Safety21

INNOVATING FOR SAFETY

US DOT National
University Transportation Center for Safety

Carnegie Mellon University

Penn University of Pennsylvania

The Ohio State University

Morgan State University

CCAC Community College of Allegheny County

The University of Texas Rio Grande Valley

Enhanced Crash Risk Estimation in Urban Environments: Integrating Multi-Source Data and Advanced Modeling Approaches for the City of Pittsburgh

PI: Sean Qian (ORCID: 0000-0001-8716-8989)

Lead Researcher: Tao Tao (ORCID: 0000-0001-9865-5985)

Researcher: Leonard Weiss (ORCIRD: 0000-0001-6857-6573)

Final Report – July 31, 2026

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, under [grant number 69A3552344811 / 69A3552348316] from the U.S. Department of Transportation's University Transportation Centers Program. The U.S. Government assumes no liability for the contents or use thereof.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
Enter the report number assigned by the sponsoring agency.		
4. Title and Subtitle Enhanced Crash Risk Estimation in Urban Environments: Integrating Multi-Source Data and Advanced Modeling Approaches for the City of Pittsburgh		**5. Report Date** July 31, 2026 **6. Performing Organization Code** Enter any/all unique numbers assigned to the performing organization, if applicable.
7. Author(s) Tao Tao, PhD, (ORCID: 0000-0001-9865-5985) Sean Qian, PhD, (ORCID: 0000-0001-8716-8989)		**8. Performing Organization Report No.** Enter any/all unique alphanumeric report numbers assigned by the performing organization, if applicable.
9. Performing Organization Name and Address Carnegie Mellon University 5000 Forbes Ave, Pittsburgh, PA 15213		**10. Work Unit No.** **11. Contract or Grant No.** Federal Grant # 69A3552344811 / 69A3552348316
12. Sponsoring Agency Name and Address Enter name and address of the organization(s) financially responsible for the work. After each agency name, enter funding type (e.g. SPR). Example: Missouri Department of Transportation (SPR)		**13. Type of Report and Period Covered** Final Report (July 1, 2025 – July 31, 2026) **14. Sponsoring Agency Code** USDOT

15. Supplementary Notes			
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.			

16. Abstract

High Injury Networks help inform the decision which streets an agency treats first, and they are almost always built from police-reported crashes alone. That record is known to miss injuries, and to miss crashes unevenly in space. We test whether a second, independently collected source names different high injury streets. Using City of Pittsburgh data for 2020 to 2024, we merge the street centerline into 4,332 corridors covering 933.5 miles, build a KABCO-aligned severity index for 14,298 emergency medical dispatches from crew-recorded disposition, transport mode and patient-handling fields corrected by a negation-aware reading of the scene text, and assign each of the 18,267 crashes and each dispatch to a single corridor. We then apply the City's rule identically to each layer, flagging a corridor when any 1,500 ft window holds five or more fatal-and-injury events or any 200 ft window holds two or more serious-or-fatal events. EMS attends fewer traffic events than police but records more injuries, 6,529 against 5,429. The two layers select overlapping but distinct networks: 128 crash corridors and 164 EMS corridors, agreeing on 100 of those corridors, which carry 73.0 percent of flagged mileage. The 64 EMS-only corridors are mostly Local and Collector streets and mark candidate under-reported segments. The procedure requires no personal identifiers and no record-level match across data sources.

17. Key Words	18. Distribution Statement
High Injury Network (HIN), Emergency medical services dispatch data, Crash under-reporting, injury severity, Corridor-level network screening	No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161. Enter any other agency mandated distribution statements. Remove NTIS statement if it does not apply.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	22	-

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

Enhanced Crash Risk Estimation in Urban Environments: Integrating Multi-Source Data and Advanced Modeling Approaches for the City of Pittsburgh

Tao Tao[1], Sean Qian[1,2], and Leonard Weiss[3]

[1]Department of Civil and Environmental Engineering, Carnegie Mellon University
[2]Heinz College of Public Policy and Information Systems, Carnegie Mellon University
[3]Department of Emergency Medicine, University of Pittsburgh School of Medicine

August 18, 2026

Abstract

High Injury Networks help inform the decision which streets an agency treats first, and they are almost always built from police-reported crashes alone. That record is known to miss injuries, and to miss crashes unevenly in space. We test whether a second, independently collected source names different high injury streets. Using City of Pittsburgh data for 2020 to 2024, we merge the street centerline into 4,332 corridors covering 933.5 miles, build a KABCO-aligned severity index for 14,298 emergency medical dispatches from crew-recorded disposition, transport mode and patient-handling fields corrected by a negation-aware reading of the scene text, and assign each of the 18,267 crashes and each dispatch to a single corridor. We then apply the City's rule identically to each layer, flagging a corridor when any 1,500 ft window holds five or more fatal-and-injury events or any 200 ft window holds two or more serious-or-fatal events. EMS attends fewer traffic events than police but records more injuries, 6,529 against 5,429. The two layers select overlapping but distinct networks: 128 crash corridors and 164 EMS corridors, agreeing on 100 of those corridors, which carry 73.0 percent of flagged mileage. The 64 EMS-only corridors are mostly Local and Collector streets and mark candidate under-reported segments. The procedure requires no personal identifiers and no record-level match across data sources.

1 Introduction

Safety resources are finite, so an agency that commits to eliminating traffic deaths and severe injuries has to decide which streets to treat first. The High Injury Network is a popular instrument the field uses to make that decision, and the City of Pittsburgh maintains one, built from police-reported crashes under an explicit rule on distance and severity (City of Pittsburgh, n.d.). Which corridors it names decides which streets get studied and which projects get funded.

This project set out to sharpen crash risk estimation for Pittsburgh by drawing on more than the crash record alone. Its premise was that a risk map built from observed crashes may not locate risk accurately, because crashes are sporadic and the record that captures them is incomplete, and it pointed to the inconsistency between the City's crash-based risk maps and what residents report through near-miss and complaint channels. The purpose of this research project is to improve the accuracy of the High Injury Network with additional data sources.

The additional source used here is the emergency medical dispatch record. Ambulances are dispatched by a different agency, on a different trigger, and their records carry scene coordinates, which makes them an independent observation of where people are hurt rather than a derivative of the crash report. Linking individual ambulance runs to individual crash reports would require personal identifiers this study does not hold, so the comparison is made at the level of the network instead.

We create a new framework to make the comparison possible. We first apply a new approach that merges the City's street centerline into corridors automatically, replacing the City's manual process. Because a dispatch record carries no severity code comparable to the KABCO scale on a crash report, we build a severity index for the dispatch file from what the attending crew recorded, its disposition, transport mode and patient-handling fields, corrected by a negation-aware reading of the free-text scene description. We then assign each crash and each dispatch to a single corridor, and we slide two window scales along every corridor, 1,500 ft for injuries accumulating along a stretch of street and 200 ft for severe outcomes clustering at a point, flagging a corridor if any window at either scale passes the City's threshold. Running the identical procedure over the crash layer and then over the dispatch layer, we obtain two High Injury Networks whose only difference is the source that fed them, and we classify every corridor by whether both flag it, only the crash layer does, or only the dispatch layer does.

The rest of the report is organized as follows. Section 2 reviews how High Injury Networks are built, what the police crash record is known to miss, and what has and has not been done with emergency records inside that practice. Section 3 sets out the data and the method, including the severity index that makes the two sources comparable. Section 4 reports the corridor network, the severity profile of the dispatch layer, and the comparison itself. Section 5 concludes.

