Quantifying What Houston's High Injury Network Misses with a Validated Design-Risk Model

Vincent Wren

University of Houston Honors College / HPE Data Science Institute

vgwrenla@cougarnet.uh.edu | ORCID 0009-0002-7667-1620

Total pages: 19

Structured Abstract

Objectives: Cities pursuing Vision Zero direct safety investment using High Injury Networks (HINs), maps of corridors where past severe crashes concentrated. Such maps are reactive, listing streets only after harm occurs. This study asks whether a model that scores streets on design characteristics, rather than their individual crash histories, can match an adopted HIN, and how much risk it misses.

Methods: 9,720 severe (killed or seriously injured) crashes from 2016 to mid-2026 are related to the design and context features of 66,922 Houston street segments by crash-frequency regression, supplemented with computer-vision features from 59,261 crowdsourced street images. Validation is spatial (council districts held out) and temporal (crash information frozen at end-2021, graded on 2023 to mid-2026 crashes).

Findings: On the held-out window, the crash-frozen model captured 51 percent of severe crashes versus 46 for the adopted HIN, comparable performance with no evidence of underperformance (interval spanning zero to 10 points). A conventional crash-hotspot map fell from 64 percent in-sample to 39, consistent with regression to the mean. Roughly half the high-design-risk network is absent from the HIN. Those streets worsened against the citywide trend (rate ratio 1.32, $p < 0.001$) while moving with comparable arterials. During the held-out window, 574 severe crashes occurred on those overlooked streets.

Novelty: To the author's knowledge, this is the first spatially and temporally validated full-network comparison of a design-risk model against a city's adopted HIN, and the first custom computer-vision pipeline on raw crowdsourced imagery for full-network crash-frequency modeling.

Practical Applications: The overlooked streets form a proactive companion map to the HIN in a three-tier screening framework. With automated speed and red-light enforcement unavailable in Texas, street design is one of Houston's most durable safety improvement levers.

1. Introduction

Cities that have adopted Vision Zero, the goal of eliminating traffic deaths and serious injuries, face a common allocation problem. Safety budgets cover only a small fraction of the street network, so streets must be ranked to decide where investment goes first, a practice known in the safety literature as network screening. The standard screening instrument is the High Injury Network (HIN), a map of the corridors where past deaths and serious injuries have concentrated (Vision Zero Network 2018). The approach is transparent and politically legible, but it is reactive. A street cannot enter the map until people have been hurt on it, and selection on observed crash counts is statistically unstable, because locations chosen for extreme counts tend to revert toward their long-run averages, a phenomenon known as regression to the mean (Hauer et al. 2002). A screening map directs future investment, so the question it must answer is where harm will concentrate next, not only where it has concentrated in the past.

An alternative approach predicts risk from the characteristics of the roadway itself. The Federal Highway Administration's systemic safety approach formalizes this logic. Rather than waiting for crashes to accumulate at particular locations, an agency identifies the roadway features associated with severe crashes and treats the streets that carry those features (FHWA 2024). What practice and the peer-reviewed literature currently lack is a direct, validated test of the two approaches against each other, using a city's full street network and the screening map the city

adopted. Without that test, the size and character of the reactive map's blind spot, meaning the dangerously built streets it has not yet listed, remain unmeasured.

This study provides that test for Houston, Texas. A crash-frequency model relates each street segment's severe-crash record to its design characteristics, such as lane count, posted speed, median type, and functional class, supplemented by streetscape features extracted with computer vision from free crowdsourced street imagery. The model is estimated from citywide crash records, but it scores each street from its design characteristics alone, so two streets built alike receive similar risk scores regardless of their individual crash histories. The model's rankings are validated on spatially and temporally held-out crashes, compared with the adopted HIN at matched mileage, and the disagreement between the two maps is treated as the object of study. The streets the model flags that the HIN does not, called the overlooked set, are sized, profiled, followed forward in time, and delivered as a Concept Proactive Safety Network following recent agency practice (Alameda CTC 2024).

The stakes are heightened in Texas, where automated enforcement is statutorily unavailable. Transportation Code Section 542.2035 (2007) bars municipal photographic speed enforcement, and House Bill 1631 (2019) ended photographic red-light enforcement. Street design is therefore among the most durable safety levers Texas cities control. Houston's own Vision Zero Action Plan commits the City to identifying high-risk roadway features correlated with severe crashes (Action 2.3), a commitment listed as underway in the City's 2022 annual report (City of Houston 2020, 2023). This study is an open implementation of that commitment.

Three research questions organize the analysis. First, can a model that scores streets on their design characteristics, rather than on their individual crash histories, match the adopted HIN at identifying where severe crashes will occur? Second, how much high-design-risk mileage does the adopted HIN miss, and what happened on those streets after the model flagged them? Third, do features extracted from free crowdsourced street imagery improve the model, and can they separate design-associated risk from otherwise unmeasured street activity?

The contribution is twofold. First, the components of this comparison each exist separately in recent literature, including proactive network-wide screening frameworks (Bonera et al. 2024), temporal consistency testing of urban screening methods (Khattak et al. 2024), and practitioner pairings of reactive and proactive networks (Alameda CTC 2024), but to the author's knowledge no prior peer-reviewed study combines them into a full-network design-risk model validated across space and time against a city's adopted HIN. Second, where recent work has used Mapillary's platform-precomputed detections as model covariates (Elayan et al. 2025), this study is, to the author's knowledge, the first to apply a custom computer-vision pipeline to raw crowdsourced imagery for a full-network, segment-level, all-mode crash-frequency model.

2. Background and Related Work

HIN practice and its limits. The High Injury Network emerged from San Francisco's 2013 WalkFirst analysis, ahead of the city's 2014 Vision Zero adoption, and is now standard practice (Vision Zero Network 2018). Houston's own editions illustrate the method. The 2018 edition, built on 2014 to 2018 data, accompanied the 2020 Action Plan; the 2022 edition (2018 to 2022 data), the comparator used throughout this study, follows a published methodology of half-mile street segmentation, crash points joined within 50 feet, and a minimum severe-crash-density threshold, with no exposure, design, or empirical-Bayes adjustment (City of Houston 2025). The

limits of count-based site selection are well documented. Sites selected on high observed counts revert toward the mean (Hauer et al. 2002; AASHTO 2010), and police-reported crash data substantially under-ascertain severe injury (Taylor et al. 2024). The screening-evaluation literature therefore grades screening methods on time periods after the one that produced them (Cheng and Washington 2008), a discipline this study applies to the Houston 2022 High Injury Network.

Systemic safety. The Federal Highway Administration's systemic approach (FHWA 2024) and NCHRP Report 893 (Thomas et al. 2018) formalize proactive, risk-factor-based screening. Practice is beginning to pair the two maps: Alameda County's 2024 report couples its HIN with a "Concept Proactive Safety Network" built from a threshold checklist of roadway factors, explicitly labeled non-predictive (Alameda CTC 2024). Montgomery County's systemic pedestrian analysis builds predictive risk scores for one mode (Kumfer et al. 2024). What the peer-reviewed literature does not appear to contain, and what this study contributes, is a full-network, fitted, validated design model tested head-to-head against a city's adopted HIN with the overlooked set as the object of study.

