CHALMERS
UNIVERSITY OF TECHNOLOGY

Harmonizing Insurance Crash Data with the IGLAD Common Data Scheme: Reconstruction and Digitization

A Weighted Compatibility Rating System for Assessing Crash Data

Master's thesis in Mobility Engineering, MPMOB

Zichen Huang

DEPARTMENT OF MECHANICS AND MARITIME SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's Thesis in Mobility Engineering

Harmonizing Insurance Crash Data with the IGLAD Common Data Scheme: Reconstruction and Digitization

A Weighted Compatibility Rating System for Assessing Crash Data

ZICHEN HUANG

Department of Mechanics and Maritime Sciences
Division of Vehicle Safety
Chalmers University of Technology
Gothenburg, Sweden 2026

Harmonizing Insurance Crash Data with the IGLAD Common Data Scheme: Reconstruction and Digitization
A Weighted Compatibility Rating System for Assessing Crash Data
ZICHEN HUANG

Supervisor: Jordanka Kovaceva, Department of Mechanics and Maritime Sciences, Vehicle Safety.
Examiner: Robert Thomson, Department of Mechanics and Maritime Sciences, Vehicle Safety.

Master's Thesis 2026
Department of Mechanics and Maritime Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46 31 772 1000

This report was written with the assistance of LLM-powered AI assistants, including grammar correction, language polishing, image generation, and typesetting.

Typeset in LaTeX
Gothenburg, Sweden 2026

Harmonizing Insurance Crash Data with the IGLAD Common Data Scheme: Reconstruction and Digitization
A Weighted Compatibility Rating System for Assessing Crash Data
ZICHEN HUANG
Department of Mechanics and Maritime Sciences
Division of Vehicle Safety
Chalmers University of Technology

Abstract

Road traffic accidents involving vulnerable road users (VRUs) remain a significant global safety concern, and the in-depth accident database Initiative for the Global Harmonization of Accident Data (IGLAD) plays an important role in advancing safety research. Insurance companies hold large volumes of crash data that could potentially enrich such databases, and effective methods for evaluating their harmonization readiness are of practical value. Utilizing the People Around the Vehicle (PAV) crash database from If P&C Insurance as an example, this project investigates the feasibility of harmonizing car-to-VRU-oriented insurance crash data with the IGLAD common data scheme. Variable-level mapping and structural harmonization strategies are explored, and the Weighted Compatibility Rating System (WCRS) is proposed to quantitatively evaluate reconstruction relevant variables at both the dataset and individual case level. Applied to PAV and existing IGLAD member datasets, the results reveal gaps in multiple variable aspects. Case-level assessments further show that reconstruction feasibility of the given case samples is constrained by the absence of measured scene geometry, objective velocity references, and VRU anthropometric data. Accordingly, targeted recommendations for data collection practice are proposed. The tools developed are designed for broader applicability and are expected to support the evaluation of future datasets contributing to IGLAD.

Keywords: Vulnerable Road Users, traffic safety, insurance crash data, accident data harmonization.

Acknowledgements

First and foremost, I would like to express sincere gratitude to my supervisor Jordanka Kovaceva for continuous guidance, constructive feedback, and unwavering support throughout the entire project. Her expertise and encouragement have played a crucial role in the direction and quality of this work.

I am also greatly grateful to my examiner Robert Thomson, for his valuable insights and oversight throughout the project.

Furthermore, sincere appreciation is expressed to Magdalena Lindman and Sofia Jonsson from If P&C Insurance for their indispensable contributions. Their willingness to share domain expertise, provide data access, and engage in thoughtful discussions enriched the depth and relevance of this work. Also, I would like to thank Jörg Bakker from Asymptotic AI for generously sharing understanding and expertise.

Special thanks are extended to all members of IGLAD who participated in and contributed to this project.

Finally, I wish to thank all those who have provided any form of assistance to this project and supported me throughout this meaningful journey!

Zichen Huang, Gothenburg, June 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AAAM	Association for the Advancement of Automotive Medicine
ABS	Anti-lock Braking System
ACEA	European Automobile Manufacturers' Association
ACT	Australian Capital Territory
ADAS	Advanced Driver Assistance System
AEB	Autonomous Emergency Braking
AIS	Abbreviated Injury Scale
BASt	Federal Highway and Transport Research Institute
CAD	Computer-Aided Design
CAN	Controller Area Network
CATARC	China Automotive Technology and Research Center
CDC	Collision Deformation Classification
CDS	Crashworthiness Data System
CIDAS	China In-depth Accident Study
CISS	Crash Investigation Sampling System
DOT	Department of Transportation
EDR	Event Data Recorder
EES	Energy Equivalent Speed
ESC	Electronic Stability Control
EU	European Union
Euro NCAP	European New Car Assessment Program
FE	Finite Element
FAT	Forschungsvereinigung Automobiltechnik e.V.
GIDAS	German In-Depth Accident Study
GPS	Global Positioning System
HBM	Human Body Model
IGLAD	Initiative for the Global Harmonization of Accident Data
IIHS-HLDI	Insurance Institute for Highway Safety and Highway Loss Data Institute
KIDAS	Korea In-Depth Accident Study
KPI	Key Performance Indicator

LDW	Lane Departure Warning
MAIS	Maximum Abbreviated Injury Scale
MBS	Multi-body System
NASS	National Automotive Sampling System
NHTSA	National Highway Traffic Safety Administration
NRSS	National Roadway Safety Strategy
OECD	Organization for Economic Co-operation and Development
PAV	If's People Around the Vehicle Crash Database
PCM	Pre-Crash Matrix
PDOF	Principal Direction of Force
POI	Point of Impact
PRT	Perception–Reaction Time
PROSPECT	PROactive Safety for PEdestrians and CyclisTs
PSU	Primary Sampling Unit
SAE	SAE International
SCP	Straight Crossing Path
TDC	Truck Deformation Classification
TTC	Time-to-Collision
UDB	German Insurers' Accident Database
US	United States
V-A	Vehicle-to-Active traveler
V-V	Vehicle-to-Vehicle
VRU	Vulnerable Road User
WCRS	Weighted Compatibility Rating System
WHO	World Health Organization

Nomenclature

Below is the nomenclature of indices, sets, parameters, and variables that have been used throughout this thesis.

Indices

v	Index for reconstruction-relevant variable
s	Index for data source
c	Index for variable category
t	Index for variable type

Sets

$\{v \in c\}_{\text{in-scope}}$	Set of in-scope variables within category c
$\{c\}_{\text{active}}$	Set of active categories included in composite scoring

Parameters

W_v	Importance weight of variable v
W_c	Weight of category c in composite score
α	Secondary-source bonus coefficient ($\alpha = 0.10$)
$R(s, t)$	Reliability weight of source s for variable type t

Variables

$P(v, s)$	Binary presence indicator for variable v from source s
$\text{type}(v)$	Variable type mapping for variable v
$\text{eff}(v, s)$	Effective contribution of source s to variable v

best_eff(v)	Highest effective contribution among all sources for variable v
second_eff(v)	Second-highest effective contribution among all sources for variable v
combined_eff(v)	Combined effective contribution after secondary-source bonus
var_score(v)	Variable score of v
var_max(v)	Maximum achievable score for variable v
cat_score(c)	Category score of category c
composite	Composite WCRS score
contrib(s, c)	Source-wise contribution of source s to category c
min_cat_score	Minimum category score across active categories
tier_S_missing	Count of uncovered Tier-S variables in scope
tier_E_missing	Count of uncovered Tier-E variables in scope
μ	Road surface friction coefficient

Contents

List of Figures

List of Tables

1

Introduction

1.1 Overview of road traffic safety

Road traffic safety has always been a serious concern, and is also one of the important public health issues worldwide. The World Health Organization (WHO) pointed out that in 2021, approximately 1.19 million people died in road traffic accidents, equivalent to about 15 deaths per 100,000 people [1]. As of 2019, road traffic injuries remained the leading cause of death for children and adolescents aged 5-29 [1]. Road traffic accidents not only seriously affect the victims and their families, but also impose economic and governance burdens on society.

The process of improving road traffic safety has not always been consistent and smooth. In the United States, the situation regarding road deaths and death rates, after stagnating for a decade, had even moved in the wrong direction between 2020 and 2021 [2, 3]. In 2021, road traffic accidents caused more than 42,000 deaths, an increase of over 10 % compared to 2020 [3]. In the EU, the number of deaths from other modes of transport has decreased largely, while in contrast, the number of fatalities in bicycle accidents increased by 2 % between 2013 and 2023 instead of declining; in 2022, approximately 1,950 cyclists died in traffic accidents [4].

Improving road traffic safety has become an important goal for governments and international organizations worldwide. Many countries have proposed more forward-looking traffic safety goals. The sustainable safety vision proposed in early 1990s in the Netherlands has guided the country's road safety improvement approach for decades [5]. In 1997, Sweden passed the Vision Zero concept through its parliament, aiming to ultimately reduce road traffic deaths to zero through systematic safety design [6]. First released by the United States Department of Transportation (U.S. DOT) in January 2022, the National Roadway Safety Strategy (NRSS) [7] introduces the Department's comprehensive approach to significantly reducing serious injuries and fatalities on the nation's highways, roads, and streets. Faced with an unacceptable crisis, it aims to set a vision and goals for road safety in the United States and represents the first step toward achieving the ultimate goal of zero road deaths [7].

Specifically, using 2019 as the baseline, the EU's current road safety objectives are to reduce both road fatalities and serious injuries by 50 % by 2030, and to bring both close to zero by 2050 [8]. The U.S. DOT also set an ambitious medium-term goal in its US Department of Transportation Fiscal Years 2022-2026 Strategic Plan to reduce motor vehicle-related deaths by 66 % by 2040 [3].

1.2 Overview of VRU

While global road safety is gradually improving overall, these benefits are not shared equally by all road users [9]. In fact, vulnerable road users (VRUs) account for a disproportionate share of road traffic injuries and fatalities, representing around half of all victims [10]. As Organization for Economic Co-operation and Development (OECD) [11] outlined in 1998, the term “VRU” refers to the group of people at the highest risk in road traffic, mainly those do not have external protective shields. Pedestrians, pedal cyclists, and motorcyclists are therefore regarded as vulnerable, as they have little or no external protection to absorb impact energy in the event of a collision [12]. Among VRU groups in road transportation, some are more vulnerable than others considering in terms of demographic characteristics and capabilities: the three most common groups are children, people with disabilities, and the elderly, especially wheelchair users [11, 13]. In recent years, the concept of micromobility has gradually been proposed, and the categories of VRUs have become more diverse. In 2021, users of personal micromobility devices, such as e-scooters, accounted for 3 % of global road traffic fatalities, respectively [14].