2 Literature review

2.1 High Injury Networks and how they are built

Vision Zero commits to eliminating traffic deaths and severe injuries, which forces a prior question: which streets first. The High Injury Network is one of the devices the field has settled on. It maps a street network's corridors, counts the people killed and injured on each, and designates the small share of the network that carries a disproportionate part of the burden. Madison, Wisconsin scores intersections and mid-block segments by a severity-weighted count and joins those above a percentile threshold into corridors (Claros et al., 2022). San Francisco, California slides overlapping quarter-mile sections along every block and keeps the highest scoring, arriving at a network of 12.2 percent of city street mileage that carries 68.2 percent of severe and fatal injuries (San Francisco Department of Public Health, 2022).

The method is standard practice. Among the 2023 Vulnerable Road User Safety Assessments, which every state completed under 23 U.S.C. 148(l), 48 percent identified high-risk areas using High Injury Networks and 92 percent drew on the state crash database (Federal Highway Administration [FHWA], 2024). Pittsburgh sits inside that practice with one deliberate departure. Many high injury networks count only serious and fatal injury. The City counts all fatal and injury crashes, including suspected minor and possible injury, to remove discrepancies and subjectivity in police crash reports and to widen the sample enough to find injury hotspots before they produce a fatality (City of Pittsburgh, n.d.).

2.2 What the police crash record misses

A network built on one source inherits that source's defects, and the police crash record has two that are well documented. The first is under-reporting. Hosseinzadeh et al. (2022) compile roughly two dozen studies linking police crash records to hospital, EMS or insurance data and report match rates spanning 29.8 percent to 74 percent. Comparing counts rather than records, Nijman (2025) found that ambulance records in one Dutch province registered about seven times as many road traffic victims as the police registers. What goes missing is patterned rather than randomly thinned. Across the studies Hosseinzadeh et al. (2022) review, the records missing from one system differ systematically by severity, crash type, age, gender and location. The last of these is the direct precedent for this report: if unreported events are patterned in space, then a network built on the crash record alone is undercounted *unevenly*, and the shortfall is in principle recoverable as a spatial signal rather than as a scalar correction.

The second defect is that severity is coded by an officer rather than a clinician. The KABCO scale is assigned on limited information at the scene, and Hosseinzadeh et al. (2022) quantify its disagreement with trauma-registry injury severity scores and emergency-department disposition. This is the defect the City's all-injury rule was designed around.

Under-reporting in Pennsylvania has a specific legal origin. Notice to police is required only where a crash injures or kills someone, or disables a vehicle enough to require towing (75 Pa.C.S. § 3746). Both conditions are judged at the scene by the people involved, so a collision that leaves the vehicles driveable and produces no roadside injury complaint falls outside the record by construction rather than by failure.

2.3 EMS as a additional source, and what has not been done with it

EMS records are an important additional source that complements the crash record. The field's method has been to link the two case by case, as the Crash Outcome Data Evaluation System did for hospital and crash records state by state (National Highway Traffic Safety Administration, 2015). Linkage to EMS records is rarer, and the obstacles are institutional as much as technical: different agencies hold the records, open extracts strip the personal identifiers a match would need, and privacy law places health information out of reach (Hosseinzadeh et al., 2022).

The closest precedent is Hosseinzadeh et al. (2022), who matched 21,358 police-reported crashes against 5,473 EMS runs in Jefferson County, Kentucky, without personal identifiers. Its lasting contribution is the taxonomy: an unmatched record is not one thing, and the outcome "0 Crash, 1 EMS", an ambulance run with no corresponding crash report, is the candidate under-reported crash. The match rates are asymmetric, 72.2 percent of EMS runs matching a crash against 18.5 percent of crashes matching an EMS run. A matched set is also a selected set, clinically as well as clerically. Patients transported by EMS carry higher injury severity scores than those arriving otherwise (Huang et al., 2016), and only 6.4 percent of crash patients attended by ambulance are high acuity (Ceklic et al., 2020). Any set produced by comparing two reporting systems is selected by both, so its composition must be checked before divergence is read as reporting failure.

A dispatch record has no severity field to set against KABCO, but it is not silent on how badly someone was hurt. Disposition, transport mode and patient-handling fields are recorded by the attending crew rather than by an officer, on a nationally standardized schema (NEMSIS Technical Assistance Center, n.d.), and alongside them sits a free-text scene description, readable only with attention to negation because dispatch text records the absence of a finding as routinely as its presence (Chapman et al., 2001).

Inside High Injury Network practice, the dispatch record has barely appeared. Only three of the state assessments used EMS data at all (FHWA, 2024), and where health data did appear it measured response rather than incidence, as in Oregon's response and transport times, computed from its state prehospital system while its high-risk areas came from a separate roadway-risk score (Oregon Department of Transportation, 2023). Missouri went furthest and still stopped short, finding that hospital data could capture under-reported crashes but never integrating it into the analysis behind the state's high-risk facilities (Sun et al., 2023). San Francisco combines police, hospital and medical-examiner records, but EMS data enter only as the means of locating hospital-identified injuries (San Francisco Department of Public Health, 2022). Pande et al. (2026) show that emergency-call and official crash records differ substantially and recommend building conventional and supplemental networks side by side, without carrying that recommendation out.

EMS has never been the network's primary input. Patient-level linkage was impossible in the Dutch study, so Nijman (2025) compared the two systems in aggregate instead, at province scale and without corridor geography. In this study, we run one rule over two independently collected layers and compare them at corridor level, which keeps the geography and still requires no personal identifiers.

3 Data and method

3.1 Data

The analysis rested on multiple important datasets. Two supply the events being compared (police-reported crashes and EMS traffic dispatches), and two establish the spatial frame on which that comparison is made: the City street centerline, which defines the units of analysis, and the municipal boundary, which fixes the study area. Table 1 summarizes the datasets.

The study period is calendar years 2020 through 2024, the most recent five-year datasets we could obtain. We used the five-year period to match the City's High Injury Network methodology.

Table 1: Datasets used in the analysis.

Data name	Description	Source
Pittsburgh street centerline	Polyline street network for the City of Pittsburgh (19,622 segments) carrying ownership, functional class, roadway type, street name, and roadway width. Supplies the geometry from which the corridors and sliding analysis windows are built.	City of Pittsburgh
Police-reported crash records	Statewide crash reporting file for Allegheny County, calendar years 2020–2024 (53,929 crashes, of which 18,570 in Pittsburgh). Supplies crash location and the KABCO maximum-injury-severity code on which the High Injury Network rule operates.	Pennsylvania Department of Transportation (PennDOT)
EMS traffic dispatch records	Unit-level dispatch records for motor-vehicle-related EMS responses in Pittsburgh, 2019–2024 (37,080 unit records spanning 19,503 distinct calls), with scene coordinates, the full dispatch-to-available timestamp chain, disposition and patient-handling fields, and free-text scene descriptions.	City of Pittsburgh Bureau of Emergency Medical Services
City of Pittsburgh boundary	Municipal boundary polygon derived from 2020 Census geography and current as of August 2023, covering the City proper (55.41 sq mi) and the rivers (2.94 sq mi). Used to restrict both event sets to the study area.	City of Pittsburgh

Street centerline. The street centerline is the backbone of the analysis: it determines what counts as a location, and therefore what can be compared. Not every street is eligible for High Injury Network treatment, since the City's methodology addresses surface streets under public jurisdiction. Segments are accordingly restricted to those owned by the City, the County, or the State and functionally classified as Local, Collector, Minor Arterial, or Principal Arterial; ramps, tunnels, limited-access highways, bridges, and interstates are removed, as are two limited-access facilities identified by the City (Ohio River Boulevard and the Crosstown Boulevard). This leaves 14,856 eligible segments totaling 933.5 centerline miles: 11,417 Local, 2,238 Collector, 1,083 Minor Arterial, and 118 Principal Arterial.