Crash-frequency methods. The negative binomial family is the standard of the crash-frequency literature (Lord and Mannering 2010). Model evaluation follows the prediction-model hierarchy: random splits are weak internal validation; held-out blocks along the dimension of intended generalization are stronger (Roberts et al. 2017); temporal validation, developing on one period and evaluating on another, is stronger still (Collins et al. 2015).

Street-view imagery. A growing literature extracts streetscape features from street-view imagery with computer vision and links them to crash outcomes. City-scale crash-frequency precedents used commercial imagery: Google Street View segmentation in Columbus (Stiles et al. 2022) and, the closest methodological antecedent, Baidu imagery under semantic segmentation and object detection in Wuhan, China (Yue 2025). Crowdsourced Mapillary imagery has entered crash research only recently, as platform-precomputed object detections used as covariates in grid-based pedestrian crash count and severity models in Lincoln, Nebraska (Elayan et al. 2025). No prior study, to the author's knowledge, runs its own vision models on raw crowdsourced imagery for a full-network, segment-level, all-mode frequency model. This study does, because the features Mapillary precomputes depend on when each photograph was captured and processed, with full-scene segmentation available only for recent imagery while Houston's coverage is dominated by older photographs. Running open vision models on the raw images yields uniform features across eras. This study documents the cost profile and the complications together.

3. Data

Network. The modeled universe is 66,922 street segments (6,421 centerline miles) inside Houston's full-purpose jurisdiction, built from OpenStreetMap and conflated with Houston Public Works layers (posted speeds, lane counts, median types, traffic counts), TxDOT roadway inventory, and American Community Survey block-group context. Divided boulevards are merged to single centerlines before crash assignment.

Crashes. 421,679 geocoded crashes (TxDOT CRIS, January 2016 through June 2026) are assigned to the nearest segment within 200 feet, with 98 percent assigning inside the cap and the remainder left unassigned. Freeway, ramp, and frontage-road crashes are filtered by CRIS road-

class and road-part codes, plus a named-freeway check for state-route mainlanes coded as US or state highways. The cap almost never binds (median snap distance 4 feet, 99th percentile 106 feet), and severe-crash counts are insensitive to it (assignments at 100/200/250 feet correlate at r >= 0.994). The outcome is killed and seriously injured, KSI (KABCO K plus A), matching the City's HIN definition: 9,720 severe crashes on the modeled network. The extract dates to June 2026. CRIS records arrive with reporting lag, so the final weeks are thin for every map equally, and all claims are rank-based.

Design features and controls. Design: functional class, lanes, posted speed, median type, sidewalk presence, one-way operation, signal count, and intersection connectivity. Controls: traffic volume (ADT), block-group median income, poverty, zero-car households, and population density. ADT from counts covers 26 percent of segments: 4 percent carry a station, and the rest inherit the station value along the same-named corridor. Roadway width is excluded, because 83 percent of it is derived as lanes times twelve feet. Operating speed is excluded as a mediator of the design effect. Race variables are never model inputs.

Predictor vintages. Design and context layers are as-published at retrieval, June 2026: OpenStreetMap geometry, signals, one-way operation, and sidewalk fallbacks (pulled June 14, 2026); Houston Public Works speed, lane, and median layers; TxDOT roadway inventory; ACS 2023 five-year estimates; ADT station readings spanning 2012 to 2026 (most recent valid reading per station). The temporal holdout therefore freezes crash information, not the design inventory: a street redesigned after 2021 enters the model with its current design, in both training and grading. A comparison against archived OpenStreetMap snapshots, bounding the resulting exposure, is reported with the temporal results (Section 5.4).

Street imagery. Mapillary hosts 3.5 million crowdsourced street-level photos within the city boundary. 94 to 100 percent of arterial miles in every council district carry imagery, with no coverage bias across income tiers (62 to 69 percent). Arterials and collectors were sampled every 50 meters, each point matched to the nearest photo within 25 meters (Mapillary positional error is 2 to 6 meters), up to twelve photos per segment: 59,261 photos across 18,133 segments (median 3 per segment, median vintage 2015, 87 percent panoramic).

4. Methods

Crash-history baseline. The reactive comparator is rebuilt from raw public data with best-practice statistical safeguards. A global test confirms that severe crashes cluster far beyond chance (Moran's I; Moran 1950), and Getis-Ord Gi* (Getis and Ord 1992) then identifies where: a street is flagged when its neighborhood's severe-crash total is higher than nearly all random reshuffles of the crash locations. Because tens of thousands of streets are tested at once, some would pass by luck alone, so false-discovery-rate control caps the expected share of false hotspots at 5 percent (Benjamini and Hochberg 1995). A street's neighborhood is defined by actual road connections, segments sharing an intersection, rather than straight-line distance. A simpler 1,000-foot distance band agrees on 93 percent of classifications.

Design-risk model. Severe-crash counts are modeled as NB2 negative binomial regression with a log-length offset, the standard family for crash counts, which are zero-heavy and more dispersed than a Poisson model allows. Overdispersion is decisive (likelihood-ratio test of alpha = 0: statistic 3,193, boundary-corrected $p < 0.001$). A zero-inflated variant offers no improvement (delta-AIC 3, and BIC favors the plain model): the plain specification already

reproduces the observed share of zero-crash segments (predicted 90.9 percent versus 90.8), and the Vuong test once standard for this comparison is invalid for it (Wilson 2015). Standard errors are cluster-robust over 88 Super Neighborhoods, because nearby streets are not independent observations (Cameron and Miller 2015). A conceptual model of covariate roles, specified in full in the repository, governs what enters the model and in what role. Operating speed is excluded as a mediator, because adjusting for the speeds a street invites would absorb the design signal a screening product needs. Functional class and the socioeconomic controls are adjustment covariates reported in a separate panel (Westreich and Greenland 2013). Coefficients are read as conditional associations used for prediction, not as identified causal effects: unmeasured pedestrian exposure and land use preclude identification, and the study's product claims (rankings, capture, the overlooked set) are prediction tasks that do not require it (Hernán, Hsu, and Healy 2019).

Validation. Spatial validation hides entire council districts: the model is fit on ten districts and graded on the eleventh, repeated so each district is held out once (11 folds; Super Neighborhood blocking as a sensitivity). Whole districts are hidden, rather than random streets, because adjacent segments are highly similar and a held-out street's neighbor would otherwise reveal it; blocking follows the dimension of intended generalization (Roberts et al. 2017). The metric is capture share: out-of-fold predictions are pooled citywide, every street is ranked by predicted severe crashes per mile, and one selection at the HIN's own mileage is graded on the share of severe crashes it contains. Standardization constants and imputation medians are computed once on the full network rather than within each training fold, where a fully inductive implementation would estimate them; standardization does not affect fitted means, and the medians are computed from covariates only, never crash outcomes. Two comparators run under identical folds: a context-and-exposure baseline (the offset plus traffic volume and the socioeconomic covariates, with no design features and no functional class) and a default-configuration gradient-boosted benchmark, an untuned nonlinear reference rather than an optimized competitor, carried into the temporal evaluation on the same footing.