Since the concept of unprotected road users, the precursor to the VRU, first appeared in research in the 1950s [15], VRU safety has become a concern and an important topic of research in road safety. For decades, researchers have conducted studies on VRUs in various fields. Du et al. [16] analyzed pedestrian and cyclist crashes in the Australian Capital Territory (ACT), Australia using an extended dataset and applying negative binomial and binary logit models, identifying road characteristics, infrastructure, and land use as significant factors influencing crash frequency and severity. Yuan et al. [17] employed a knowledge mapping approach to review VRU collision avoidance research across perception, behavior cognition, and decision-making, highlighting the adoption of deep learning. Singh et al. [18] conducted a systematic review of multiple studies on e-scooter injuries, identifying falls as the primary injury mechanism and upper limb fractures and head injuries as the most common outcomes, with low helmet use highlighting the need for improved safety measures and regulation. Based on data from the Guangzhou Public Security Traffic Management Integrated System, Zhou et al. [19] used linear regression analysis to quantitatively analyze the occurrence characteristics and severity influencing factors of e-bike related accidents, providing empirical evidence for the formulation of e-bike safety policies in the context of Chinese cities.

“Active travel” is a term frequently used by researchers and policymakers in recent years, which has a wider definition of: a way of travel in which travelers directly propel themselves forward through sustained physical activity [20]. Although exploration into broadening the concept of active travelers has been ongoing, in the field of traffic safety, active travelers generally refer to pedestrians and cyclists. It is essential to distinguish between the concepts of active traveler and VRU. As discussed above, VRU includes pedestrians, various non-motor vehicle drivers or riders, and motorcyclists among motor vehicle drivers. All active travelers belong to VRU, and the terms VRU and active traveler are used interchangeably in some research contexts.

This project will be based on VRU accidents in road traffic. Specifically, analytical work and in-depth studies will use accident cases involving motor vehicles and active travelers and e-scooters.

1.3 Overview of accident data levels

Road accident data can be classified according to the spatial scale at which crashes are aggregated and analyzed. The prevailing level framework in the safety literature distinguishes between macroscopic and microscopic levels, with a mesoscopic level increasingly recognized as a distinct analytical tier [21, 22].

At the macroscopic level, crashes are grouped into spatial units such as traffic analysis zones or counties, with analysis centered on zone-level variables including roadway characteristics, traffic demand, and sociodemographic factors to inform planning-level interventions [21, 22]. At the microscopic level, individual segments or intersections serve as the unit of analysis, where geometric design, traffic volume, and roadway characteristics are examined to support site-specific engineering purposes [21, 22]. A third tier, the mesoscopic level, integrates intersections and adjacent roadway segments into a single analytical unit, offering a methodological bridge between the two established levels [22].

These three levels are defined by spatial aggregation, not by the investigative depth of individual crash records. In-depth accident data occupy a distinct dimension. As Ljung Aust [23] makes explicit, official police-reported databases are statistically representative but contain limited causal information, whereas in-depth case studies, conducted on-scene by multidisciplinary teams through driver interviews and vehicle kinematics reconstruction, yield detailed causal information but are rarely representative of the broader crash population. In-depth accident data therefore do not map onto any single spatial tier; rather, it addresses the causal dimension that statistical databases at all three levels lack.

1.4 Overview of in-depth accident data

In-depth accident data plays a key role in the development and evaluation of vehicle safety systems in traffic safety research. It can be used not only to analyze individual crash cases, but also to provide detailed information on accident types, offering a valuable understanding of accident mechanisms [24]. In-depth accident data collected from different sources in different countries provides detailed information about individual cases, their environment, participants, collisions, injuries, and safety systems [25].

Globally, researches on in-depth accident data have evolved various representative datasets for their specific traffic environments, and representative examples among them can be the databases from German, Chinese and American research organizations.

1.4.1 In-depth database from different countries

In Germany, building on early work began in Hanover in 1973, German In-Depth Accident Study (GIDAS), established in 1999, is Europe's leading representative and one of the world's largest in-depth traffic accident research projects [26, 27]. GIDAS is jointly operated by the Federal Highway Research Institute (BASt) and Forschungsvereinigung Automobiltechnik e.V. (FAT) to conduct in-depth investigations of traffic accidents from a purely scientific perspective, with the aim of achieving a sustainable improvement in road safety [27].

In China, the China Automotive Technology and Research Center (CATARC) and other automotive enterprises have been conducting the China In-depth Accident Study (CIDAS) since 2011, providing foundational data for formulating and revising relevant policies and standards, to optimize the revision of the China New Car Assessment Program (C-NCAP), and to enhance and build crash testing capabilities [28].

In the US, National Highway Traffic Safety Administration (NHTSA) established the National Automotive Sampling System (NASS) in the 1970s to support motor vehicle safety research and policy development [29]. For nearly four decades, the Crashworthiness Data System (CDS)—one of two NASS subsystems—collected in-depth data from passenger vehicle crashes through trained investigation teams conducting on-site scene documentation, vehicle inspections, victim interviews, and medical record reviews [30, 31]. However, the sampling frame and primary data collection sites had remained essentially unchanged for over 30 years, becoming increasingly misaligned with contemporary safety priorities [30]. Congress authorized a comprehensive modernization effort beginning in 2012, resulting in the Crash Investigation Sampling System (CISS), which replaced CDS in 2016 [29]. CISS retains a three-stage stratified sampling structure—selecting Primary Sampling Units (PSUs), police jurisdictions, and police accident reports—with updated sampling frames and selection procedures [30]. PSUs are geographically bounded to limit technician travel time, the multi-phase design accommodates budget fluctuations without complete reselection, and Pareto sampling at subsequent stages manages jurisdictional frame changes while minimizing sample disruption [30, 31]. Following each crash selection, trained technicians document scene evidence, inspect vehicles, and conduct confidential victim interviews alongside medical record reviews [32]. Statistical weighting procedures accounting for unequal selection probabilities, non-response adjustment, post-stratification adjustment, and PSU weight calibration generate nationally representative population estimates from approximately 4,000 annual investigations conducted across 32 data collection sites as of 2023 [30, 33].

1.4.2 IGLAD

The various in-depth traffic accident data formats in different studies around the world make cross-dataset comparisons challenging. The differences and gaps exist in aspects such as variable definitions, data structures, and data quality, which to some extent obstruct the merging of data and comparative researches. Despite the difficulties, multiple studies have explored methods for harmonizing different accident databases. For instance, the EU project PROactive Safety for PEdestrians and CyclisTs (PROSPECT) [34] established a standardized accident scenario classification system by analyzing accident databases from multiple European countries. Initiative for the Global Harmonization of Accident Data (IGLAD) is an international effort which provides a harmonized global in-depth accident database. Using a unified data scheme, it aims to consolidate in-depth accident data from multiple sources, establishing a globally comparable standardized data framework [35]. Initiated jointly by Mercedes-Benz AG, the European Automobile Manufacturers' Association (ACEA) and several research institutions in 2010, the consortium now comprises 24 members, and continues to seek new crash data and new members [35]. According to the latest (Date 2026-02-25) IGLAD codebook [36], 14 organizations from 12 different countries have provided IGLAD with data from original databases, including the examples above.

Various studies in the field of traffic safety have been conducted based on IGLAD. Liers and Pett [37] used IGLAD data for cross-regional crash comparisons, showing its capability to analyze scenarios, injury severity, and reconstruction parameters, and linking higher speeds and VRU involvement in Asia to greater crash severity. In accidents related to VRU, Elawad et al. [38] applied IGLAD crash data to validate safety envelope models for bus–VRU interactions, showing that minimum safe distance–based critical sets correspond well to real crash conditions and can support collision avoidance evaluation. Furthermore, IGLAD also provides inspiration and reference when establishing other new accident datasets in the world. The experience gained from establishing Korea In-Depth Accident Study (KIDAS) database [39] demonstrates that even with limited information on accidents, adopting the IGLAD standard format and integrating multiple data sources (e.g. vehicle crash worthiness test data, crash reconstruction simulations, and insurance company data) enables the creation of a comprehensive accident database compliant with international standards.

1.5 Overview of insurance crash data

Insurance companies possess a large amount of crash data and play an important role in road traffic safety research. Insurance crash data has certain advantages in many aspects. Data collected during the processes of insurance policies and claims can be utilized to formulate strategies aimed at enhancing traffic safety [40]. In accident cases between VRUs and passenger cars, insurance data can provide more information that may not be available in police or hospital databases, for instance, Isaksson-Hellman and Werneke's study analyzed 882 bicycle-car collisions from a dataset by If P&C Insurance, of which only half appeared in the official police-reported data; the dataset provides the possibility to study more accidents of varying severity [41].

Insurance accident data has been widely applied in various directions of traffic safety research, including: the analysis of crash and injury trends, such as statistics on motor third-party liability claims from Insurance Europe; provision of consumer information and vehicle safety assessment based on insurance losses and injury outcomes, such as Insurance Institute for Highway Safety and Highway Loss Data Institute (IIHS-HLDI) Insurance losses by make and model [40, 42]; development and enrichment of crash databases, such as crash databases from the German Insurers' Accident Database (UDB) [40, 43]; injury risk analysis using crash recorder data, such as crash recorder databases from Folksam and AXA [40].

1.6 Overview of crash reconstruction

Scientific reconstruction of road traffic accidents has become an indispensable tool for forensic investigation, liability determination, and the development of evidence-based road safety policies. Crash reconstruction is defined by SAE International as a scientific process that utilizes principles of physics and empirical data to analyze the physical, electronic, video, audio, and testimonial evidence from a crash to determine how and why the crash occurred [44]. The outputs of this process could typically include vehicle velocities and trajectories, velocity changes, deformation energies, and occupant load profiles that span the pre-impact, impact, and post-impact phases [45]. Crash reconstruction can be divided into pre-impact, impact, and post-impact phases, which differ not only in timing but also in analytical focus: pre-impact reconstruction commonly aims to determine the unknown pre-impact velocities, impact reconstruction estimates collision-related velocity changes such as Delta-V, while post-impact reconstruction often starts from scene evidence, including skid marks, debris, final rest positions, and vehicle deformation, to infer the vehicles' motion immediately after separation [46].

A broad spectrum of methodologies has been established to address the diversity of collision scenarios encountered in practice, and can be broadly categorized into impulse-momentum models, multi-body system (MBS) simulations, and finite element (FE) analysis [47]. These approaches differ considerably in their computational demands and analytical scope: FE methods offer greater fidelity by correlating kinetic energy losses with structural deformation [48], while MBS simulations enable efficient reconstruction of vehicle and occupant kinematics, and are often employed in combination with FE analysis—with MBS-derived kinematic conditions serving as inputs to subsequent FE injury prediction [49]. The quality of scene documentation is equally consequential, and geomatics techniques such as unmanned aerial vehicle photogrammetry and laser scanning have progressively supplanted traditional methods, enabling high-precision acquisition of scene geometry, vehicle deformation, and victim morphology [49].