Police-reported crashes. PennDOT's crash reporting system records every crash on a public roadway meeting the State's reporting threshold of injury, fatality, or a vehicle requiring towing (75 Pa.C.S. § 3746). The data are distributed as annual files organized into related tables (crash, person, vehicle, roadway, and flag records); this analysis draws on the crash-level table, using the crash record number, decimal latitude and longitude, municipality code, and maximum injury severity. Severity follows the KABCO scale, coded 1 through 4 for fatal (K), suspected serious (A), suspected minor (B), and possible (C) injury, with separate codes for uninjured, injured-but-unknown-severity, and unknown. Of 53,929 Allegheny County crashes in 2020–2024, 18,570 carry the Pittsburgh municipality code, 18,521 of which are geocoded, and 18,267 fall inside the study area. Figure 1 maps them by severity, and Table 4 gives the breakdown alongside the EMS layer.

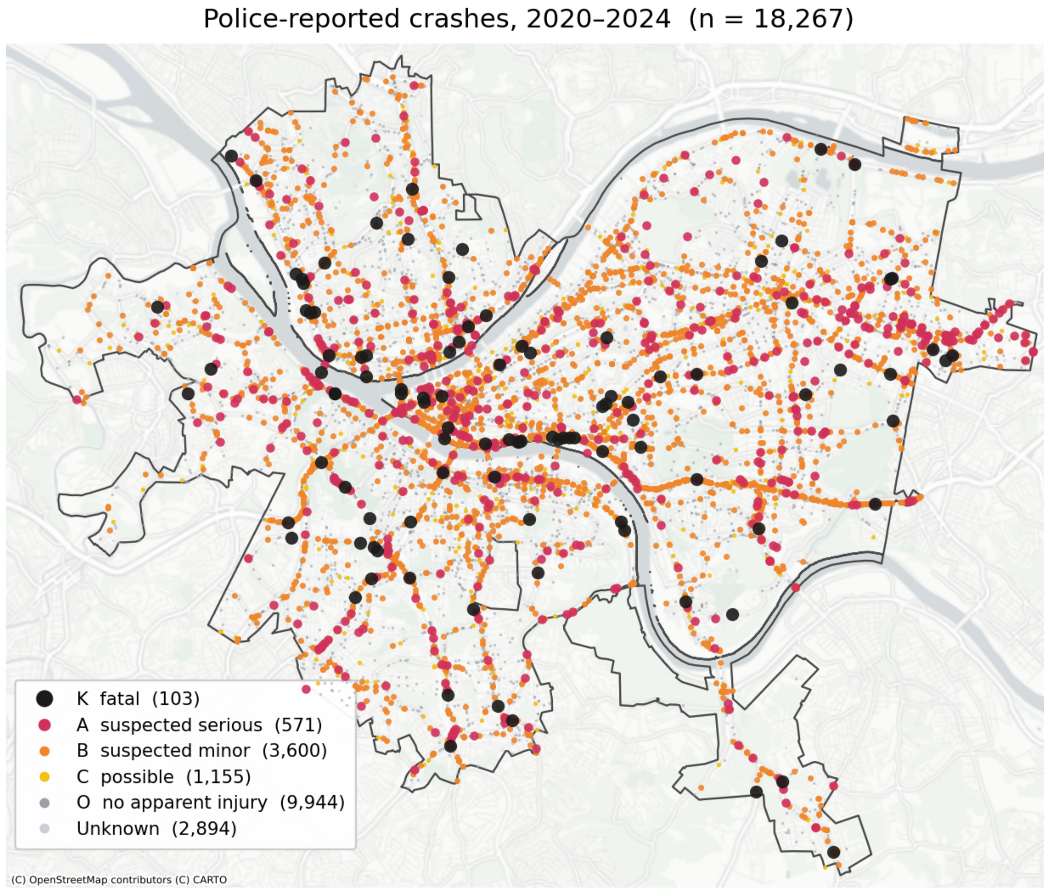

Figure 1: Police-reported crashes by KABCO severity, 2020–2024, over the study area. Severity is read directly from the crash report.

EMS traffic dispatches. The EMS file contains one row per responding unit rather than one row per event. Because a traffic call routinely draws more than one unit (1.90 on average, and as many as 14), the 37,080 unit records represent 19,503 distinct calls, and the file must be collapsed to incident level before any spatial count is meaningful. Each record carries the dispatching unit and station, the full timestamp chain from dispatch through en route, arrival, and departure, scene coordinates, the unit's disposition, several patient-handling and transport-mode fields, and a free-text scene description.

One property of this file shapes the method that follows: it carries no injury severity field comparable to KABCO. Severity is recorded only indirectly, through disposition and patient-handling codes that are populated for transported patients, and through the free text. A severity index must therefore be constructed before the EMS layer can be put through the same rule as the crash layer, and Section 3.2.2 describes its construction. Restricting to 2020–2024 and to incidents with usable scene coordinates inside the study area leaves 14,298 EMS incidents for analysis.

City boundary. The boundary layer restricts both event sets to the study area. It is distributed as three polygons: the City proper, the rivers, and Mt. Oliver Borough, a separate municipality of 0.34 square miles fully enclosed by Pittsburgh. Mt. Oliver is excluded from the mask, which therefore covers 58.34 square miles. The exclusion keeps the two event sets under a single spatial definition, since crashes are independently filtered by PennDOT municipality code and would omit the borough in any case, and it matches the scope of the analysis, since the City holds no jurisdiction over Mt. Oliver streets and the City centerline covers little of the borough's interior.

3.2 Method

3.2.1 Corridors and sliding windows

The High Injury Network is expressed in terms of corridors. A segment breaks at every intersection, so a single corridor is split across dozens of records and a concentration of injury that an engineer would describe as one dangerous stretch is divided among several of them. The rule in Section 3.2.3 counts injuries inside a window of fixed length, which is meaningful only if the geometry the window slides along runs continuously through intersections. The corridor is that continuous unit, and this section builds it. Figure 2 illustrates the idea on a schematic network, and Figure 3 summarizes the construction as a pipeline.

Building the corridor network. The 14,856 eligible segments are assembled into an undirected multigraph whose nodes are segment endpoints, that is, intersections and dead ends, and whose edges are the segments themselves. This yields 12,833 nodes and 14,856 edges. Corridors are then grown by walking the graph. Starting from an unclaimed edge, the walk extends in both directions, and at each node it examines the connected edges that no corridor has yet taken. A candidate is admissible only if it carries the same functional class as the corridor being built and if the angle it makes with the current edge is within 60 degrees of straight, so a candidate that turns more sharply is treated as a different street