Temporal validation separates training and grading in time. The models are refit on 2016 to 2021 crashes, the Gi* map is recomputed on pre-2022 crashes, and the HIN enters as adopted (2018 to 2022 data). The grading window, January 2023 through June 2026, lies outside every model-training and HIN-selection window. The 2022 crash year is excluded from grading because it overlaps the HIN's selection window; grading therefore begins only after every map's selection data closes. This follows the two-period design of the screening-evaluation literature (Cheng and Washington 2008; Collins et al. 2015). Uncertainty comes from Super Neighborhood block-bootstrap intervals: neighboring segments and their crashes are spatially dependent, so blocks rather than segments are resampled. The freeze applies to crash information, not to the roadway inventory, which is current-vintage (Section 3); the test is a crash-history freeze, not a full deployment replication, and Section 5.4 bounds what that distinction could leak. Imagery freeze integrity is tested by masking imagery features on segments whose photographs could postdate 2021. Finally, the overlooked set receives the site consistency test (Cheng and Washington 2008): its before-to-after change in severe-crash rate is adjusted by the citywide trend and tested formally with a set-by-period negative binomial interaction, cluster-robust by Super Neighborhood, against two comparators, all other streets and matched off-HIN arterials and collectors.

Imagery features. Mapillary's own precomputed detections were evaluated first and rejected, because what the platform computes for a photo depends on when that photo was processed: full-scene segmentation exists only for recent imagery, while Houston's coverage is dominated by 2012 to 2017 photographs. Using those features would have confused what a street looks like with when its neighborhood was photographed. Running open vision models on the raw imagery yields the same features for every era instead. Because 87 percent of sampled photos are 360-degree panoramas that perspective-trained models mishandle, each panorama is sliced into four 90-degree flat views. Two open models run per view. SegFormer-B2, fine-tuned on street scenes (Xie et al. 2021; Cordts et al. 2016), labels every pixel and yields scene shares: road, sidewalk, building, vegetation, sky. Faster R-CNN (Ren et al. 2015) on COCO (Lin et al. 2014) counts objects: people, bicycles, vehicles, and traffic lights, with a guard that discards person boxes filling over 15 percent of the frame, which are the photographer in walking captures. Shares are averaged and counts summed across a panorama's views, then averaged over each segment's photos. Five extracted features enter as additional covariates in the otherwise unchanged specification, defining v2: person and vehicle counts (log-transformed) and sidewalk, vegetation, and building pixel shares, with observed-subset standardization, mean-fill, an explicit missingness flag, and a panorama-share measurement control. v1 is the identical model without them. Inference ran overnight on one consumer GPU (RTX 3070, 2.7 images per second). v2 was designed to measure pedestrian exposure directly; its performance in that role is reported in Section 5.6. Two checks support the imagery claims: an availability ablation (v1 plus only the missingness flag and panorama share) separates visual content from coverage geography, and the temporal refit applies a strict leakage bound, masking all imagery features on any segment with any post-2021 photograph within the 25-meter matching radius.

The overlooked set and the Concept PSN. All streets are scored by the validated model and ranked by predicted severe crashes per mile. The final Concept PSN is generated from v2 fit on the full 2016 to 2026 crash period; the frozen refits serve only the temporal evaluation. High-risk sets are taken at the HIN's mileage and at top 5 and 10 percent of street-miles, and crossed with the adopted HIN. The Concept Proactive Safety Network (PSN) smooths the selection into corridors and assigns three priority tiers, following agency presentation conventions (Alameda CTC 2024). Same-street gaps up to a quarter mile are bridged, isolated fragments under half a mile are dropped, and every segment is tiered as on both networks, PSN only, or HIN only.

5. Results

5.1 Baselines

Severe crashes cluster far beyond chance (Moran's I = 0.181, p = 0.001). The Gi* reconstruction places 54 percent of severe crashes on its top 546 miles, the mileage its significance threshold selects. The adopted HIN carries 52 percent on 589 miles. The two crash-based maps agree only partially: they overlap on only 43 percent of segments, and simply switching the Gi* ranking from crash counts to crashes per mile reshuffles its top set substantially (Jaccard 0.56). Crash-based screening is sensitive to method choices that have no analogue in a fitted model.

5.2 The design model

Table 1 reports incidence-rate ratios in two panels; all are conditional associations used for prediction (Section 4). The design features (Panel A) behave as the safety literature expects: streets with more lanes, higher posted speeds, greater connectivity, and more signals carry higher

severe-crash rates, and one-way operation lower (Figure 1). Sidewalk presence shows 1.4 to 1.7, a pedestrian-exposure proxy flagged in every output and never read as a design harm. Sidewalks mark where people walk, and no city dataset counts walking. The same artifact appears in the street-view literature (Yue 2025). Among the adjustment covariates (Panel B, no causal reading; Westreich and Greenland 2013), major arterials carry 2.9 times the severe-crash rate of local streets at the same measured traffic, a class-level stratification pattern, and the income coefficient (0.81) additionally absorbs reporting differences and is never interpreted. Diagnostics: residual Moran's I falls from 0.181 to 0.021; maximum VIF 5.9; race-inclusion sensitivity moves no IRR more than 14.5 percent; a refit on measured-ADT segments flips no significant effect's sign. The model over-predicts the citywide total by 10 percent, concentrated in the top decile (13 to 20 percent), but observed crashes rise at every step across the ten predicted-risk deciles out of fold, so the miscalibration is order-preserving and every product claim is rank-based.

TABLE 1. Incidence-rate ratios from the design model, v1 (NB2; continuous effects per 1 SD; cluster-robust 95 percent confidence intervals).

Panel A: design features	**IRR**	**95% CI**	**p**
Sidewalk both sides *	1.71	1.47 to 1.99	<0.001
Sidewalk partial *	1.54	1.40 to 1.70	<0.001
Sidewalk one side *	1.37	1.24 to 1.53	<0.001
Intersection connectivity	1.22	1.16 to 1.29	<0.001
Lanes	1.21	1.16 to 1.27	<0.001
Signals	1.16	1.13 to 1.19	<0.001
Posted speed	1.10	1.05 to 1.16	<0.001
Median type (four levels)	0.90 to 1.19	n.s.	>0.22
One-way operation	0.74	0.60 to 0.92	0.005

Panel B: adjustment covariates (no causal reading)	**IRR**	**95% CI**	**p**
Major arterial (vs local)	2.88	2.19 to 3.79	<0.001
Arterial (vs local)	2.56	1.99 to 3.29	<0.001
Collector (vs local)	2.37	1.90 to 2.97	<0.001
Minor street (vs local)	1.58	1.21 to 2.06	<0.001
Traffic volume (log ADT)	1.16	1.09 to 1.23	<0.001
Neighborhood poverty	1.11	1.05 to 1.16	<0.001

Panel B: adjustment covariates (no causal reading)	IRR	95% CI	p
Population density	1.06	1.03 to 1.10	<0.001
Zero-car households	1.06	1.01 to 1.11	0.010
Median household income	0.81	0.75 to 0.88	<0.001
Missingness flags (ADT, lanes, income)	0.53 to 0.93	(varies)	

* Pedestrian-exposure proxy, not a design harm (Section 5.6). All entries are conditional associations used for prediction, not identified causal effects.