1.7 Aims and objectives

This project aims at identifying the gaps between insurance data and IGLAD common data scheme in terms of available variables, data quality, and structure. Analyzing insurance data and evaluating the feasibility of transforming it to be IGLAD-compatible is the main objective. To achieve this, the study will review and examine both the IGLAD common data scheme and a set of insurance crash data, and will explore methods for harmonizing possible new crash datasets in IGLAD together with feasible format transformations. Harmonizing insurance data with other data sources is expected to provide a more comprehensive and accurate understanding of road traffic safety and to improve the level of in-depth accident investigation. Accordingly, the effective connection, integration, and harmonization of insurance accident data with structured accident databases constitutes the core of this project. This project will be closely related to crash reconstruction, and identifying the variables relevant to reconstruction and their possible sources will form the basis of the main outcome of this project. Through understanding the field of crash reconstruction and the identification of reconstruction related variables in crash data, particularly input variables, a Weighted Compatibility Rating System (WCRS) will be proposed, constructed, and presented in the form of an interactive dashboard. The WCRS will be a quantitative tool to support researchers to evaluate the readiness of given crash data sources to harmonize with IGLAD or how well individual cases can support IGLAD-compatible reconstruction.

Additionally, the reconstruction readiness of specific insurance crash cases will be analyzed, and the viability of conducting complete reconstructions of these cases will be studied. Based on these analyses, recommendations for data collection practices for the involved stakeholders to improve future in-depth crash data harmonization will be proposed. This project will also enrich the diversity of sources of in-depth accident data through a series of methods that help IGLAD find potential new members.

2

Methods

This chapter includes an introduction to the datasets used, a portion on structural-Leveled variable harmonization process, and the proposal of a method for evaluating reconstruction relevant variables.

2.1 PAV crash database

According to 2 § of Traffic Injuries Act (1975: 1410) of Sweden [50], Motor (Third Party) Liability Insurance must cover all motor vehicles registered in the road traffic register that have not been decommissioned, as well as other motor vehicles used in the country's traffic. If P&C Insurance (hereinafter referred to as "If") insures about 25 % of all cars in Sweden [51], and aims to contribute to making roads safer for both drivers and VRUs such as pedestrians and cyclists [52]. Originally established in 2010 and and has undergone continuous revisions since then, If's People Around the Vehicle (PAV) crash database contains information for crashes between motor vehicle and active travelers. As introduced in the previous chapter, the scope of active treveler may vary. The definition of active travelers in PAV includes not only pedestrians and cyclists, but also skateboarders, scooter users, and wheelchair users [53].

The PAV data covers all of Sweden, is identified through claims information, and encompasses a wide range of crash scenarios [53]. Each case is analyzed and annotated based on accident descriptions from drivers and active travelers. Police reports are also used when available, and witness information is supplemented as needed [53]. When an active traveler is injured, injury information is collected from statements and medical journals, if available [53]. Information about the damage to the motor vehicle in accidents comes from photographic documentation of these locations [53]. The PAV data collection is a qualitative process [53].

For each case, the information collected is coded according to the coding protocol. Coding consists of entry values or free texts, depending on the specific circumstances of each case. Each case contains approximately 200 coded variables. Injury information from hospital reports and medical journals is used for coding according to Abbreviated Injury Scale (AIS), Association for the Advancement of Automotive Medicine (AIS), Association for the Advancement of Automotive Medicine (AAAM)), 2016, version 6 [54]. For motor vehicles, photographs of vehicle damage are used and coded according to Collision Deformation Classification (CDC (CDC), SAE J224-2022 [55]. For descriptions of the road environment, various map services (such as Google Maps) are used.

To ensure the highest data accuracy and maintain coding consistency, PAV has a rigorously adopted set of quality processes. The quality assurance route ensures coding continuity; the inter-rater reliability check (Q-check) is conducted about four times annually [53].

In this project, a subset of insurance crash data provided by If from PAV crash database is used.

The structure of PAV consists of three tables, which respectively are:

- Accident
- Injury
- Impact

The Accident table contains most of the information about the accident, such as the sequence of events, the location and traffic environment of the accident, the weather and lighting conditions, and demographics [53]. The unique ID and accident information based on the unique active traveler are included in each row. Multiple rows will be created, and one row for each active traveler, if multiple active travelers are involved in a collision with the same vehicle; a sequence number associates different active travelers with the same accident [53].

In the Impact table, collisions involving active travelers are coded, including those with vehicles, their own transportation equipment, the ground, or other objects, with each row containing information about a single collision event; each collision has a unique ID (ImpID), which is associated with a case ID [53]. Any additional information regarding each impact, as well as the possible CDC coding of vehicle damage, is also provided here [53].

In the Injury table, all personal injuries sustained by active travelers that have been or can be medically diagnosed are coded, including minor injuries such as abrasions; information about active travelers experiencing only pain is not coded as AIS injury but added as a separate entry [53]. Each row contains information about a unique injury, and each injury has a unique injury ID, linked to a case ID [53].

In addition, images showing crash details and photos of vehicle damage during inspections (screenshots of the information analyzed and coded in each case) will be added to a separate Figures file to facilitate revisiting specific cases during report writing [53].

2.2 Preliminary assessment

This section presents the preliminary work required before conducting variable-level harmonization and development of the rating system. Through the following explorations, differences and gaps between PAV and the IGLAD common data scheme in terms of available variables are shown.

2.2.1 Review of IGLAD common data scheme

The primary purpose of this project is to contribute to evaluating the feasibility of new datasets from potential members for the IGLAD dataset through a series of initiatives. Therefore, the IGLAD common data scheme is reviewed via the IGLAD codebook [36].

At present, containing 133 mandatory variables, the IGLAD common data scheme is organized around four interrelated relational tables:

- Accident (ACCIDENT)
- Participant (PARTICIPANT)
- Occupant (OCCUPANT)
- Safety System (SAFETYSYSTEM)

Each table captures a distinct analytical dimension of a road traffic accident, and together they form a hierarchical, case-based structure in which individual accident cases serve as the primary unit of analysis. The accident number (CASENR), formatted as [yy][XX][0000] to encode the member year and country of origin, appears as the first variable in all four sections, serving as the universal key that links records across the entire database. Additional linking variables are introduced progressively at lower levels: the participant number (PARTNR) is shared by PARTICIPANT, OCCUPANT, and SAFETYSYSTEM; the occupant number (OCCNR) further distinguishes individuals within OCCUPANT; and the system number (SYSNR) identifies each safety system entry within SAFETYSYSTEM.

ACCIDENT contains 29 variables describing the general circumstances of each crash event. These cover temporal and spatial attributes such as the year, month, weekday, time of occurrence, and GPS coordinates, along with road environment characteristics including road type, surface condition, lane separation, lighting, and weather. Accident-level variables further document the collision type, accident type, main contributing factor, and overall accident severity, with aggregate count variables for participants, vehicles, persons, and injured persons supporting cross-table quality control.

PARTICIPANT contains 58 variables recording information on each vehicle or non-motorized road user involved in the accident. Beyond vehicle attributes such as make, model, registration year, dimensions, engine type, and seating capacity, the table captures detailed reconstruction data for both primary and secondary collision events, including CDC/TDC deformation coding, kinematic parameters (initial speed, delta-v, EES, deceleration), contributing factors, pre-crash scenario classification, and quality check variables for reconstructed data.

OCCUPANT contains 40 variables pertaining to each individual involved in the accident, whether a driver, passenger, or pedestrian. Demographic information (age, gender, weight, height) is recorded alongside injury outcomes at multiple levels of detail: police reported severity, MAIS score, and AIS scores across nine body regions. The table also documents restraint system use and the deployment status of multiple airbag types, as well as child restraint systems, helmets, and headrest protection systems.

SAFETYSYSTEM contains six variables and records the active safety systems fitted to each involved vehicle. For each system, the type (such as ABS, ESC, AEB, LDW), operational status at the moment of accident initiation, and whether the system was activated by the vehicle during the event are all documented, enabling research into the real-world effectiveness of these technologies.

In addition to the four core sections, the coding manual also includes eight appendices (Appendix A-H) as supplementary information. This includes: Airbag Coding Examples (A), Description of Active Safety Systems (B), Sampling Procedure (C), ACCTYPE (accident type classification) for left-hand traffic (D), Guidelines for IGLAD accident sketches (E), Codebook change log (F), Country specific data collection (G), and IGLAD publications (H).

Data entry and reasonableness checks have become standardized using UNIDATO software platform since Phase 2 of the IGLAD project, which serves as a common data collection tool for all participating organizations, automating quality control before dataset merging and publication [36].

2.2.2 Variable mapping comparison: five-level classification

A thorough review of both the PAV manual [53] and IGLAD codebook [36] is conducted. Given the differences in the variable structure of the datasets in PAV and IGLAD, a variable-level comparison was necessary to determine whether each IGLAD variable had a corresponding variable in PAV, the degree of correspondence, and whether recoding or harmonization might be required before any data merging.

To compare the two datasets, each one of the 133 mandatory IGLAD variables was reviewed against PAV and assigned to one of five levels defined in Table 2.1.

Table 2.1: Interpretation of variable mapping levels.

Level	Interpretation
Yes	Strong indicative match, directly mappable.
Partly	Partial indicative match, mappable with recoding or value-set adjustment; or, due to structural or crash-type constraints.
Possibly	Potential match, mappable only with rule-based inference from other variables.
No	No counterpart in the other dataset.
NA	Not applicable for mapping due to structural or crash-type constraints.

2.3 Framework-level variable harmonization

Exploratory efforts on framework-level harmonization and are presented below: the selected variable examples from Section 1 (ACCIDENT) are harmonized. The following subsections illustrate the mapping strategies and recoding procedures for key variables in this section.

2.3.1 Weekday (WDAY)

In PAV, variables SKADEMAN and SKADEDAG encode the month and day of the crash. Since the year of crash occurrence is known for all PAV cases, i.e., the IGLAD variable "Year of accident (YEAR)" would be available for all cases, this variable can be coded by consulting external sources (such as calendars).

2.3.2 Time (TIME)

This IGLAD variable can be directly mappable. However, the information of the hour (hh) and minute (mm) that occurred in the crash is coded separately under two different variables TT and MM in PAV. It is worth noting that if, in a certain PAV case, only one of TT and MM is unknown, the variable should be coded as "9999 - unknown" here. Figure 2.1 visually presents the mapping and recoding relationships.