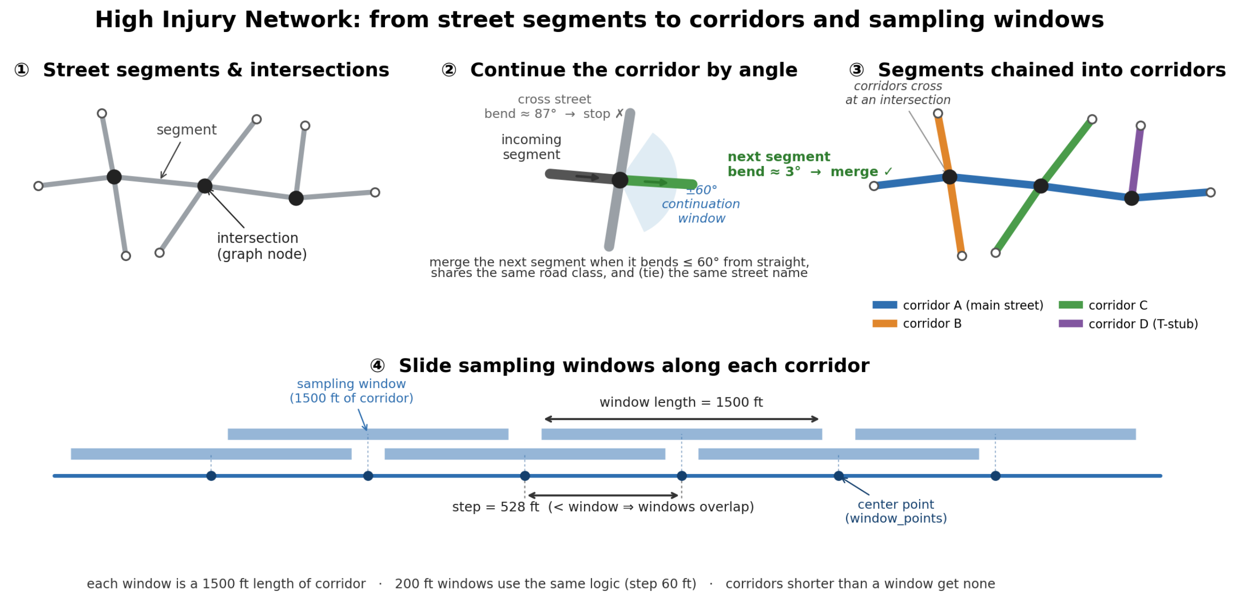

Figure 2: The corridor and its sampling windows, on a schematic network. Panel 1 shows the street segments and the intersections that break them. Panel 2 shows the merge test applied at an intersection. Panel 3 shows the resulting corridors, which cross rather than turn at a shared intersection. Panel 4 shows overlapping windows sliding along one corridor. The geometry is illustrative, not the Pittsburgh network.

rather than a continuation. Among admissible candidates the walk prefers one carrying the same street name, and settles any remaining tie by taking the straightest. It stops when no admissible candidate remains. Because a segment can belong to only one corridor, the outcome depends on the order in which seed edges are drawn; that order is randomized under a fixed seed so the construction is reproducible.

Three rules therefore define a corridor: continuity of alignment, constancy of functional class, and preference for a continuous name. The first two are what make the resulting object behave like a street rather than a routing path, since a corridor cannot turn a corner and cannot promote itself from a local street to an arterial partway along.

Sliding windows. The rule is not applied to whole corridors, because a corridor may run for a mile and a concentration of injury within one block of it should not be diluted by the length of the rest. Each corridor is instead sampled by windows of fixed length that slide along it. Two scales are used, matching the two limbs of the City rule set out in Section 3.2.3. The coarse scale is a window of 1,500 ft advanced in steps of 528 ft, one tenth of a mile, and is meant to capture corridor-length concentrations of injury of any severity. The fine scale is a window of 200 ft advanced in steps of 60 ft, and is meant to isolate the tight clusters of serious and fatal injury that a longer window would average away. Windows overlap heavily at both scales, so a cluster is not missed by falling across a boundary.

A window is an interval of arc length along the corridor, located by distance from the corridor's start, and events are counted by their own measured position along the same line, as described in Section 3.2.3. Windows begin half a window length from the corridor's start and continue while a full window still fits, so a corridor shorter than the window receives no window at that scale and cannot be flagged by that limb of the rule. This makes the two scales reach different amounts of the network, which Section 4.1 quantifies.

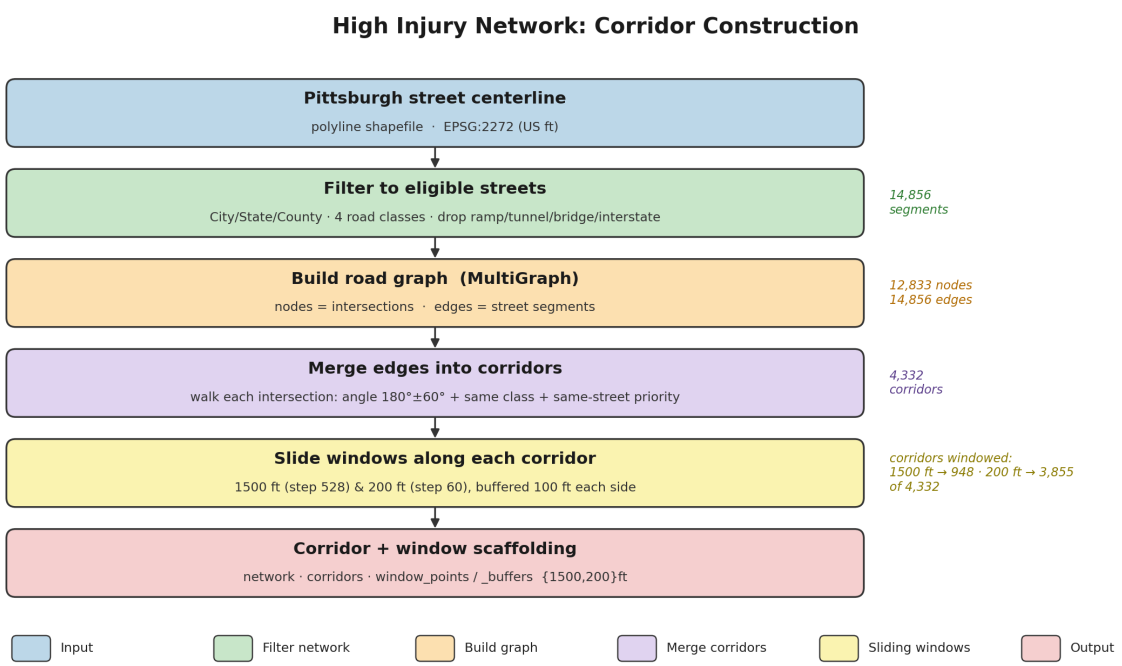

Figure 3: Construction of the corridor network and its sliding windows, from the filtered street centerline to the scaffolding on which the High Injury Network rule is evaluated. Counts at the right report the size of each intermediate product.

3.2.2 A KABCO-aligned severity index for EMS dispatches

The High Injury Network rule is defined on injury severity, so the two layers can be run through the same rule only if both carry severity on the same scale. Crash records do, through the KABCO code. The dispatch file does not, and no single field in it states how badly anyone was hurt. Severity is instead spread across each responding unit's disposition, a set of patient-handling fields recorded only when a patient is transported, and a free-text description of the scene. The index below combines those three sources. It uses only what the responders recorded, with no imputation from crash data, so the comparison in Section 3.2.3 stays independent, and where the evidence is ambiguous it takes the lower severity, so EMS counts understate rather than overstate injury. Figure 4 summarizes the construction.