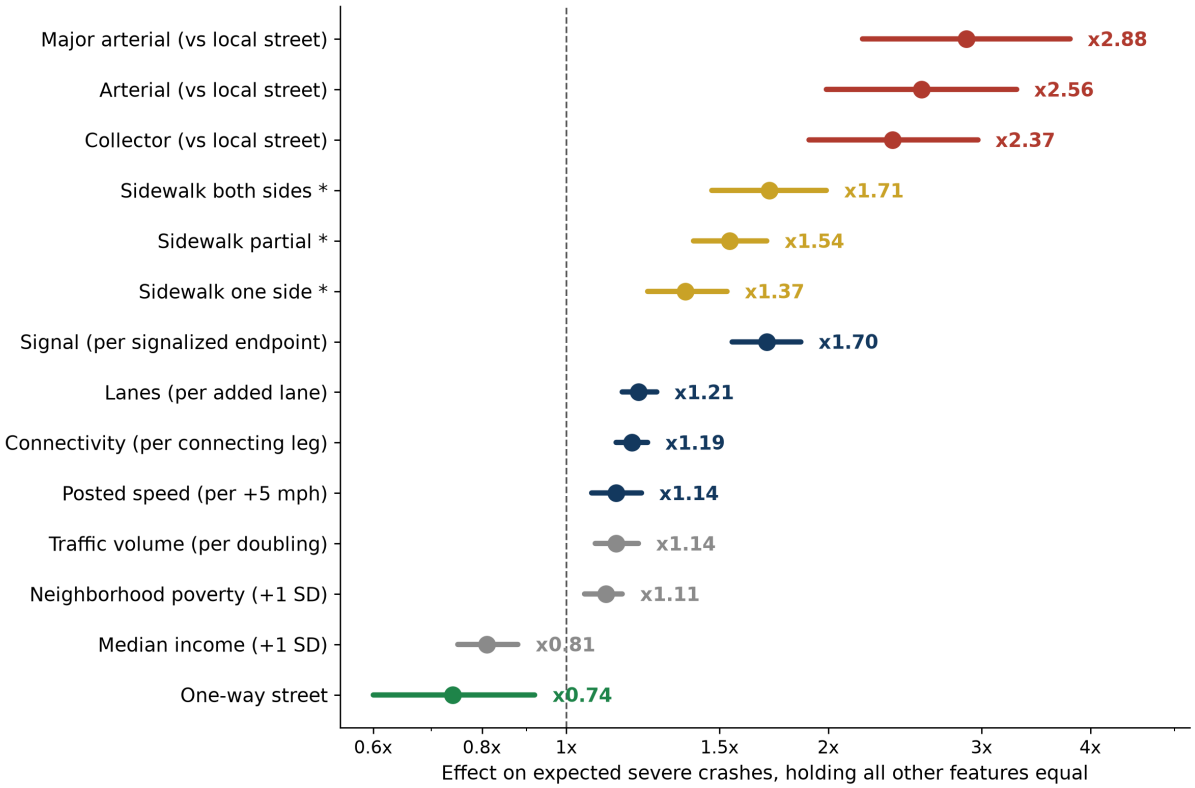

FIGURE 1. Estimated effects on a street's severe-crash rate (forest plot of Table 1). Asterisked sidewalk rows are a pedestrian-exposure proxy, not a design harm (Section 5.6), and red and grey rows are adjustment covariates with no causal reading. Continuous effects are shown in natural units (per added lane, per signalized endpoint, per 5 miles per hour, per doubling of traffic, per connecting leg), converted from Table 1's per-SD ratios; the SES adjustment rows remain per 1 SD.

5.3 Spatial validation

On districts withheld from fitting (out-of-fold predictions pooled citywide, one selection at 589 miles), capture is 47 percent for v1, versus 42 percent for the context-and-exposure baseline and

50 percent for the untuned gradient-boosted reference. Super Neighborhood blocking agrees. The references are graded partly (the HIN, selected on 2018 to 2022 crashes) or entirely (Gi*) on the same crashes that selected them, so matching them out of fold is a conservative result.

5.4 Temporal validation

The temporal evaluation (Table 2) grades every map on the 3,328 severe crashes of January 2023 through June 2026, a window outside every model-training and HIN-selection window: the HIN's 2018 to 2022 selection window closes before the grading window opens, and the models' crash information ends in 2021. With crash information frozen at end-2021, the imagery model (v2) captures 51 percent at 589 miles; the adopted HIN 46 percent; v1 48 percent; the context-and-exposure baseline 43 percent; the untuned gradient-boosted reference, fit on the same pre-2022 data, 53 percent; and the pre-2022 Gi* map 39 percent. The nonlinear benchmark keeps a two-point advantage over the interpretable model, indicating additional predictive structure in the same feature set. The negative-binomial specification remains the primary screening model because it provides interpretable feature associations and a transparent, auditable scoring rule; gradient boosting is retained as a performance benchmark. The Gi* map's fall, from the 64 percent it captures of its own pre-2022 crashes at the same 589-mile footing to 39 on the future, is the expected regression-to-the-mean pattern (Hauer et al. 2002; AASHTO 2010), though temporal change and map-construction differences may also contribute.

The five-point gap carries sampling uncertainty. A Super Neighborhood block bootstrap places the model-minus-HIN difference between -0.2 and +9.8 points with 95 percent confidence, a range consistent with anything from a near tie to a ten-point model advantage and inconsistent with meaningful underperformance. The interval does not quite exclude zero, so under the interpretation rule set before any results were seen, the model is not claimed to outperform the HIN. Formal equivalence is not claimed either; no equivalence margin was set in advance. The supported conclusion is that the two maps performed comparably, with no evidence that the design-based map does worse.

Training-window sensitivities (2016 to 2019, excluding the pandemic anomaly; 2018 to 2021, single injury-definition regime) both hold at 48 percent. Freeze integrity extends to the imagery through a strict leakage bound: 13.6 percent of covered segments have at least one post-2021 photograph anywhere within the 25-meter matching radius, and the bound row in Table 2 withholds imagery features from all of them, so no post-freeze photograph can inform that row by construction. Capture under this deliberately restrictive sensitivity is 49 percent, compared with 51 unmasked and 46 for the HIN. The sensitivity discards those segments' legitimate pre-freeze imagery along with any later photographs (the corpus median capture year is 2015).