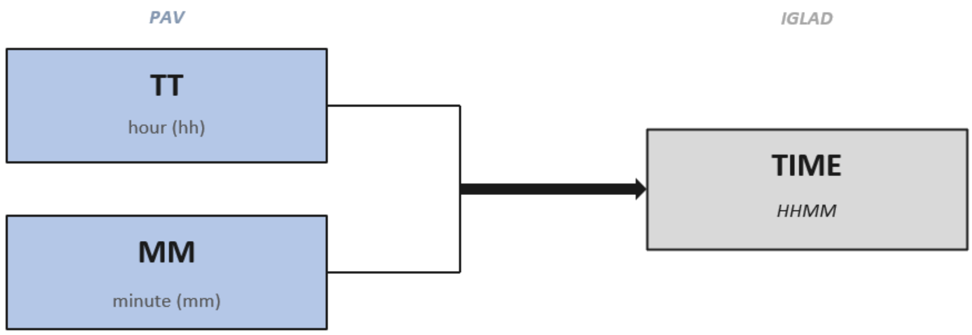


Figure 2.1: Mapping of TT and MM (PAV) to TIME (IGLAD).

2.3.3 Accident description (ACCDESC)

This IGLAD variable not only requires a general description, but also incorporates all relevant technical and medical characteristics into both the occurrence of the accident and its consequences. The variable FORLOPP in PAV only contains the description of events and their sequence of the crash, therefore extra information should be gained from other PAV variables. Variables that can provide supplementary information for the mapping are listed below:

- The possible influence of alcohol can be mapped from ALKOHOLVRU (only for active traveler participant) ;
- The type of collision can be mapped from CRASHTYPE and KONSIT;
- The description of possible visibility obstructions can be mapped from SIKTHINDERBIL (for driver participant) and SIKTHINDERVRU (for active traveler participant);
- The information of special weather and road conditions can be mapped form VADER and VAGLAG separately;

- The severity of the injuries can be mapped from multiple variables in table ALLINJURYVARIABLES.

Figure 2.2 visually presents the mapping relationships.

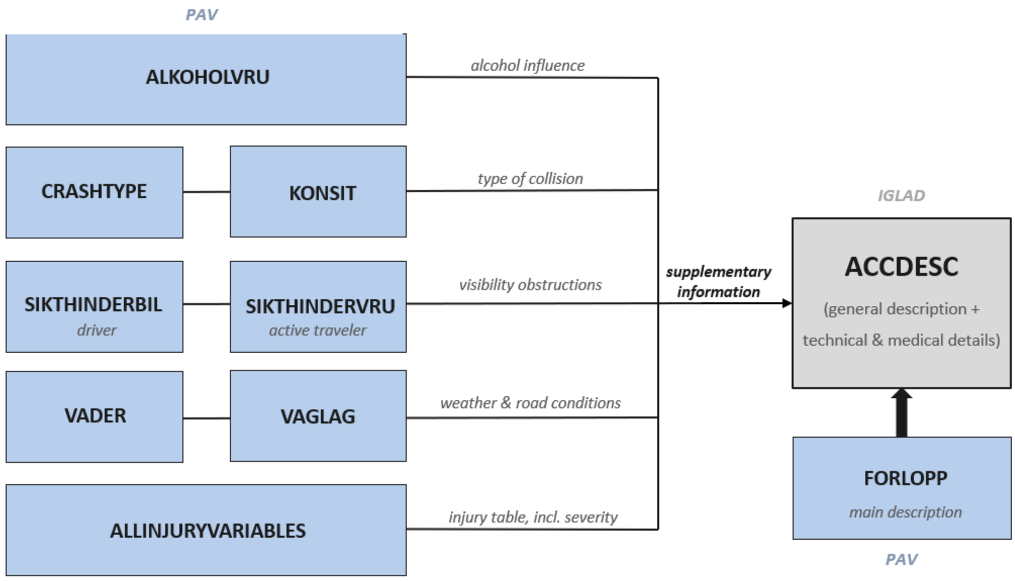


Figure 2.2: Mapping of multiple PAV variables to ACCDESC (IGLAD).

2.3.4 Accident type (ACCTYPE)

This IGLAD variable can be mapped mainly based on KONSIT in PAV. However, However, in ACCTYPE, the circumstances leading to an accident or conflict are described using a set of detailed coding rules, distinguishing between different types of the accident site, relative positional relationship with the intersection, sight obstruction occurrence, and whether lane changes happened. In contrast, in PAV, the conflict situation in KONSIT does not include such information; road types, crosswalks, parking areas, intersections, and other conditions are coded separately as other variables. Variables that can provide supplementary information for the mapping are listed below:

- Information on whether a participant changes lanes during a conflict can be mapped from FILBYTEBIL (for driver participant) and FILBYTEVRU (for active traveler participant);
- The possible visibility obstructions can be mapped from SKYMDSIKTBIL (for driver participant) and SKYMDSIKTVRU (for active traveler participant);
- Information about active traveler crossing the road in conflicts can be mapped from KORSVRU;
- The information of participants' placements relative to roundabouts, intersections or parking areas can be mapped form PLATSBIL (for driver participant) and PLATSVRU (for active traveler participant);

- Whether participants in the conflict have information about overtaking can be mapped from OMKORBIL (for driver participant) and OMKORVRU (for active traveler participant).

Figure 2.3 visually presents the mapping relationships.

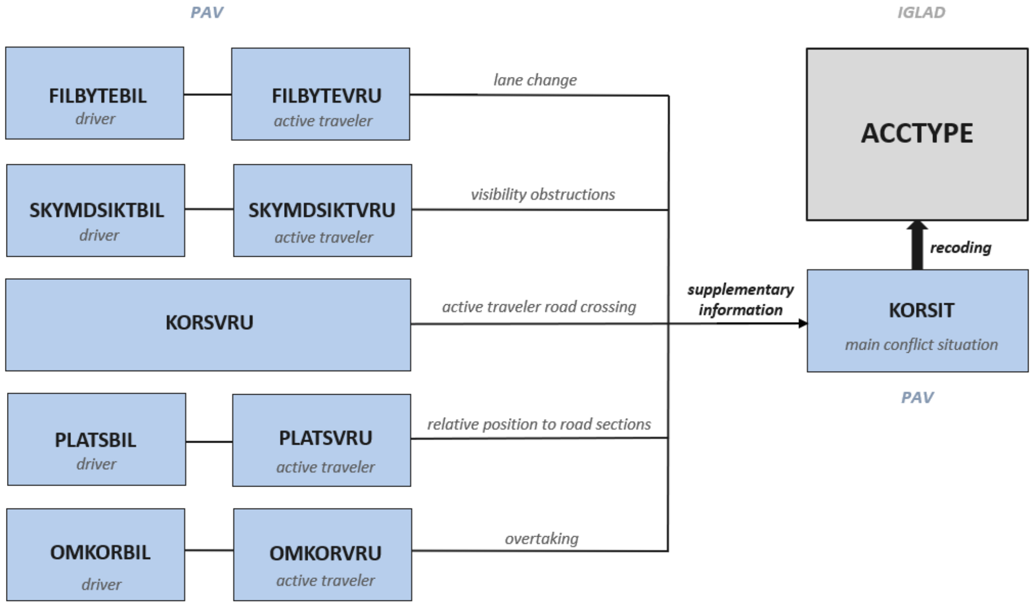


Figure 2.3: Mapping of multiple PAV variables to ACCTYPE (IGLAD).

2.3.5 Road type (ROADTYPE)

This IGLAD variable can be mapped from VAGTYP in PAV. The classification and categorization of road types vary significantly from country to country. In IGLAD, ROADTYPE provides a general definition and establishes its coding rules, ranging from principal arterial to local road, supplemented by codes for car parks, private roads, footpaths, cycle paths, traffic-calmed areas, unsealed roads, and the residual categories “88888 - other” and “99999 - unknown”. The actual meaning of each defined label depends largely on the corresponding rules adopted by each member organization when converting their national administrative and design classifications into the IGLAD format.

PAV collects and covers accidents in Sweden; therefore, its coding rules are based on the domestic road classification in Sweden. This subsection proposes a mapping scheme on road type related variables between PAV and IGLAD based on Swedish road network categorization, and provides suggestions for the recoding of PAV variable VAGTYP for reference.

In Sweden, road types are primarily defined based on administrative responsibility and functional usage. According to Trafikverket [56], the road network consists of municipal roads, state roads, and a large proportion of private roads (enskilda vägar), which are typically maintained by property owners or road associations. In parallel, the regulatory framework defined in Vägmärkesförordning (2007:90) [57] provides standardized categorizations of road environments through traffic signs, distinguishing between built-up areas (tätort), rural areas, and specific road types such as motorways (motorväg) and expressways (motortrafikled).

In contrast, ROADTYPE classifies road environments using a set of categorical defined labels. The mapping process involves translating Swedish administrative and regulatory classifications into functionally equivalent IGLAD categories. The detailed correspondence between PAV VAGTYP codes and IGLAD ROADTYPE categories for the Swedish context is presented in Table 2.2 below.

Table 2.2: Proposed mapping between PAV VAGTYP code values and IGLAD ROADTYPE defined labels.

Code value and description in VAGTYP (PAV)	**Possible defined label in ROADTYPE (IGLAD)**
0. Not applicable	88888 - Other
1. Motorway	1 - principal arterial
2. Expressway	1 - principal arterial; 2 - secondary arterial
3. County road	3 - collector
4. Street in residential or urban area	4 - local; 7 - foot path; 8 - cycle, rambling path; 9 - traffic-calmed area (4–7 kph)
5. Cobblestone street in residential or urban area	4 - local; 7 - foot path; 8 - cycle, rambling path; 9 - traffic-calmed area (4–7 kph)
6. Private, minor road	6 - private road
7. Other unpaved road	10 - unsealed road without pavement, earth road, gravel road
8. Large open area (not parking)	88888 - other
9. Small open area (not parking)	88888 - other
10. Private driveway	6 - private road
11. Small area	88888 - other
999. Unknown	99999 - unknown

It is worth noting during the mapping process that one VAGTYP code value may correspond to multiple ROADTYPE defined labels. In the recoding of PAV cases, when this occurs, information from other variables, such as FORLOPP, LATITUD, and LONGITUD, should be used to assist in determining the new code. In addition, VAGTYP does not include codes for accidents that occur in parking areas; this information is captured in the variable PTYP. In case that PTYP is not coded as "not applicable" in a given case, then after recoding during case harmonization, ROADTYPE should be coded as "5 - car park".

Figure 2.4 visually presents the mapping relationships.