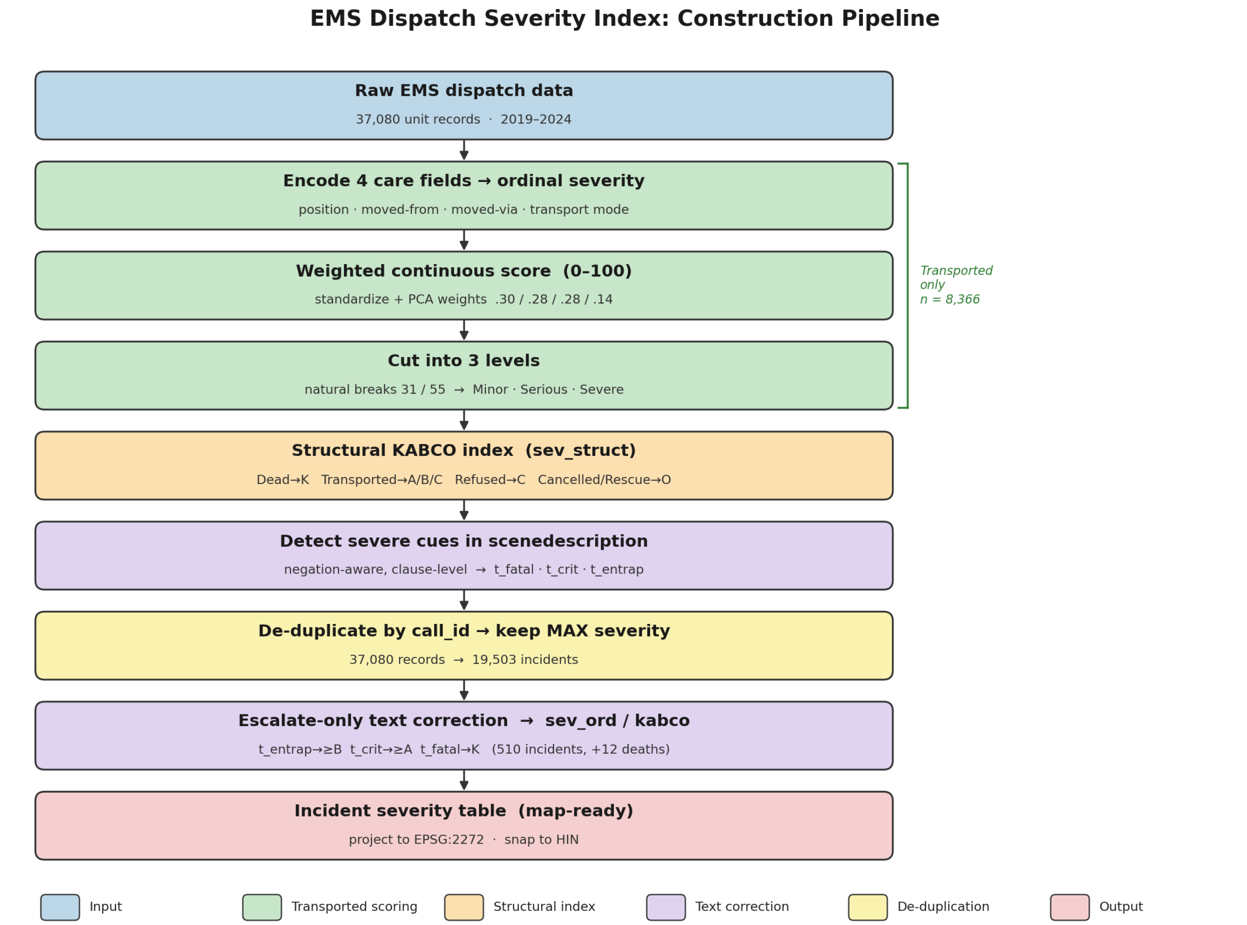

Figure 4: Construction of the EMS dispatch severity index, from raw unit records to a KABCO-coded incident table. The continuous score is defined on transported patients only; the disposition spine and the text correction apply to every record.

The structural index. Four coded fields carry clinical signal: the patient's position, how the patient was moved from the scene, the means by which the patient was moved to the ambulance, and the mode of transport to hospital. Each is encoded ordinally, with higher values indicating greater incapacitation, and administrative values such as "Not Known" are mapped to missing rather than to zero, so an unrecorded field does not read as an uninjured patient. These fields are populated only for transported patients, 8,366 of the 37,080 unit records. For those, the encoded fields are standardized and combined into a score running from 0 to 100, weighted by the normalized first principal component loadings from a factor analysis on the transported subset: 0.296, 0.284 and 0.280 for the three patient-handling fields, and 0.141 for transport urgency. Weights are renormalized per record over whichever fields are present, so a record missing one field is still scored, and 8,339 of the 8,366 receive a score. One-dimensional k-means cuts the score into three bands at 31.2 and 54.8, which are then nested into a spine given by the unit's disposition to produce the index in Table 2. Two categories require comment. *Cancelled* and *Rescue* units are disregarded or assisting units whose patient, if any, is counted on the transporting unit's record, so they carry no injury of their own. A transported patient too sparsely documented to score is placed in the middle band, which is both the modal and the conservative choice.

Table 2: Structural severity index: how each EMS record is placed on the KABCO scale.

KABCO	Meaning	Basis in the dispatch record
K	Fatal	Disposition *Dead*
A	Suspected serious injury	Transported, care score above 54.8
B	Suspected minor injury	Transported, care score between 31.2 and 54.8, or transported but unscored
C	Possible injury	Transported, care score below 31.2, or disposition *Refused*
O	No apparent injury	Disposition *Cancelled* or *Rescue*

A negation-aware text correction. Because the care fields exist only for transported patients, a serious outcome recorded under any other disposition is invisible to the structural index, and a fatality coded *Cancelled* because the transport was called off is the clearest case. The scene description carries that signal, but keyword matching would misread it, since severity words in this file appear more often negated than not: crews routinely record that there was no entrapment, or that a patient denies injury. The correction therefore reads each description clause by clause and accepts a cue only from a clause containing no negation token. Three narrow cue families apply: unambiguous indications of death raise the record to K, critical injury such as ejection or cardiac arrest raises it to at least A, and genuine entrapment or extrication raises it to at least B. The layer is escalate-only, never lowering a severity, and the structural value is retained alongside the corrected one so its effect stays auditable.

De-duplication to incident level. Because a traffic call draws more than one unit, records are grouped by call and represented by the unit with the highest structural severity, with the text cues pooled across every unit on the call and scene coordinates back-filled from any of them. The 37,080 unit records reduce to 19,503 incidents, of which 17,983 carry usable scene coordinates. Section 4.2 reports what the finished index yields.

3.2.3 A parallel EMS High Injury Network

The two layers now share a spatial frame from Section 3.2.1 and a severity scale from Section 3.2.2. What remains is to place events on corridors and apply the City's High Injury Network rule to each layer separately, producing a crash-derived network and an EMS-derived network that differ only in the events that built them.

Assigning events to corridors. Each crash and each EMS incident is assigned to exactly one corridor, so no event is counted twice and no corridor borrows a neighbour's events. Assignment follows the City's convention of attributing a crash at an intersection to the major street, and takes two paths. An intersection zone is placed at every network node joining more than two segments, with a radius scaled from the widths and functional classes of the roads that meet there; the 5,968 zones range from 28 to 113 ft in radius, with a median of 45 ft. An event falling inside a zone goes to the highest-class corridor meeting there, with ties broken by the higher traffic volume and then by proximity. An event outside every zone is a mid-block event and goes to the nearest corridor within 150 ft. Traffic volume enters only as that tie-break, taken as the highest annual average daily traffic recorded within 40 ft of the corridor, which is available for 2,212 of the 4,332 corridors. Each assigned event is then located by its position measured along the corridor. Section 4.3 reports how much of each layer the step retains.

The City rule. A corridor qualifies for the High Injury Network if either of two conditions holds anywhere along it, as set out in Table 3. The two limbs answer different questions. The coarse limb asks whether injuries accumulate along a stretch of street, and the fine limb asks whether the most severe outcomes cluster at a single point such as one intersection. A corridor is flagged if a single window at either scale passes, so the rule is a maximum over windows rather than an average, and one dangerous block is not diluted by a quiet mile.

Table 3: The two limbs of the City of Pittsburgh High Injury Network rule, applied identically to the crash layer and the EMS layer.

Limb	Window	Events counted	Threshold
Corridor scale	1,500 ft, step 528 ft	Fatal and injury, KABCO K, A, B and C	at least 5
Cluster scale	200 ft, step 60 ft	Serious and fatal, KABCO K and A	at least 2

Counting is measure-based. An event contributes to a window if its measured position along the corridor falls inside that window's interval, so the perpendicular tolerance is set once, by the 150 ft assignment radius, rather than a second time by the width of a buffer. Serious and fatal events count toward both limbs, since they are injuries as well. On the crash side the two severity groups are read directly from the KABCO code. On the EMS side they are read from the index built in Section 3.2.2, with severity O excluded, so a dispatch that found no apparent injury never contributes to either limb.