Freeze integrity for the roadway inventory receives the same treatment. Because design layers are current-vintage (Section 3), a street redesigned after 2021, possibly in response to the very crashes being predicted, could carry information from the grading period into the frozen model. A comparison against archived OpenStreetMap snapshots bounds that exposure. Comparing every way underlying the model's top miles and the HIN between the 2021-12-31 and 2026 snapshots, 12 percent of the model's top-ranked miles (9 percent of the HIN's) show a value-to-value change in class, lanes, speed, or one-way status, an upper bound that includes tag corrections; most remaining churn is mapping enrichment or id re-cutting, not construction. Excluding every value-changed segment from the temporal selection reduces capture from 51 to 49 percent, still above the HIN's 46, and refitting with those segments also excluded from

training gives the same 49. Predictors without archives (medians, sidewalks, signals, city layers, and context variables) cannot be reconstructed historically.

The site consistency test: the model scores used no post-2021 crash information. The overlooked set used here (302 miles) is the frozen model's version; the deployed model's version appears in Section 5.5. It was defined against the adopted 2018 to 2022 HIN and evaluated on the grading window. It worsened from 0.41 to 0.55 severe crashes per mile-year in the grading window while the citywide rate rose 8 percent. On those miles, 574 severe crashes accumulated. A set-by-period negative binomial interaction (cluster-robust by Super Neighborhood) makes the departure from the citywide trend formal: rate ratio 1.32 (95 percent CI 1.16 to 1.49). Against matched off-HIN arterials and collectors the interaction is null (0.99; 0.84 to 1.16): the deterioration is a functional-class phenomenon, off-HIN arterials worsening while the HIN's streets improve relative to trend (0.84) and the Gi* streets revert (0.61), and the flagged streets move with their class. The model's contribution is therefore locational rather than trajectory-level: it identified, in advance and from design characteristics alone, which streets concentrate the deteriorating class's harm, at more than three times the average street's severe-crash rate.

TABLE 2. Temporal holdout: the grading window (January 2023 through June 2026) lies outside every model-training and HIN-selection window.

Screening map (selection window)	Capture, top 546 mi	Capture, top 589 mi
Design model with imagery, v2 (fit on 2016-2021)	48%	**51%**
v2, strict imagery bound (any post-2021 photo within 25 m masked)	47%	49%
Design model v1 (fit on 2016-2021)	46%	48%
Gradient-boosted reference (untuned, fit on 2016-2021)	51%	53%
Official HIN as adopted (2018-2022 data; 589 mi)		46%
Context-and-exposure baseline (no design or class)	40%	43%
Crash-hotspot map (Gi*, pre-2022 crashes)	37%	39%

Block-bootstrap 95 percent interval (Super Neighborhood blocks) for (v2 minus HIN) at 589 mi: -0.2 to +9.8 points. In-sample references, each map graded on the crashes that built it: the pre-2022 Gi* ranking on its own pre-2022 crashes captures 64 percent at the same 589-mile footing; the adopted HIN on its own 2018 to 2022 selection window, 59 percent. (The 546-mile FDR hotspot set of Section 5.1 is a different object: 54 percent at 546 miles.)

Site consistency test (severe crashes per mile-year, grading window). The vs city column divides each set's post/pre ratio by the citywide ratio of 1.08:

Set	Miles	2016-21 rate	2023-26 rate	Ratio	vs city
Overlooked (pre-fit, off-HIN)	302	0.41	0.55	1.36	**1.26**
Official HIN	589	0.83	0.75	0.91	0.84
Gi* pre-2022 top miles	589	0.98	0.64	0.65	0.61
All streets	6,421	0.14	0.15	1.08	1.00

Interaction rate ratio for the overlooked set, grading window: 1.32 (1.16 to 1.49) versus all other streets; 0.99 (0.84 to 1.16) versus matched off-HIN arterials and collectors (see text).

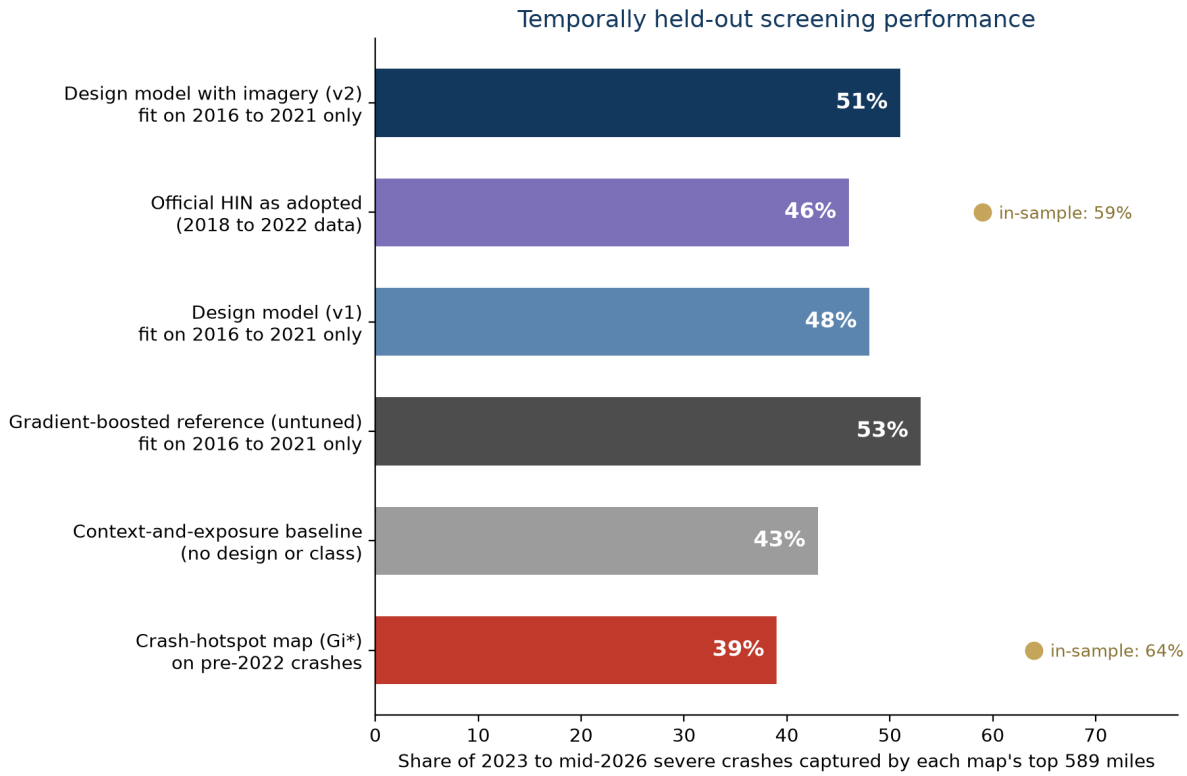

FIGURE 2. Temporally held-out screening performance for the principal Table 2 comparators at 589 miles. Gold dots show each crash-based map's in-sample capture on its own selection-window crashes (HIN: 2018 to 2022; Gi*: pre-2022) and the subsequent decline, consistent with regression to the mean. The models are evaluated on the held-out period.