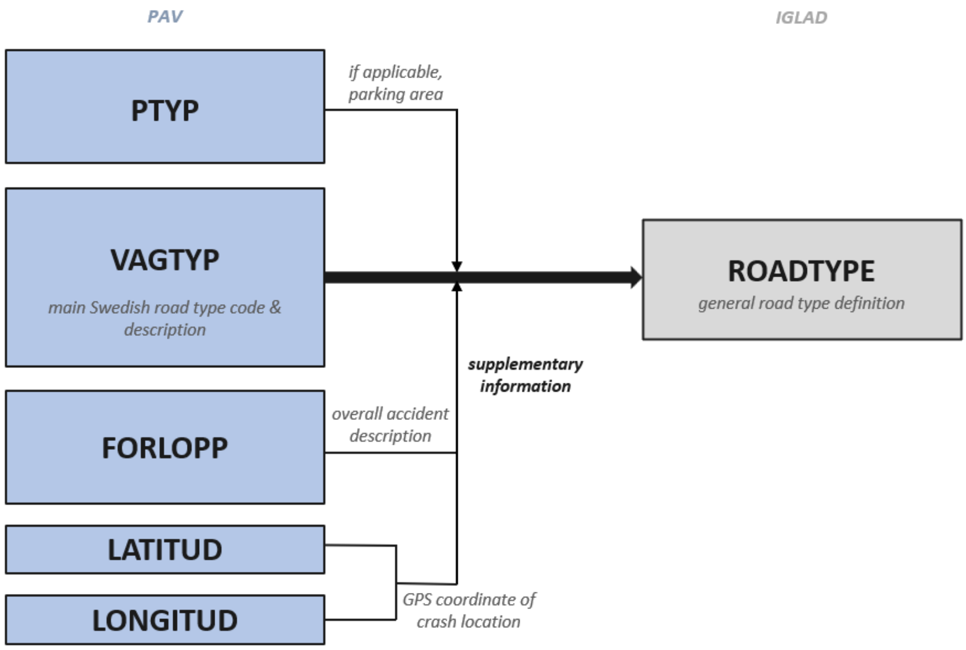


Figure 2.4: Mapping of multiple PAV variables to ROADTYPE (IGLAD).

2.4 Reconstruction input variable evaluation

In this section, the Weighted Compatibility Rating System (WCRS) is introduced. The WCRS is a quantitative tool developed to evaluate reconstruction readiness of crash data at two complementary levels: at dataset level, it assesses how comprehensively a given crash data source covers the variables required for IGLAD-compatible reconstruction; at case level, it quantifies how well the evidence documented for an individual crash case supports IGLAD-compatible reconstruction. The underlying premise is that crash datasets differ substantially in their coverage of reconstruction-relevant variables and in the objectivity of the sources supplying those variables; a structured scoring framework is therefore needed to make such differences transparent and comparable across sources. The system operates on a two-dimensional reconstruction relevant variable grid (Figure C.2): rows represent the 75 reconstruction-relevant variables organized into ten thematic categories; columns represent the 14 data sources that can supply evidence for those variables. Each cell records whether the variable is available from the corresponding source as a binary presence indicator $P(v, s) \in \{0, 1\}$, where v denotes a reconstruction-relevant variable (a row) and s denotes a data source (a column); a cell is counted as present if any non-empty value appears in the input template.

The design of the WCRS and the construction of the reconstruction relevant variable grid involve an independent and systematic review of primary source documents. The IGLAD codebook [36] and PAV manual [53] are examined in full and in parallel, informing the identification of reconstruction-relevant variables, the assignment of applicability flags, and the delineation of variable categories — decisions for which no directly transferable precedent exists in the literature. The variable grid, the two-dimensional reliability matrix $R(s,t)$ indexed by both source and variable type, and the combination of source columns spanning six methodological families are therefore original constructs of the present work, shaped by direct engagement with the structural realities of the datasets under study rather than adopted from any single external framework. Building on this foundation, the framework further incorporates literature-informed principles and project-specific scoring decisions.

The selection of reconstruction-relevant variables is informed by the IGLAD-PCM methodology developed by Spitzhüttl et al. [58] primarily, which defines the minimum information requirements for creating Pre-Crash Matrices (PCM) from IGLAD data and develops compensation methods for missing or unknown information. The study distinguishes reconstruction inputs from accident sketch content as the two prerequisites for pre-crash simulation. The reconstruction inputs include the coded IGLAD dynamics variables (initial speed, collision speed, deceleration, and deceleration distance), together with inputs that have no direct IGLAD variable and must be estimated or reconstructed, such as the road surface friction coefficient and the vehicle-to-opponent contact point; the sketch supplies trajectory and scene information. These requirements provide the conceptual basis for distinguishing reconstruction-critical variables from variables that mainly contribute contextual or descriptive information. However, the variable importance tiers used in the WCRS (Speed-critical, Essential, Important, and Contextual) and their associated numerical weights are not adopted directly from the PCM study; they represent a self-developed, framework-specific design. The PCM requirements serve as a qualitative reference for identifying variables with higher reconstruction relevance, while the final tier assignments and weight values are methodological choices introduced to operationalize reconstruction readiness within the WCRS.

A separate importance tier is introduced for vehicle initial speed, vehicle crash speed and VRU crash speed because these variables represent central kinematic inputs and outputs in reconstruction-oriented analysis. Their elevated importance is consistent with PCM-based reconstruction methodologies, in which speed forms part of the dynamics-data prerequisites for pre-crash simulation and reconstruction, alongside deceleration and trajectory information [58]. Furthermore, Gabauer and Gabler [59] found that delta-V, a velocity-change-based kinematic measure derived from EDR data, was a reasonable predictor of occupant injury and performed comparably to more complex severity metrics. These findings highlight the analytical importance of objective kinematic information in crash analysis. Within the WCRS, they provide supporting rationale for treating speed-related variables as a distinct high-priority group. Nevertheless, the Tier-S classification and its associated numerical weight remain original methodological design choices introduced in this project.

Source reliability is inspired by the data-recording technology hierarchy discussed by Vida and Török [60, 61]. Their studies distinguish traditional data recording, EDR and EDR+ technologies and examine how increasingly advanced data sources affect accident accessability. The 2023 study compares accessability across these technology levels using 124 accident cases, while the 2025 study further quantifies the relationship between accident category, recording technology and accessability using an ordered logit model on a related 123-case dataset. Within the WCRS, this hierarchy is used as an ordinal reference, whereby more comprehensive and technologically advanced data sources are generally considered to provide higher reconstruction value and support more reliable accident assessment. The numerical reliability values used in WCRS, including the five-level scale from 1.00 to 0.30, are are framework-specific operational designs rather than values derived directly from the cited studies.

A two-dimensional reliability matrix $R(s, t)$ is constructed so that each source is credited only for the types of variables it can meaningfully document. When multiple sources cover the same variable, the best source provides the base score and the second source contributes a small corroboration bonus reflecting the evidential value of independent cross-validation. The bonus coefficient is a conservative design parameter of the present framework rather than an empirically calibrated value.

Category scores are aggregated into a composite score as a weighted average over applicable categories, with categories not applicable to a given crash type excluded from both the score sum and the weight denominator so that irrelevant categories neither inflate nor depress the result. The score presentation implemented in the dashboard module, in which the percentage-based composite is shown together with category-level color bands, is inspired by the way Euro NCAP demonstrates vehicle safety performance to the public [62]. Euro NCAP enforces all-round performance through balance criteria, whereby insufficient performance in an individual assessment area restricts the overall star rating despite a higher weighted score [63]; WCRS does not apply such a downgrade; the composite remains a re-normalized weighted average, and uneven coverage is instead surfaced diagnostically through the min_cat_score flag and the per-category color bands. In WCRS the Euro NCAP star rating is not used and is replaced by a percentage-based score presented with color-coded category interpretation.

As a main deliverable of this project, the dashboard is detailed in Section 3.3.

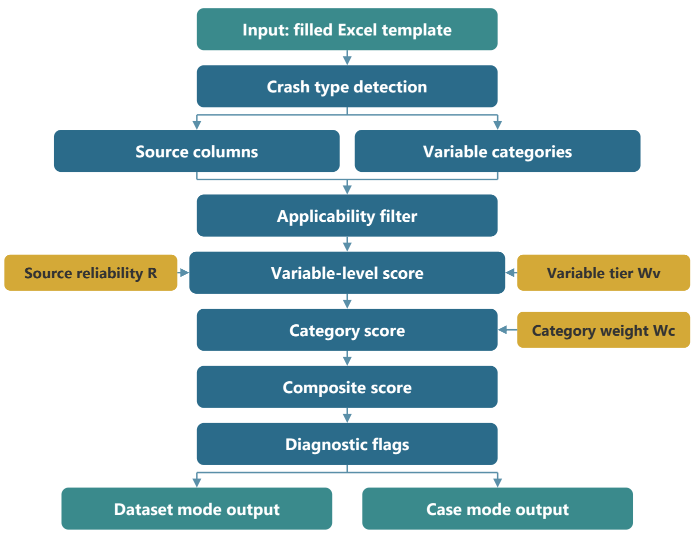


Figure 2.5: Flow chart of WCRS assessment process (Image generated by Claude for PowerPoint, Anthropic, 2026).

2.4.1 Scoring formula

2.4.1.1 Variable-level score

For each variable v and source s, the effective contribution is:

$$\text{eff}(v,s) = R\Big(s,\ \text{type}(v)\Big) \times P(v,s) \tag{2.1}$$

Sources are ranked by their effective contribution in descending order. The primary source (highest $\text{eff}(v,s)$) provides the base score. The secondary source contributes an additional 10 % of its own effective contribution as a corroboration bonus, reflecting the evidential value of independent cross-validation without allowing secondary sources to dominate the score:

$$\text{combined\_eff}(v) = \min\Big(1.0,\ \text{best\_eff}(v) + \alpha \times \text{second\_eff}(v)\Big) \tag{2.2}$$

where $\text{best\_eff}(v)$ is the maximum $\text{eff}(v,s)$ over all sources and $\text{second\_eff}(v)$ is the second-highest. The coefficient $\alpha = 0.10$ (10 %) was selected as a conservative design value reflecting the assumption that corroboration by a second independent source adds limited marginal evidential value beyond the primary source; sensitivity to this parameter was not formally evaluated and remains a direction for future work. Only the top two sources are considered; third and subsequent sources are ignored to prevent score inflation from multiple low-quality sources. The coefficient is stored as the constant `SECONDARY_SOURCE_WEIGHT` in `weights.py`.

The variable score is:

$$\text{var\_score}(v) = W_v \times \text{combined\_eff}(v) \tag{2.3}$$

The maximum achievable variable score is $\text{var\_max}(v) = W_v$, corresponding to the case where the best available source has $R = 1.00$ and $P = 1$.

2.4.1.2 Applicability and in-scope variables

Variables are classified as *in-scope* for a given case based on the crash type detected or assigned:

- Variables marked **V** (motor-vehicle participant only) are always in scope.
- Variables marked **A** (active traveler only) are in scope only when the crash type is V-A or Mixed.
- Variables marked **B** (both) are always in scope.