Two networks and their comparison. Running the rule over the assigned crashes yields the crash-HIN, and running the same rule over the assigned EMS incidents yields the EMS-HIN. Every corridor appearing in either is then placed in one of three classes. *Concordant* corridors clear the bar on both sources and are the highest-confidence priorities, since two independent reporting systems agree. *Crash-only* corridors clear it on reported crashes but not on EMS incidents. *EMS-only* corridors clear it on EMS incidents but not on reported crashes, and are the object of this study: places where the emergency response record shows a concentration of injury that the police-reported crash record does not.

One point of interpretation should be fixed before the results. The crash-HIN reported here is a reproduction of the City's rule on the corridor network of Section 3.2.1, not a copy of the City's published High Injury Network, which was built on the City's own segment geometry. The comparison in Section 4 is therefore internal and like-for-like: both networks come from the same corridors, the same assignment, the same window geometry and the same thresholds, and the only thing that differs between them is which events were fed in. That is what licenses reading a divergence as a property of the two reporting systems rather than an artefact of construction.

4 Results

4.1 The corridor network

Merging the 14,856 eligible centerline segments yields **4,332 corridors** spanning 933.5 miles. Coverage is complete and exclusive: every eligible segment belongs to exactly one corridor, none is assigned twice and none is dropped.

Corridors are short, and many of them are a single segment. The median runs 0.13 miles, about 685 ft, and the mean 0.22 miles, so the distribution is heavily right-skewed: a quarter of corridors are under 0.07 miles while the longest runs 4.8 miles. A corridor absorbs 3.43 segments on average and up to 56, but 1,705 corridors, 39.4 percent of the total, are a single segment. A short residential block that meets nothing continuing straight, in the same functional class, at either end has nothing to merge with, and most of Pittsburgh's local street grid is of that kind. Figure 5 maps the result.

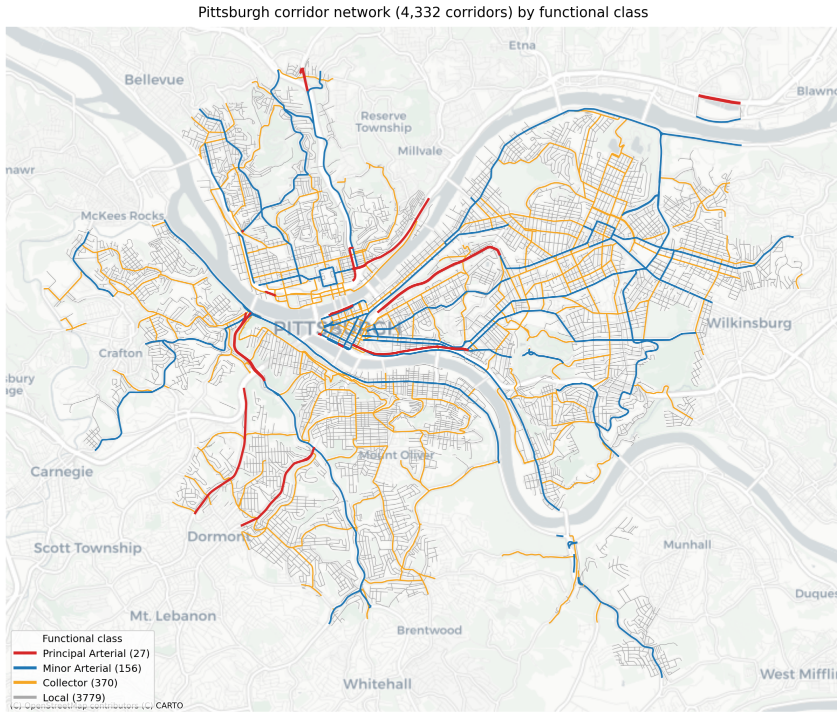

Figure 5: The 4,332 corridors, coloured by functional class. Counts in the legend are corridors, not segments, so they differ from the segment counts reported in Section 3.1: merging chains many local segments into one corridor, while an arterial often survives as a single long corridor.

Corridor length decides which limb of the rule can reach a corridor. Because a corridor shorter than a window receives no window at that scale, the two limbs of the rule see different parts of the network. The 200 ft scale places 70,206 windows on 3,855 corridors, 89.0 percent of them, covering 921.7 of the 933.5 miles. The 1,500 ft scale places 3,308 windows on only 948 corridors, 21.9 percent. That sounds severe, but those 948 corridors carry **549.7 miles, 58.9 percent of the network**, because length is exactly what qualifies a corridor for the coarse scale.

4.2 Severity of EMS traffic dispatches

Applied to the 14,298 incidents in the study area, the index places **6,529 of them, 45.7 percent, at some level of injury**, with 1,203 at the serious-or-fatal end. The remainder are dispatches that found no apparent injury, which is what one expects of a service that responds to a report rather than to a confirmed casualty. Table 4 sets the profile against the crash record for the same place and period, and Figure 6 shows how the incidents fall across the city.

Table 4: **Severity profile of the two event sets, study area, 2020–2024.**

	Police-reported crashes		EMS dispatches	
Severity	**Events**	**%**	**Incidents**	**%**
K fatal	103	0.6	46	0.3
A suspected serious	571	3.1	1,157	8.1
B suspected minor	3,600	19.7	3,220	22.5
C possible	1,155	6.3	2,106	14.7
Injury subtotal (K–C)	**5,429**	**29.7**	**6,529**	**45.7**
O no apparent injury	9,944	54.4	7,750	54.2
Unknown	2,894	15.8	19	0.1
Total	**18,267**	**100.0**	**14,298**	**100.0**

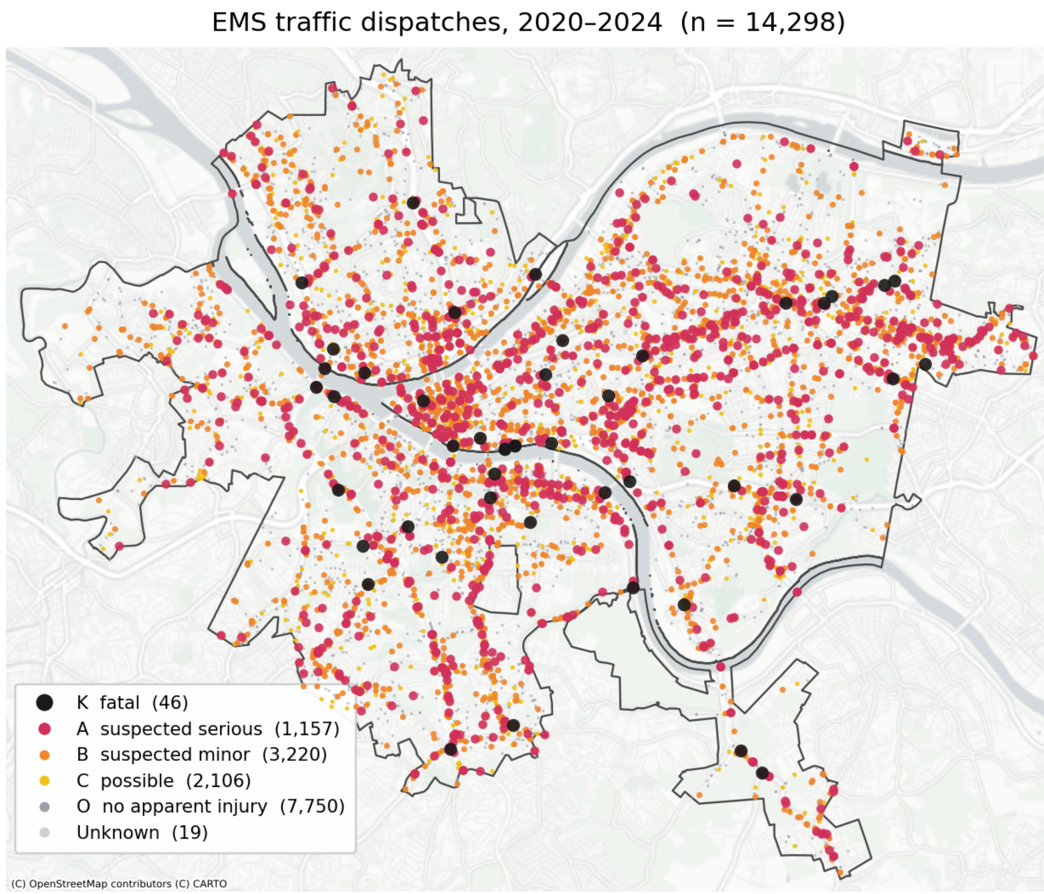

Figure 6: EMS traffic dispatches by KABCO severity, 2020–2024, over the study area. Severity comes from the index built in Section 3.2.2. Drawn on the same scale and palette as Figure 1 so the two are directly comparable.