5.5 The overlooked set and its profile

At the HIN's mileage, 49 percent of the high-design-risk network is absent from the adopted HIN (3,234 segments, 318 miles under the baseline model; the imagery model's set is nearly identical at 3,240 segments), robust at 42 to 51 percent across thresholds. Those streets had already recorded 1,399 severe crashes, spread too thin to clear count-density thresholds. Under the deployed score the overlooked set skews lower-income (median $61,405 versus $69,625

citywide). A design-only sensitivity attributes part of that skew to the model's neighborhood-context adjustment (design-only median $80,313), so the profile is reported as a description of the deployed score with the mechanism disclosed, not as a design finding.

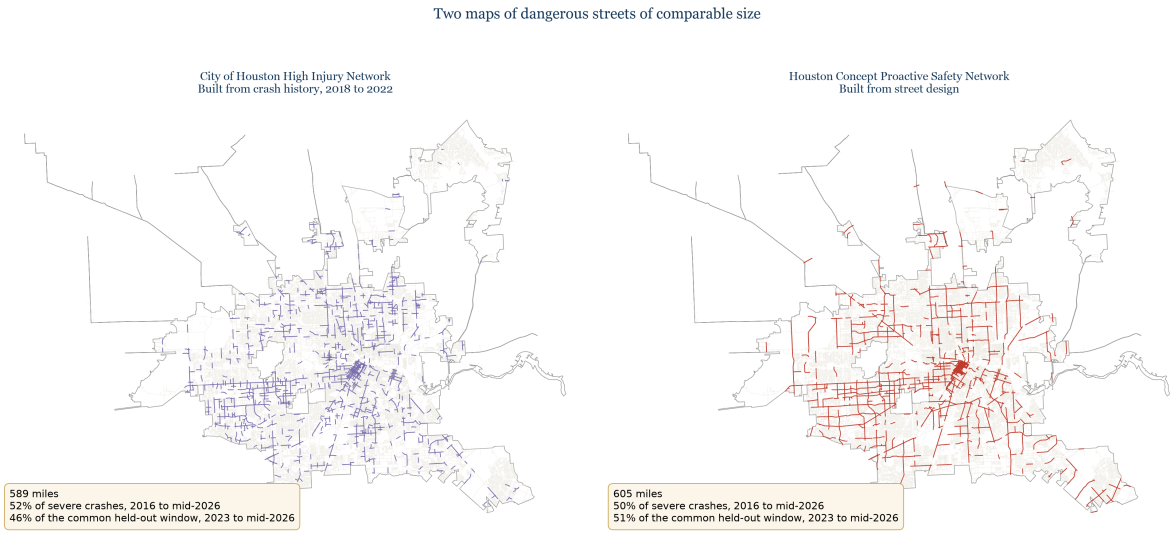

FIGURE 3. Two maps of dangerous streets of comparable size. Both begin from the same 589-mile selection budget. The adopted HIN (left, 589 miles) is built from crash history. The design-based network (right, 605 miles after corridor smoothing) scores every street without using each street's own crash record; its corridors read as longer and more continuous because of the smoothing described in Section 4. The two networks share 304 miles. The stat boxes report each map's capture of past crashes (2016 to mid-2026) and of the held-out window (2023 to mid-2026): the HIN holds more of the past, and the design network performed comparably on the crashes that came after (Table 2).

5.6 What imagery added

Imagery improves fit decisively (AIC 43,740 to 43,418) and out-of-fold capture from 47 to 49 percent, consistently (v2 beats v1 in 10 of 11 folds, sign test $p = 0.012$). The dominant contributor is the vehicle count (1.24 per standard deviation over and above ADT), replicating the central vehicle-count finding of the street-view literature (Yue 2025) and supplying a proxy for traffic activity on the 74 percent of streets without counters. Person counts, by contrast, are null (1.00), an instrument limitation of sparse, low-resolution, mid-2010s-vintage imagery rather than a finding about walking. The sidewalk coefficients accordingly attenuate only partially (1.71 to 1.53), leaving the exposure artifact directionally confirmed but unresolved. An imagery-vintage control moves no coefficient more than 4.5 percent.

An availability ablation separates content from coverage, because imagery coverage is not random. Sampling targeted arterials and collectors means the availability flag itself encodes road class and activity. Adding only the availability information to v1 (missingness flag and panorama share) leaves out-of-fold capture unchanged at 47 percent. Adding the extracted visual features carries the full gain to 49. Within the imagery-covered subset, where availability is constant, v2 exceeds v1 by three points (61 to 64 percent at 589 miles). The gain is visual content, not coverage geography.

5.7 The Concept Proactive Safety Network

Smoothed into corridors at the HIN's mileage, the PSN spans 605 miles carrying 50 percent of 2016 to 2026 and 51 percent of post-2021 severe crashes (Table 3). The 605-mile PSN consists of 304 miles shared with the HIN and 301 PSN-only miles. Adding the 285 HIN-only miles produces a three-tier combined screening framework of 890 miles: both networks (act first), PSN only (the proactive additions), HIN only. The HIN-only tier carries 7.1 percent of citywide predicted design risk versus the PSN-only tier's 23.7 percent: nearly half the adopted HIN's mileage receives substantially lower predicted design-risk scores under the measured covariates, while an equal mileage of high-design-risk streets sits off the list. The model may be missing a genuine design or exposure variable on the HIN-only tier. The claim is about measured design risk, not about those streets being safe.

TABLE 3. Houston's combined three-tier screening framework.

Tier	Miles	Severe 2016-26	Severe 2022-26	Share of predicted design risk
On both networks (act first)	304	3,500	1,501	28.3%
PSN only (proactive additions)	301	1,379	682	23.7%
HIN only (high historical harm, lower modeled design risk)	285	1,549	611	7.1%