Out-of-scope variables are excluded from both numerator and denominator of the category score, so that a V-V case is not penalized for the absence of VRU-specific variables.

2.4.1.3 Category score

The category score for category c is the ratio of achieved to maximum variable scores over all in-scope variables:

$$\text{cat\_score}(c) = \frac{\sum\limits_{v \in c,\,\text{in-scope}} \text{var\_score}(v)}{\sum\limits_{v \in c,\,\text{in-scope}} \text{var\_max}(v)} \tag{2.4}$$

A category for which all variables are out of scope has $\sum \text{var\_max}(v) = 0$; it is treated as inactive and excluded from the composite score denominator.

2.4.1.4 Composite score

The composite score is a weighted average of active category scores, re-normalized to the sum of weights for active categories:

$$\text{composite} = \frac{\sum\limits_{c\,:\,\text{active}} W_c \times \text{cat\_score}(c)}{\sum\limits_{c\,:\,\text{active}} W_c} \tag{2.5}$$

Weights sum to 1.00 and are enforced by an assertion in `weights.py`. The renormalization ensures that V-V cases, which have no active VRU Properties category, are not structurally penalized relative to V-A cases.

2.4.1.5 Diagnostic flags

In addition to the composite score, `scorer.py` computes three diagnostic quantities that support qualitative interpretation of the result alongside the composite percentage:

- min_cat_score: the minimum category score across all active categories. Displayed on the scorecard alongside the composite to flag whether any single category is a serious outlier.
- tier_S_missing: the count of in-scope Tier-S ($W_v = 4$) variables for which no source yields $R \times P > 0$. A non-zero value is highlighted in red on the dashboard, indicating that at least one speed-output variable has no evidential coverage at all.
- tier_E_missing: the count of in-scope Tier-E ($W_v = 3$) variables with no coverage, similarly flagged.

2.4.2 Variable importance weights

2.4.2.1 Four-tier scheme

Variables are assigned to one of four WCRS importance tiers according to their assumed role, and Table 2.3 shows the assignment below. The three lower tiers

— Essential ($W_v = 3$), Important ($W_v = 2$), and Contextual ($W_v = 1$) — are used as WCRS design categories informed by related work on accident-data completeness, assessability, and kinematic-data correctness, while the numerical weights are defined within the present study. The fourth tier, Tier S ($W_v = 4$), is introduced specifically for selected speed-related kinematic variables and is justified in Section 2.4.2.2.

Table 2.3: Variable importance tiers and weights.

Tier	W_v	Definition
S	4	Speed-critical: selected speed-related kinematic variables for which source objectivity strongly affects interpretability.
E	3	Essential: variable strongly affects reconstruction-oriented interpretation or IGLAD-compatible coding completeness.
I	2	Important: variable affects reconstruction, severity interpretation, or contextual assessment, but may sometimes be estimated from external information, simplified assumptions, or compensation methods.
C	1	Contextual: descriptive or supporting variable that provides useful context but has lower direct influence on the reconstruction-readiness score.

2.4.2.2 Rationale for Tier S

The introduction of a dedicated Tier S with weight $W_v = 4$, separate from the Essential tier at $W_v = 3$, is a WCRS-specific design and therefore requires specific justification.

When impact speed is provided by an objective source such as an EDR (reliability $R = 1.00$), the reconstruction is substantially a process of verification rather than estimation. When it is available only from a witness statement ($R = 0.50$) or not at all, the reconstruction problem is fundamentally different and becomes more dependent on estimation or compensation.

With $W_v = 3$ (Essential), the maximum achievable variable score for police-only coverage of impact speed ($R = 0.70$) would be $3 \times 0.70 = 2.10$ out of 3.00, which may make the difference between estimated and directly recorded kinematic evidence less visible in the WCRS output. With $W_v = 4$, the same scenario yields $4 \times 0.70 = 2.80$ out of 4.00, and the absolute gap between police coverage and EDR coverage ($4.00 - 2.80 = 1.20$ vs. $3.00 - 2.10 = 0.90$ under Tier E) is amplified, making the contribution of objective kinematic evidence more visible in the composite score.

2.4.2.3 Commonly assumed parameters

Several variables in the WCRS framework may sometimes be estimated from standard assumptions, external sources, or simplified reconstruction practice when direct case-specific evidence is unavailable:

- *Road surface friction coefficient* μ may be estimated from road condition, surface type, or reconstruction assumptions when no direct measurement is available. It is therefore assigned Tier I rather than Tier E.

- *Driver perception–reaction time (PRT)* is often represented by an assumed value when no objective case-specific evidence is available. It is assigned Tier C.
- *Longitudinal gradient* may be treated as negligible in cases where no evidence suggests a relevant slope (although this assumption should be checked when road geometry is important). It is assigned Tier C.
- *Vehicle curb weight* can often be estimated from manufacturer specifications; *Gross vehicle weight* can also be recoverable with external information. They are therefore assigned Tier I rather than Tier E. The anchor variable *Vehicle make/model/year* retains Tier E.
- *Body height* and *weight* may be estimated from external statistical data such as population-based anthropometric reference data by gender and age group [64] when no case-specific measurements are available. They are assigned Tier I rather than Tier E.

2.4.3 Source reliability weights

The reliability of a data source is not a single scalar but depends on the type of variable the source is supplying. An EDR provides objective kinematic measurements but contains no road geometry information; a vehicle registration authority provides exact vehicle specifications but no crash-event information. The WCRS therefore defines a summary of six variable types:

- **K** — Kinematic: speeds, decelerations, trajectories, steering/brake inputs
- **D** — Damage / physical evidence: deformation, skid marks, rest positions
- **E** — Environment / road context: road type, surface, weather, GPS, speed limit
- **V** — Vehicle properties: make, model, dimensions, equipment fitted
- **O** — Occupant/VRU properties: demographics, belt use, helmet, posture
- **I** — Injury: AIS codes, MAIS, police severity classification

The reliability scale (0.30–1.00) is a rule-based scoring convention defined for the WCRS, with ordinal anchors informed by the technology hierarchy described above. The values are interpreted as relative evidential weights within this dashboard, not as empirically estimated probabilities of correctness:

- **1.00** — Direct objective event-based measurement or authoritative coding for the relevant variable type, such as EDR kinematic data, video-derived trajectory evidence, official vehicle records, or medical injury (AIS) coding.
- **0.85** — Instrumented measurement, professional documentation, or authoritative external records that provide strong but not necessarily direct event-based evidence, such as scene surveys, forensic documentation, photographic documentation, weather/road DB (databases), or telematics.
- **0.70** — Authoritative on-site observation or professional judgment, such as police narrative report of speeds or damage and forensic damage coding.
- **0.50** — Supportive human report or visual documentation that may give interpretation but lacks direct measurement, such as witness statements.

- **0.30** — Retrospective or indirect information, such as retrospective contextual lookup, where the information may still be useful but is less reliable than direct or professional documentation.

The "0.30 floor" is a conservative design that distinguishes weak retrospective information from complete absence of data, while still assigning it substantially lower evidential value than objective event-based recordings or professional documentation. Even retrospective self-reported data carries informational signal and should be credited rather than discarded.

Moreover, the reliability of the source is not always fixed; it is related to the type of variable information extracted from it. Zero entries in $R(s,t)$ represent source–type combinations that are methodologically not applicable within the design. A reliability matrix $R(s,t)$ is defined as shown in Table 2.4.

Table 2.4: Source reliability matrix $R(s,t)$. Bold values indicate the highest reliability for each variable type. "—" denotes an inapplicable combination ($R = 0$).

Source	**K**	**D**	**E**	**V**	**O**	**I**
Witness on-site	0.50	0.50	0.50	0.30	0.50	0.30
Witness later (claim)	0.30	0.30	0.30	0.30	0.30	0.30
CCTV / surveillance	**1.00**	0.85	0.70	—	0.50	—
Dashcam / body-cam	**1.00**	0.85	0.70	—	0.50	—
Vehicle registration	—	—	—	**1.00**	—	—
Police / forensics	0.70	0.85	0.85	0.70	0.70	0.70
External (weather/road DB)	—	—	0.85	0.70	—	—
EDR log	**1.00**	0.85	—	0.70	0.85	—
CAN log	**1.00**	0.85	—	0.85	0.70	—
Telematics	0.85	—	0.70	0.50	—	—
Scene measurements	0.85	**1.00**	0.85	0.85	0.85	—
Photographs	0.85	0.85	0.70	0.85	0.50	—
Map services (retrospective)	—	0.30	0.85	—	—	—
Medical journals	0.85	—	—	—	0.85	**1.00**

2.4.4 Category weights

The ten category weights W_c are defined as transparent design parameters to reflect the assumed relative importance of each variable group for reconstruction-oriented use cases (Table 2.5). It is worth noting that they are not empirically calibrated coefficients, but an explicit and adjustable weighting scheme used within the WCRS.

Table 2.5: Category weights W_c. Weights sum to 1.00.

Category	W_c	**Rationale**
Pre-crash motion	0.16	Primary input to PCM generation and ADAS scenario construction
Physical scene & kinematics	0.15	Backbone of kinematic reconstruction; required for momentum and energy balance

continued on next page

Table 2.5 continued.

Category	W_c	Rationale
CDC/TDC primary collision	0.15	Directly coupled to delta-V; primary severity quantification
Occupant properties	0.12	Required for occupant kinematics, restraint analysis, and HBM (human body model) scaling
Injury severity	0.12	Primary outcome variable; MAIS is the headline metric
Vehicle properties	0.08	Substantially recoverable from make/model/year
Road & environment	0.07	Largely recoverable retrospectively from external databases
VRU properties	0.07	Applicable only to V-A crashes; weight redistributed for V-V
CDC/TDC secondary collision	0.05	Conditional on secondary collision; excluded if not applicable
Human factors	0.03	PRT routinely assumed; TTC derived from PCM
Total	**1.00**	

2.4.5 Score Interpretation

The composite score is a percentage in $[0, 100]$ and is displayed on the dashboard scorecard in a color that reflects three bands derived from the `_score_colour` function (Table 2.6).

Table 2.6: Dashboard color bands for composite and category scores.

Color	Threshold	Interpretation
Green	$\geq 80\,\%$	Strong indicative coverage; most major reconstruction relevant/input variables appear to be supported.
Amber	$\geq 50\,\%$	Partial indicative coverage; important gaps may remain.
Red	$< 50\,\%$	Weak coverage; significant variables are missing or poorly supported.

The same color scheme is applied independently to each category score, so that the scorecard immediately reveals which specific categories are responsible for driving the composite below a satisfactory level.

In addition to the composite, the three diagnostic flags provide qualitative guardrails that the composite alone cannot supply. A dataset that covers Occupant Properties and Injury Severity comprehensively but has no kinematic evidence at all may still obtain a composite near the amber threshold; the tier_S_missing flag, displayed as a red pill when non-zero, signals this structural deficiency directly.