EMS records fewer events but more injuries. The comparison in Table 4 is the central descriptive result of this report. EMS attends *fewer* traffic events than the police record, 14,298 against 18,267, yet it records **more injuries in absolute terms**, 6,529 against 5,429, a ratio of 1.20. At the serious end the gap widens: 1,203 EMS incidents reach K or A against 674 crashes, a ratio of 1.78. The share is starker still, since injury accounts for 45.7 percent of EMS incidents but only 29.7 percent of crashes. The two systems are not sampling the same thing. The police record is built around the reportable crash, which includes a large body of property-damage events, while EMS is dispatched to a suspected casualty and its no-injury category is the residue of calls that turned out not to need one.

Two qualifications keep this from being read as a simple accuracy ranking. The A categories are not the same instrument: on the crash side A is an officer's judgment of suspected serious injury at the scene, while on the EMS side it is a care score above an empirical threshold, so the 1.78 ratio measures the two definitions as much as the underlying injuries. And the fatality counts run the other way, **46 EMS against 103 crash**, because a traffic death is often pronounced at hospital or at the scene by a unit that never transports, and the crash record counts deaths within 30 days of the collision, which reaches outcomes no dispatch record can see. Neither source dominates the other. Each sees a part of the injury burden the other misses, which is the premise the network comparison in Section 4.3 tests spatially.

4.3 The two networks compared

Roughly nine in ten events in each layer reach a corridor. The two layers behave almost identically under the assignment step, which is itself reassuring. Of the 18,267 crashes, 16,174 are assigned, 8,021 by the intersection rule and 8,153 as mid-block. Of the 14,298 EMS incidents, 12,664 are assigned, 5,383 by the intersection rule and 7,281 as mid-block. The unassigned remainder, 11.5 percent of crashes and 11.4 percent of EMS incidents, lies more than 150 ft from any eligible corridor. These are events on parking lots, private property, and the ramps, bridges and limited-access facilities excluded in Section 3.1, together with ordinary geocoding error.

One rule applied to two sources produces two networks of different size. Running the City rule over the assigned crashes flags **128 corridors**, covering 151.0 miles. Running the identical rule over the assigned EMS incidents flags **164**, covering 174.4 miles. Between them the two networks name 192 corridors and 188.1 miles, 20.1 percent of the 933.5-mile network, and they agree on 100 of those corridors. Table 5 gives the classification and Figure 7 maps it.

Table 5: Corridor classification, crash-HIN against EMS-HIN.

Status	Corridors	%	Miles	% of miles	Median length (mi)
Concordant (both sources)	100	52.1	137.3	73.0	1.16
Crash-only	28	14.6	13.7	7.3	0.51
EMS-only	64	33.3	37.1	19.7	0.44
Total flagged	**192**	**100.0**	**188.1**	**100.0**	**0.67**

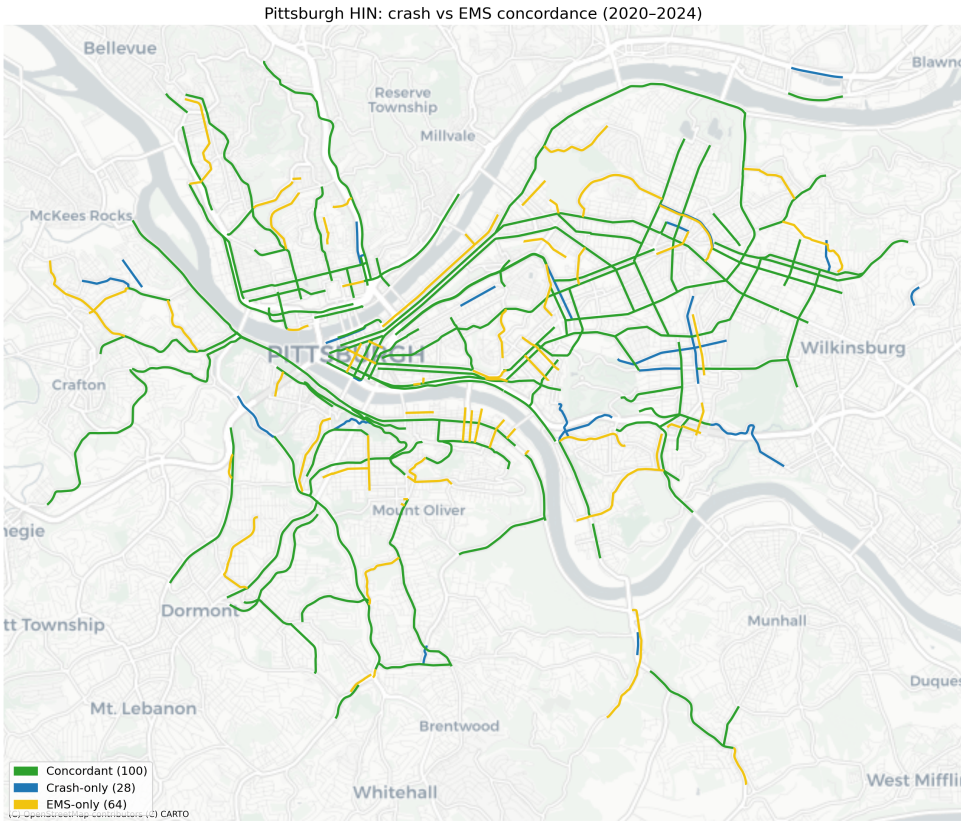

Figure 7: Concordance between the crash-derived and EMS-derived High Injury Networks. Green corridors are flagged by both sources, blue by reported crashes only, and yellow by EMS incidents only.

The agreement is real but asymmetric. The two sources agree on the busiest part of the network. The 100 concordant corridors are only half the flagged corridors but **73.0 percent of the flagged mileage**, with a median length of 1.16 miles against 0.44 and 0.51 for the two disagreeing classes. These are the long arterial and collector corridors, and they are the highest-confidence priorities in the whole exercise, because two reporting systems built on different instruments, by different agencies, for different purposes, independently identify them.

The disagreement is not symmetric. **78.1 percent of the crash-HIN also appears in**

the EMS-HIN, but only 61.0 percent of the EMS-HIN appears in the crash-HIN. The EMS network is both larger and more nearly a superset: it recovers most of what the crash record finds and adds a substantial set of its own. That is the pattern one would expect if the crash record were missing injury events rather than the EMS record inventing them, though it is consistent with other explanations considered below.

EMS-only corridors are local streets. The single clearest finding of the comparison is where the two sources part company. Table 6 breaks each class of corridor down by functional class, and the pattern is stark. Concordant corridors are arterial and collector: 87 of 100 are Minor Arterial or Collector, and only 6 are Local. **EMS-only corridors invert this: 31 of 64 are Local and a further 26 are Collector, leaving just 7 on the arterial network.**

Table 6: Functional class composition of each corridor class.