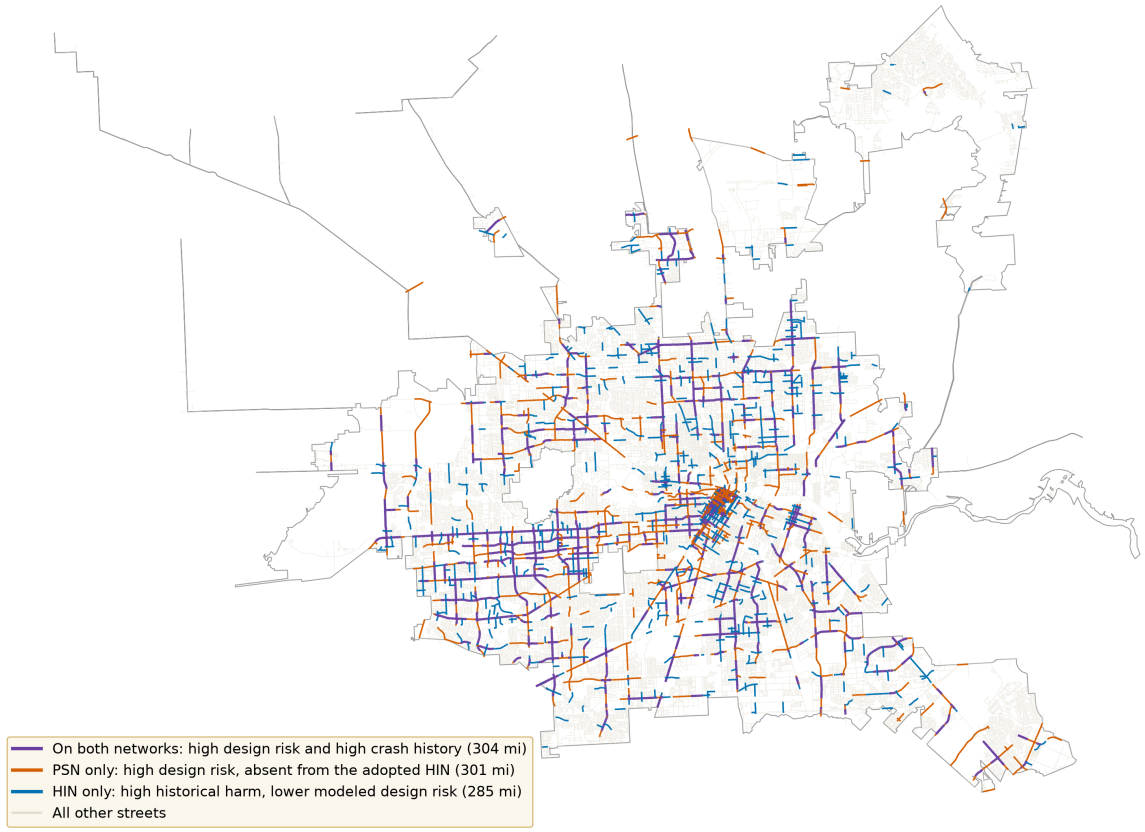

FIGURE 4. The Houston Concept Proactive Safety Network (605 smoothed miles), tiered against the adopted HIN (the 2022 network, built on 2018 to 2022 data, the comparator throughout this paper).

6. Discussion

For screening practice. The HIN performs as crash-history screening should, and nothing here argues for abandoning it. The finding is that a design-based model, scoring streets on their design rather than their individual crash histories, performs comparably to the adopted product on temporally held-out crashes and identifies a complementary network of roughly 300 miles that concentrates harm on an arterial class poorly covered by the reactive list. That class worsened relative to the citywide trend while moving with comparable arterials. The practical proposal is the pairing that agency practice is already moving toward (Alameda CTC 2024; FHWA 2024): a reactive list for where harm has landed, a proactive list for where design invites it, and a three-tier priority scheme where they are combined. For Texas cities, with automated enforcement statutorily unavailable (Transportation Code Section 542.2035; HB 1631), the proactive list is an especially actionable complement to the HIN.

Method sensitivity as a finding. Count-density screening proved unstable in two independent ways: the held-out decline of the raw hotspot map (64 to 39 on the common holdout at the same footing), and the 43 percent overlap between two crash-based maps built from the same crashes. A fitted model with stated covariates is comparable in capture and less exposed to those instabilities, which matters for a document that directs multi-year capital programs.

The natural extension is Empirical Bayes. The Highway Safety Manual's standard for network screening combines exactly these two ingredients: a model's prediction for a street, weighted together with the street's own crash history, the weight set by the model's dispersion parameter (Hauer et al. 2002; AASHTO 2010). This study deliberately keeps the ingredients separate, because the objective is to isolate the incremental predictive contribution of design and context variables independent of site crash history, and the overlooked set exists only under that separation. An EB combination would likely capture more than either component and is the appropriate deployment tool for prioritizing established streets, while the crash-history-free model remains the tool for screening what the crash record cannot yet see. Quantifying the combination on this network is left as future work.

Cost and reproducibility. The pipeline uses open data, open models, and one consumer GPU. The imagery phase cost approximately nothing. The crowdsourced-imagery caveats are documented rather than elided: coverage is excellent on arterials, vintage is heterogeneous, panoramas require slicing, and pedestrian counting fails at thumbnail resolution. Agencies replicating this work should treat vehicle counts as reliable and person counts as requiring newer, higher-resolution imagery.

Limitations. Pedestrian exposure remains substantially unmeasured, and the sidewalk coefficient remains an exposure proxy. Land use is unmeasured citywide. Individual coefficients are therefore interpreted as conditional predictive associations rather than causal effects, and the study's product claims (rankings, capture, the overlooked set) are prediction tasks that do not require causal identification (Hernán, Hsu, and Healy 2019). Police-reported severity is substantially under-ascertained. Ratio estimates are protected under stated assumptions, while absolute counts are not. Traffic volumes are imputed on 74 percent of segments; the imputations carry flags, and a measured-subset refit is reported as sensitivity. Point predictions in the top tail run 13 to 20 percent high. Rankings, not counts, are the supported claim. The forward test's worsening is relative to the citywide trend. Against matched off-HIN arterials and collectors the interaction is null, so the overlooked-set result identifies which streets carry a deteriorating class's harm, not street-level departures from the class trajectory. Design and context layers are current-vintage (June 2026), so the crash-history-frozen model sees post-2021 roadway changes in training and grading alike. Section 3 states the vintages, Section 5.4 bounds the exposure against archived OpenStreetMap snapshots, and re-scoring changed segments with their archived 2021 attributes is left as future work. Comparative performance claims are descriptive; no equivalence margin was set in advance, so equivalence is not formally established. The equity profile of the overlooked set is partly produced by the deployed score's own context adjustment, and is reported with that mechanism disclosed. Findings are one city, 2016 to 2026. Transfer is untested.

7. Conclusion

A design-based severe-crash model for Houston, validated across space and time, performed comparably to the City's adopted High Injury Network at anticipating severe crashes outside both products' crash-information windows, with no evidence of underperformance, scoring each street on its design rather than its own crash record (RQ1: 51 versus 46 percent on the common holdout). At matched mileage, roughly half the high-design-risk network is absent from the adopted HIN. Those streets concentrate the harm of an off-HIN arterial class that worsened relative to the citywide trend (rate ratio 1.32, formally significant) while moving with

comparable arterials; during the held-out window, 574 severe crashes occurred on those overlooked miles (RQ2). Free crowdsourced imagery measurably sharpened the model through vehicle counts, a gain an availability ablation attributes to visual content rather than coverage geography (RQ3), while the pedestrian instrument's failure is documented for the field. With automated speed and red-light enforcement statutorily unavailable, design is among the most durable levers Texas cities control. The Concept Proactive Safety Network shows where it applies before the crash history arrives.