2.4.6 Validation

Sensitivity of Tier-S speed variables to source quality is crucial for the designed system. Table 2.7 illustrates the scoring of vehicle crash speed (#2 in Table 3.4, $W_v = 4$, type K) under five representative source combinations.

Table 2.7: Variable scores for vehicle crash speed under different source combinations ($W_v = 4$, type K).

Source combination	best_eff(v)	$\alpha \times$ second_eff(v)	combined_eff(v)	**Score / 4**
EDR only	1.00	0.000	1.000	4.00
EDR + Police	1.00	0.070	1.000†	4.00
Police + Witness (on-site)	0.70	0.050	0.750	3.00
Police only	0.70	0.000	0.700	2.80
Witness (on-site) only	0.50	0.000	0.500	2.00

† Capped at 1.00 by Eq. (2.2).

The example illustrates a clear five-way discrimination within the designed scoring logic. The absolute gap between witness-only (2.00) and EDR (4.00) is 2.00 points, compared to 1.50 points under Tier E ($W_v = 3$); the gap between police-alone (2.80) and EDR (4.00) is 1.20 points, compared to 0.90 points under Tier E. This amplification ensures that the Pre-crash Motion category score is meaningfully sensitive to the presence or absence of objective kinematic evidence.

3

Results

3.1 Mapping results and analysis

Established with the help from If, Table 3.1 shows the results of this mapping. The first column lists all variables from the four sections of the IGLAD codebook [36], organized by their respective data tables. In the second column, each PAV variable was evaluated against its IGLAD counterpart and categorized as a defined level in Table 2.1 (Section 2.2.2). The third column displays the original comments from the If researchers.

Table 3.1: Mapping of PAV variables to the IGLAD common data scheme for data comparability assessment.

IGLAD Codebook	Mapping PAV	Comment
1 Accident (ACCIDENT)		
1.1 CASENR	NA	
1.2 MEMYEAR	NA	
1.3 YEAR	Yes	
1.4 MONTH	Yes	
1.5 WDAY	Possibly	Could be mapped by using DATE
1.6 TIME	Yes	Could be mapped by using TT + MM
1.7 GPSLAT	Yes	
1.8 GPSLONG	Yes	
1.9 COUNTRY	Yes	
1.10 ACCDESC	Partly/Possibly	FORLOPP does not contain all this information. Most information could be mapped from other PAV variables though
1.11 COLLTYPE	Yes	
1.12 ACCTYPE	Partly	Needs mapping to different PAV variables. The values differ between PAV and IGLAD
1.13 ACCTYPEA	Yes	
1.14 ACCTYPEB	Yes	
1.15 MAINFACT	Possibly	Needs a set of rules and mapping to different PAV variables
1.16 ROADTYPE	Possibly	The values differ between PAV and IGLAD
1.17 ROADSURF	Yes	
1.18 ROADCOND	Yes	
1.19 LANESEPAR	No	
1.20 LIGHTCON	Yes	Could be mapped by using LJUS + GATUBELYS
1.21 WEATHER1	Partly	The values differ between PAV and IGLAD

continued on next page

Table 3.1 continued.

IGLAD Codebook	Mapping PAV	Comment
1.22 WEATHER2	Partly	The values differ between PAV and IGLAD
1.23 LOCATION	Yes	Depends on the definition of rural/urban
1.24 ACCSEV	Yes	
1.25 PCMAVAIL	Yes	No PCM-data is available
1.26 PARTCNT	Partly	Number of occupants in motor vehicle is unknown
1.27 VEHICNT	Yes	
1.28 PERSCNT	Yes	
1.29 INJRCNT	Yes	
2 Participant (PARTICIPANT)		
2.1 CASENR	NA	
2.2 PARTNR	Yes	
2.3 PARTTYPE	Yes	
2.4 VEHMAKE	Yes	
2.5 MODEL	Yes	
2.6 REGYEAR	Yes	
2.7 VEHMASS	Yes	
2.8 GVWEIGHT	Yes	
2.9 VLENGTH	Yes	
2.10 VWIDTH	Yes	
2.11 VHEIGHT	Yes	
2.12 ENGINE	Yes	
2.13 POWER	Yes	
2.14 SEATS	Yes	
2.15 TRAILER	Yes	
2.16 VLIM	Yes	
2.17 OPPON1	Yes	
2.18 NROPPON1	Yes	
2.19 CDC1DIRE	Yes	
2.20 CDC1AREA	Yes	
2.21 CDC1LONG	Yes	Needs a set of rules to be mapped
2.22 CDC1VERT	Yes	
2.23 CDC1TYPE	Yes	
2.24 CDC1EXTT	Yes	
2.25 CDC1PERC	No	
2.26 INISPEED1	Yes	Today based on witness statements if available, or on-board car data if available. (Not common in car-to-VRU crashes)
2.27 DECEL1	No	
2.28 DECDIST1	Possibly	
2.29 DEFANG1	No	
2.30 COLSPEED1	Yes	Today based on witness statements if available, or on-board car data if available. (Not common in car-to-VRU crashes)
2.31 DELTAV1	NA	
2.32 EES1	NA	
2.33 OPPON2	Yes	
2.34 NROPPON2	Yes	
2.35 CDC2DIRE	Yes	
2.36 CDC2AREA	Yes	
2.37 CDC2LONG	Yes	Needs a set of rules to be mapped
2.38 CDC2VERT	Yes	
2.39 CDC2TYPE	Yes	
2.40 CDC2EXTT	Yes	

continued on next page

Table 3.1 continued.

IGLAD Codebook	Mapping PAV	Comment
2.41 CDC2PERC	No	
2.42 INISPEED2	Yes	Today based on witness statements if available, or on-board car data if available. (Not common in car-to-VRU crashes)
2.43 DECEL2	No	
2.44 DECDIST2	Possibly	
2.45 DEFANG2	No	
2.46 COLSPEED2	Yes	Today based on witness statements if available, or on-board car data if available. (Not common in car-to-VRU crashes)
2.47 DELTAV2	NA	
2.48 EES2	NA	
2.49 FACTOR1	Possibly	Needs a set of rules and mapping to different PAV variables
2.50 FACTOR2	Possibly	Needs a set of rules and mapping to different PAV variables
2.51 FACTOR3	Possibly	Needs a set of rules and mapping to different PAV variables
2.52 TYPEDISTRACT	Partly	The values differ between PAV and IGLAD
2.53 SRCDISTRACT	Partly	The values differ between PAV and IGLAD
2.54 SCENARIOTYPE	Yes	Needs a set of rules to be mapped
2.55 SRCRECO	No	
2.56 CHECK_RECO1	NA	
2.57 CHECK_RECO2	NA	
2.58 CHECK_RECO3	NA	
3 Occupant (OCCUPANT)		
3.1 CASENR	NA	
3.2 PARTNR	Yes	
3.3 OCCNR	Yes	
3.4 OCCTYPE	Partly	VRU & Driver. No data for vehicle passengers
3.5 AGE	Partly	VRU & Driver. No data for vehicle passengers
3.6 GENDER	Partly	VRU & Driver. No data for vehicle passengers
3.7 WEIGHT	No	
3.8 HEIGHT	Partly	VRU Height
3.9 INJSEVER	Yes	
3.10 MAIS	Yes	
3.11 AISREGIO1	Yes	
3.12 AISREGIO2	Yes	
3.13 AISREGIO3	Yes	
3.14 AISREGIO4	Yes	
3.15 AISREGIO5	Yes	
3.16 AISREGIO6	Yes	
3.17 AISREGIO7	Yes	
3.18 AISREGIO8	Yes	
3.19 AISREGIO9	Yes	
3.20 BELT	NA	
3.21 BELT_USE	NA	
3.22 AIRBF	NA	
3.23 AIRBF_DEPL	NA	
3.24 AIRBTC	NA	
3.25 AIRBTC_DEPL	NA	
3.26 SIDEB	NA	
3.27 SIDEB_DEPL	NA	

continued on next page

Table 3.1 continued.

IGLAD Codebook	Mapping PAV	Comment
3.28 KNEEB	NA	
3.29 KNEEB_DEPL	NA	
3.30 AIRBSR	NA	
3.31 AIRBSR_DEPL	NA	
3.32 AIRBR	NA	
3.33 AIRBR_DEPL	NA	
3.34 AIRBFC	NA	
3.35 AIRBFC_DEPL	NA	
3.36 HEADREST	NA	
3.37 HEADREST_DEPL	NA	
3.38 CHILDSEAT	NA	
3.39 BOLCHILD	NA	
3.40 HELM	Yes	
4 Safety System (SAFETYSYSTEM)		
4.1 CASENR	NA	
4.2 PARTNR	NA	
4.3 SYSNR	NA	
4.4 SYSTYPE	Partly	AEB only
4.5 SYSUSE	Partly	AEB only
4.6 DEPLACT	Partly	AEB only

In total, 68 variables (51.1 %) are directly mappable (Yes), while 34 variables (25.6 %) are marked as not applicable (NA). Among the remaining variables, 22 show partial or potential correspondence, including Partly (13, 9.8 %), Possibly (8, 6.0 %), and Partly/Possibly (1, 0.8 %). In addition, nine variables (6.8 %) are currently unavailable in PAV, and are therefore classified as No.

To present the visualization of the mapping outcome, a Python-based variable mapping dashboard is developed as one of the preliminary interactive tools in the project (Section 3.2.1).

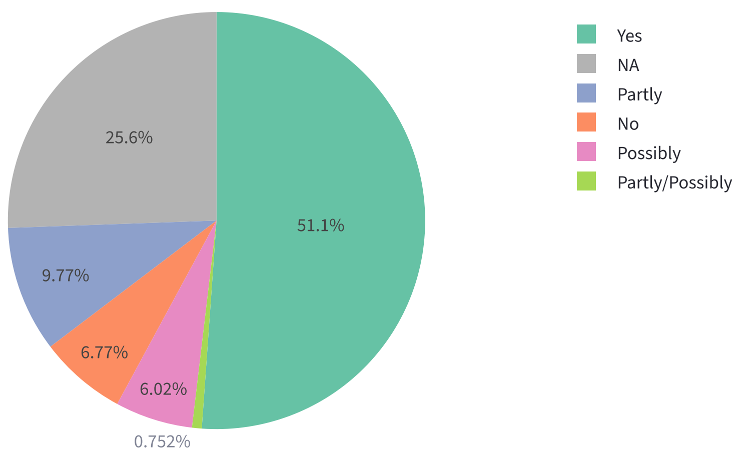


Figure 3.1: Overall composition of variable mapping results from PAV to IGLAD (all results).

Figure 3.2: Distribution of variable mapping results by IGLAD section.