Functional class	Concordant	Crash-only	EMS-only
Principal Arterial	7	3	1
Minor Arterial	41	8	6
Collector	46	11	26
Local	6	6	31
Total	**100**	**28**	**64**

The two records therefore agree about arterials and disagree about neighbourhood streets. This is coherent with how the two systems are triggered. A collision on an arterial is likely to involve a tow, block a lane and draw a police response, so it enters the crash record almost by default. An injury on a local street where the vehicles remain driveable may never generate a police report at all, because Pennsylvania's reporting threshold turns on injury, fatality or a vehicle requiring towing, and the judgment is made at the scene. An ambulance, by contrast, is dispatched on the report of a casualty regardless of what happens to the vehicles. The corridors where the two sources diverge are exactly the corridors where that difference in trigger would bite hardest.

Both limbs of the rule flag the EMS-only corridors. Of the 64, **31 were flagged by the 1,500 ft all-injury limb alone**, 25 by the 200 ft serious-or-fatal limb alone, and 8 by both. That the fine limb accounts for more than a third on its own matters, because it means these are not merely corridors accumulating minor incidents: 33 of the 64 hold at least two serious-or-fatal EMS incidents within a single 200 ft window. The concentration can be considerable. The strongest candidate, a 0.6-mile local corridor, carries 47 injury incidents in one 1,500 ft window and 6 serious-or-fatal in one 200 ft window while never clearing the crash threshold. Across the 64 corridors, EMS records more events than the crash record does, 18.9 against 14.9 per corridor on average, exactly reversing the ratio on Crash-only corridors, where the figures are 7.9 EMS against 23.5 crashes.

Crash-only corridors point to gaps in EMS involvement, not to crash over-reporting. The 28 crash-only corridors are the smallest class by both count and mileage, 13.7 miles against the EMS-only 37.1. Their existence shows the divergence runs in both directions and that the EMS layer is not simply a superset produced by a looser threshold. Their most likely reading is EMS coverage rather than crash over-reporting, since a corridor can accumulate reportable crashes whose injuries were minor enough that no ambulance was dispatched, or where transport was declined. They average 23.5 crashes against 7.9 EMS incidents, which is consistent with that account.

5 Conclusion

In this report, we asked whether a High Injury Network built from emergency medical dispatch records selects the same corridors as one built from police-reported crashes, and we answered that question without linking the two records. We built a single corridor geometry from the City street centerline, merging 14,856 eligible segments into 4,332 corridors over 933.5 miles, and we constructed a KABCO-aligned severity index for the dispatch file so that we could grade both layers on one scale. We assigned each of the 18,267 crashes and 14,298 EMS incidents recorded between 2020 and 2024 to a single corridor, and we then applied the City of Pittsburgh rule identically to each layer: we flag a corridor if any 1,500 ft window holds at least five fatal-and-injury events, or any 200 ft window at least two serious-or-fatal events. Because the procedure requires no personal identifiers and no record-level match, it is portable to any agency that holds dispatch locations without holding a data-sharing agreement.

The two layers select different networks. The crash record flags 128 corridors covering 151.0 miles and the EMS record 164 covering 174.4 miles, and between them they name 192 corridors, a fifth of the network's mileage, agreeing on 100. That agreement is concentrated where the burden is: the concordant corridors are half the flagged count but 73.0 percent of the flagged mileage, and they form the arterial and collector spine that both reporting systems identify independently. The divergence is concentrated on neighbourhood streets. Of the 64 corridors flagged by EMS alone, 31 are Local and 26 Collector, against 6 Local among the 100 concordant corridors, which is the pattern a reporting threshold turning on injury, fatality or a vehicle requiring towing would produce, since an ambulance is dispatched on the report of a casualty whatever happens to the vehicles. The 28 crash-only corridors show that the divergence runs in both directions and that the EMS layer is not simply a looser threshold.

One limitation qualifies these findings. We compare the two sources at the level of the network rather than the record, so an EMS-only corridor marks a place where the dispatch record shows a concentration of injury that the crash record does not, and it does not establish that any particular ambulance run is missing from the crash file. Confirming that would require the personal identifiers this study does not hold. The EMS severity scale is likewise a constructed proxy. We infer it from disposition, transport mode and patient-handling fields, escalate it from cued clauses in the scene text and never lower it, so its categories carry the definitions of the fields that built them rather than a clinical assessment, and the layer cannot see the deaths the crash record counts within 30 days of the collision.

Acknowledgments

The authors gratefully acknowledge Panini Chowdhury from the City of Pittsburgh for his valuable advice and for providing a part of the data sets used in this study. The authors also thank the City of Pittsburgh for its support and assistance. The views and opinions expressed in this work are solely those of the authors and do not necessarily represent the views, opinions, or official position of the City of Pittsburgh.

References

Ceklic, E., Tohira, H., Ball, S., Brown, E., Brink, D., Bailey, P., Whiteside, A., & Finn, J. (2020). Motor vehicle crash characteristics that are predictive of high acuity patients: An analysis of linked ambulance and crash data [Advance online publication]. *Prehospital Emergency Care*. https://doi.org/10.1080/10903127.2020.1771487

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., & Buchanan, B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, *34*(5), 301–310. https://doi.org/10.1006/jbin.2001.1029

City of Pittsburgh. (n.d.). *High Injury Network (HIN) development methodology*. Department of Mobility and Infrastructure, City of Pittsburgh, PA.

Claros, B., Bill, A., & Noyce, D. A. (2022). High injury network: City of Madison, Wisconsin. In *International Conference on Transportation and Development 2022* (pp. 61–69). American Society of Civil Engineers. https://doi.org/10.1061/9780784484333.006

Federal Highway Administration. (2024). *A preview of states' Vulnerable Road User (VRU) safety assessments* (Report No. FHWA-SA-24-034). U.S. Department of Transportation.

Hosseinzadeh, A., Karimpour, A., Kluger, R., & Orthober, R. (2022). Data linkage for crash outcome assessment: Linking police-reported crashes, emergency response data, and trauma registry records. *Journal of Safety Research*, *81*, 21–35. https://doi.org/10.1016/j.jsr.2022.01.003

Huang, C.-Y., Rau, C.-S., Chuang, J.-F., Kuo, P.-J., Hsu, S.-Y., Chen, Y.-C., Hsieh, H.-Y., & Hsieh, C.-H. (2016). Characteristics and outcomes of patients injured in road traffic crashes and transported by emergency medical services. *International Journal of Environmental Research and Public Health*, *13*(2), 236. https://doi.org/10.3390/ijerph13020236

National Highway Traffic Safety Administration. (2015). *Crash Outcome Data Evaluation System (CODES): Program* (Report No. DOT HS 812 178). U.S. Department of Transportation. https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812178

NEMSIS Technical Assistance Center. (n.d.). *NEMSIS: The National EMS Information System*. https://nemsis.org/

Nijman, S. (2025). OA2047. Medical data of road traffic victims, what do they teach us? *European Journal of Public Health*, *35*(Supplement 5), ckaf165.010. https://doi.org/10.1093/eurpub/ckaf165.010

Oregon Department of Transportation. (2023). *Oregon vulnerable road users safety assessment*. Oregon Department of Transportation.

Pande, A., Majumdar, A., & Kantorowski, L. (2026). *Equitable estimation of accurate high injury networks (HINs) for vulnerable road users*. Mineta Transportation Institute. https://doi.org/10.31979/mti.2025.2459

75 Pa.C.S. § 3746, Immediate notice of accident to police department (n.d.). https://www.legis.state.pa.us/WU01/LI/LI/CT/HTM/75/00.037.046.000..HTM

San Francisco Department of Public Health. (2022). *Vision Zero high injury network: 2022 update. A methodology for San Francisco, California*. Center for Data Science, San Francisco Department of Public Health.

Sun, C., Brown, H., Edara, P., Stilley, J., & Reneker, J. (2023). *Vulnerable road user (VRU) safety assessment* (Report No. cmr 23-015). Missouri Department of Transportation.