Acknowledgments

The author thanks [advisor name and title, TODO] for guidance and review; the Pharis Fellowship of the University of Houston Honors College and the HPE Data Science Institute for supporting this work; and the office of Council Member Panzarella (District C) for facilitating data access. Any errors are the author's alone. [Wording pending advisor approval.]

Disclosure of Generative AI Use

Generative AI (Claude Code, Anthropic) was used, under the author's direction, for development of analysis code, data visualization, assistance compiling literature search results, and drafting and revising manuscript text. Per TRB policy, this use is also disclosed in the Editorial Manager submission form. The author made all methodological decisions, recorded them in a dated public project log before estimation, and verified all analytical outputs and every reference against original sources. The author takes full responsibility for the content. All code and derived data: github.com/WrenVin/PharisFellowshipVisionZero.

Data Availability

Derived segment-level data, all analysis code, and generated reports are public in the repository above. Raw crash records are available from TxDOT CRIS on request. Mapillary imagery is available under its terms of service.

References

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

Alameda County Transportation Commission. 2024. High-Injury Network and Proactive Safety Network Report.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society, Series B 57 (1): 289-300.

Bonera, Michela, Benedetto Barabino, George Yannis, and Giulio Maternini. 2024. "Network-Wide Road Crash Risk Screening: A New Framework." Accident Analysis and Prevention 199: 107502.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." Journal of Human Resources 50 (2): 317-372.

Cheng, Wen, and Simon Washington. 2008. "New Criteria for Evaluating Methods of Identifying Hot Spots." Transportation Research Record 2083: 76-85.

City of Houston. 2020. Vision Zero Action Plan.

City of Houston. 2023. Vision Zero Annual Report 2022.

City of Houston. 2025. Houston Vision Zero High Injury Network GIS layers and metadata (2022 and 2025 editions). Planning and Development Department, COHGIS Open Data Portal.

Collins, Gary S., Johannes B. Reitsma, Douglas G. Altman, and Karel G. M. Moons. 2015. "Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement." Annals of Internal Medicine 162 (1): 55-63.

Cordts, Marius, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. "The Cityscapes Dataset for Semantic Urban Scene Understanding." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3213-3223. https://doi.org/10.1109/CVPR.2016.350

Elayan, Mohammad, Sagun Karki, and Jason Hawkins. 2025. "Better Safety Analyses Through Smarter Data: Adding Open-Street-View and Traffic-Calibrated Location-Based Services Data to Pedestrian Crash Analysis in Lincoln, NE." Transportation Research Record 2680 (7): 261-274. https://doi.org/10.1177/03611981251391714

FHWA. 2024. Systemic Safety User Guide. Washington, DC: Federal Highway Administration.

Getis, Arthur, and J. Keith Ord. 1992. "The Analysis of Spatial Association by Use of Distance Statistics." Geographical Analysis 24 (3): 189-206.

Hauer, Ezra, Douglas W. Harwood, Forrest M. Council, and Michael S. Griffith. 2002. "Estimating Safety by the Empirical Bayes Method: A Tutorial." Transportation Research Record 1784: 126-131.

Hernán, Miguel A., John Hsu, and Brian Healy. 2019. "A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks." Chance 32 (1): 42-49.

Khattak, Muhammad Wisal, Hans De Backer, Pieter De Winne, Tom Brijs, and Ali Pirdavani. 2024. "Comparative Evaluation of Crash Hotspot Identification Methods: Empirical Bayes vs. Potential for Safety Improvement Using Variants of Negative Binomial Models." Sustainability 16 (4): 1537.

Kumfer, Wesley, Jesse Cohn McGowan, Krista Nordback, Mike Vann, Bo Lan, Brian Frizzelle, and Libby Thomas. 2024. "Systemic Predictive Safety Analysis of Pedestrian Crashes for Montgomery County's Vision Zero Program." Transportation Research Record 2678 (11): 2165-2180. https://doi.org/10.1177/03611981241247178

Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and C. Lawrence Zitnick. 2014. "Microsoft COCO: Common Objects in Context." In Computer Vision - ECCV 2014, 740-755. Cham: Springer. https://doi.org/10.1007/978-3-319-10602-1_48

Lord, Dominique, and Fred Mannering. 2010. "The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives." Transportation Research Part A 44 (5): 291-305.

Moran, P. A. P. 1950. "Notes on Continuous Stochastic Phenomena." Biometrika 37 (1/2): 17-23.

Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2015. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." In Advances in Neural Information Processing Systems 28, 91-99.

Roberts, David R., Volker Bahn, Simone Ciuti, Mark S. Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, Jose J. Lahoz-Monfort, Boris Schroder, Wilfried Thuiller, David I. Warton, Brendan A. Wintle, Florian Hartig, and Carsten F. Dormann. 2017. "Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure." Ecography 40 (8): 913-929. https://doi.org/10.1111/ecog.02881

Stiles, Jonathan, Yuchen Li, and Harvey J. Miller. 2022. "How Does Street Space Influence Crash Frequency? An Analysis Using Segmented Street View Imagery." Environment and Planning B: Urban Analytics and City Science 49 (9): 2467-2483. https://doi.org/10.1177/23998083221090962

Taylor, Nandi L., Mike Dolan Fliss, Sharon E. Schiro, and Katherine J. Harmon. 2024. "Comparative Analysis of Injury Identification Using KABCO and ISS in Linked North Carolina Trauma Registry and Crash Data." Traffic Injury Prevention 25 (7): 912-918. https://doi.org/10.1080/15389588.2024.2361052

Texas Legislature. 2007. House Bill 922: Limitation on Municipalities (Automated Speed Enforcement). Codified at Texas Transportation Code Section 542.2035. 80th Legislature.

Texas Legislature. 2019. House Bill 1631: Prohibition of Photographic Traffic Signal Enforcement Systems. 86th Legislature.

Thomas, Libby, Laura Sandt, Charles Zegeer, Wesley Kumfer, Katy Lang, Bo Lan, Zachary Horowitz, Andrew Butsick, Joseph Toole, and Robert J. Schneider. 2018. Systemic Pedestrian Safety Analysis. NCHRP Research Report 893. Washington, DC: Transportation Research Board. https://doi.org/10.17226/25255

Vision Zero Network. 2018. "HIN for the WIN."

Westreich, Daniel, and Sander Greenland. 2013. "The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients." American Journal of Epidemiology 177 (4): 292-298.

Wilson, Paul. 2015. "The Misuse of the Vuong Test for Non-Nested Models to Test for Zero-Inflation." Economics Letters 127: 51-53.

Xie, Enze, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. 2021. "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers." In Advances in Neural Information Processing Systems 34 (NeurIPS 2021), 12077-12090.

Yue, Han. 2025. "Investigating Streetscape Environmental Characteristics Associated with Road Traffic Crashes Using Street View Imagery and Computer Vision." Accident Analysis and Prevention 210: 107851. https://doi.org/10.1016/j.aap.2024.107851