessed April 20, 2022.

Getting Involved with the EAR Program

To take advantage of a broad variety of scientific and engineering discoveries, the EAR Program involves both traditional stakeholders (State department of transportation researchers, University Transportation Center researchers, and Transportation Research Board committee and panel members) and nontraditional stakeholders (investigators from private industry, related disciplines in academia, and research programs in other countries) throughout the research process.

Learn More

For more information, see the EAR Program website at https://highways.dot.gov/research/exploratory-advanced-research. The site features information on research solicitations, updates on ongoing research, links to published materials, summaries of past EAR Program events, and details on upcoming events.

EAR Program Results

As a proponent of applying ideas across traditional research fields to stimulate new problem-solving approaches, the EAR Program strives to develop partnerships with the public and private sector. The program bridges basic research (e.g., academic work funded by National Science Foundation grants) and applied research (e.g., studies funded by State DOTs). In addition to sponsoring projects that advance the development of highway infrastructure and operations, the EAR Program is committed to promoting cross-fertilization with other technical fields, furthering promising lines of research, and deepening vital research capacity.

EXPLORATORY ADVANCED **RESEARCH**

U.S. Department of Transportation
Federal Highway Administration

Recommended citation: Federal Highway Administration, *Multidisciplinary Initiative to Create and Integrate Realistic Artificial Datasets* (Washington, DC: 2023) https://doi.org/10.21949/1522003

FHWA-HRT-23-058
HRTM-30/06-23(Web)E