The proportional breakdown is visualized as a pie chart in Figure 3.1, which confirms that direct matches dominate the overall result while the NA group forms a substantial secondary share; Figure 3.2 provides a section-level view of the same results. It reveals that NA outcomes are heavily concentrated in the OCCUPANT section, while the PARTICIPANT and ACCIDENT sections show comparatively higher rates of direct or partial matches.

From the above results, if the variable is only counted as mapped when the status is directly mappable (Yes), then at the structure level, the mapped coverage from PAV to IGLAD is only 51.1 %. However, upon further analysis of the outcomes mapped as not applicable (NA), which account for the second largest proportion, these 34 variables can be categorized into two categories.

The first category of these seven NA variables are not descriptive variables but structural identifiers used to uniquely link records across different tables. These variables, such as CASENR, PARTNR, and OCCNR, ensure the internal consistency and relational integrity of the database rather than representing accident characteristics or data attributes. Therefore, they are inherently tied to the IGLAD database structure itself and can only be meaningfully assigned once a dataset is fully integrated into the IGLAD scheme. During the mapping process of an external dataset (e.g., PAV) to IGLAD, such variables cannot be directly derived or translated from the source data and are consequently treated as NA. Their values are generated only after the data have been formally incorporated into the IGLAD database system.

The remaining 27 NA variables belong to the second category, resulting from the unique and specific nature of accidents (car-to-active-traveler) collected in PAV. Among them, 20 variables from the OCCUPANT section pertain to the presence and deployment of passive safety features (seat belts, airbags, child seats, etc.) protecting vehicle occupants at the time of crash. These systems have almost no opportunity to be activated in car-to-active-traveler accidents. Therefore, collecting this information in such accidents is of bare significance.

Among the seven variables from the Participant (PARTICIPANT) section, the four related to delta-v and energy equivalent speed (EES) are largely inapplicable. Delta-v has traditionally served as a predictor of injury risk for vehicle occupants in vehicular crashes, operating on the premise that a larger change in vehicle velocity corresponds to greater inertial loading on the restrained occupant within the vehicle compartment [65]. This logic does not extend to unprotected road users such as pedestrians and cyclists, whose post-impact kinematics, involving free-body projection, rotation, and secondary ground contact, are governed by fundamentally different mechanics. Indeed, Jurewicz et al. [66] explicitly noted that "the delta-v principle does not provide a good estimation of injury" in vehicle-pedestrian collisions, and VRU-focused research has consequently adopted closing speed or vehicle impact speed as the primary severity predictor [67]. EES faces a more fundamental constraint: its computation requires measuring permanent crush deformation of vehicle structural elements and applying vehicle-specific stiffness coefficients as input [68]. Since pedestrians and cyclists possess no structured, energy-absorbing body, neither crush profiles nor stiffness parameters can be defined for them, rendering EES physically undefinable in this crash type. Finally, the currently applied three checks for the kinematic rationality of the IGLAD reconstruction data also refer to accidents between two or more vehicles [69]. Although its basic physical principles remain valid, this method cannot be used for data validation of car-to-active-traveler accidents.

Based on the above analysis, excluding variables that are not applicable to PAV yields revised statistical results. The resulting distribution is presented in Figure 3.3. Using the same criterion as above, where only variables marked as directly mappable (Yes) are counted, the mapped coverage increases to 68.7 % after excluding NA variables.

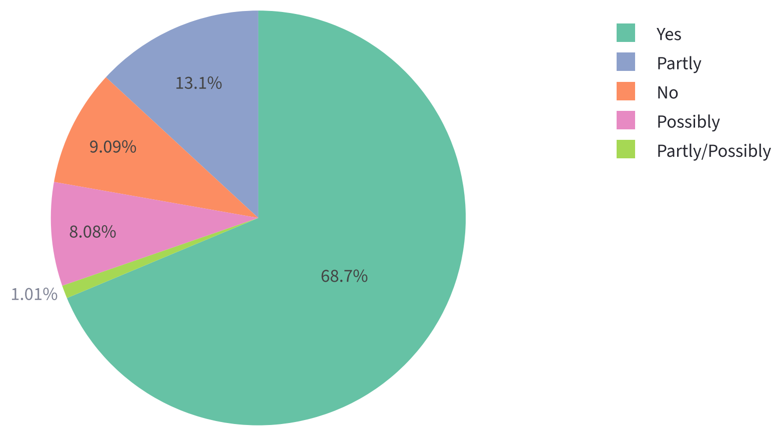


Figure 3.3: Composition of variable mapping results from PAV to IGLAD (NA results excluded).

A further interpretation suggests that if the variables currently classified as Partly, Possibly, and Partly/Possibly can be resolved through structural-level harmonization between the two datasets, the mapped coverage could potentially reach an ideal upper bound of $1 - 9.09\,\% = 90.91\,\%$.

The mapping results confirms that harmonizing PAV with IGLAD is feasible, and the progression is promising.

3.2 Tools for preliminary mapping process

In the preliminary process of mapping PAV to IGLAD at the structural level, several tools were proposed and applied. Broader applicability is taken into consideration when designing, so they have the capability to be used for projects that allow potential mapping of other possible research datasets with IGLAD. Three pages are integrated into a single multi-page browser-based interactive application and navigated via a page selector. Figures demonstrating whole-page layouts for this section are collected in Appendix A.

3.2.1 IGLAD variable mapping dashboard

The mapping dashboard is a two-page interactive application Built on the Streamlit [70] framework and rendered interactive visualizations through Plotly [71]. The application visualizes variable-level mapping coverage between any crash database and the IGLAD scheme. In this section, all illustrations of the display page designs are based on the PAV data as an example. The tool can read Excel files provided by users in a specific content format (Table 3.1) via file upload or local path input, with column detection for the IGLAD section, variable code, mapping status, and optional comment fields through the sidebar. Users can switch between two mapping definitions “All except No and NA” and “Yes only” with all statistics updating accordingly across both pages (Figure A.1).

The first page, **All Variables**, presents the full codebook population (Figure A.2). Three headline KPI metrics at the top display the total IGLAD variable count, the number of mapped variables under the selected definition, and the overall coverage percentage. A section-level summary table below breaks down each of the four IGLAD sections by target variable count, mapped count, coverage percentage, and per-status counts.

Two interactive charts complement the table: a stacked bar chart showing status composition across sections, and a pie chart showing the overall distribution across all six status categories. Sidebar controls support additional filtering by section, mapping status, and variable keyword search. An optional data quality panel, disabled by default, surfaces diagnostic information including unrecognized section labels, non-standard status values, and duplicate variable entries. These visualization features are shown in Figure 3.1 and Figure 3.2 (Section 3.1) when evaluating the mapping outcome results of PAV.

The second page, **NA Excluded**, shares the same layout and interaction design but restricts all computations to variables that are structurally applicable to the database, excluding those marked as NA (Figure A.3). The KPI metrics, summary table, and charts all recalculate against this reduced population, enabling direct comparison with the All Variables view to distinguish coverage gaps attributable to non-applicable variables from those that are genuinely unmapped. The same visualization features on the second page are shown in Figure 3.3 (Section 3.1) when evaluating the mapping outcome results of PAV.

The Comments panel (Figure 3.4) is also a feature on both pages, where variables are organized by IGLAD section, and each entry supports a hover-based tool-tip that reveals the annotation recorded during the mapping process, supporting efficient inline review of partially or tentatively mapped variables.

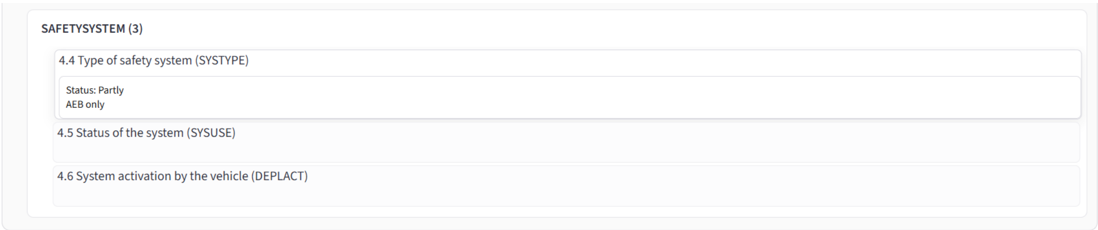


Figure 3.4: Comments panel on IGLAD Variable Mapping Dashboard with section-grouped entries and hover tool-tips (result of PAV dataset presented).

3.2.2 IGLAD codebook Appendix G viewer

The Appendix G Viewer is a Streamlit-based [70] page provides structured access to the country-specific dataset metadata documented in Appendix G of the IGLAD codebook [36]. The tool fetches the latest version of the official IGLAD codebook (publicly available) directly from `iglad.org` at runtime, extracts the Appendix G tables using `pdfplumber`, and caches the parsed result in memory. A manual "Refresh from URL" button in the sidebar allows users to force re-fetching (Figure 3.5).

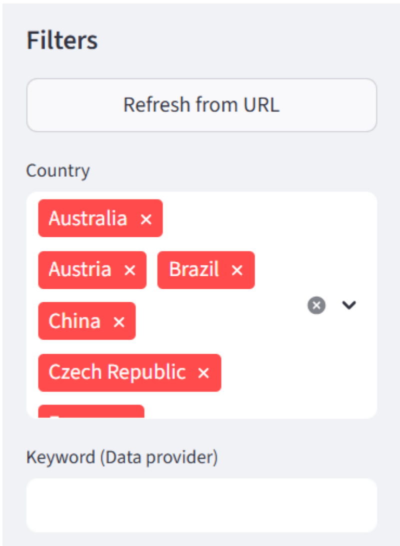


Figure 3.5: Sidebar in IGLAD codebook Appendix G viewer.

Figure A.4 presents the layout of the complete web page of the IGLAD codebook Appendix G viewer. Data source URL and the last-modified timestamp of the fetched PDF are displayed beneath the page title, providing explicit traceability of data provenance. The upper section of the page presents a filterable dataset registry listing all contributing datasets alongside their country and data provider. Summary metrics show the current count of matching datasets and distinct countries. Sidebar controls allow filtering by country via a multi-select widget and by provider name via keyword search. The lower section provides a detailed view for any dataset selected from a drop-down. An overview panel displays key identifiers (country, data provider, and the corresponding page number in the codebook), together with any noted change since the previous codebook version. Below this, all Appendix G metadata fields are rendered as a three-column grid of labeled buttons. Hovering over any button reveals a popover showing both the field description and the dataset-specific value extracted from the PDF, enabling rapid inspection of individual dataset characteristics without manually navigating the source document (Figure 3.6).

Figure 3.6: Detailed information button in IGLAD codebook Appendix G viewer with hover-to-preview function.

3.3 WCRS

The WCRS is implemented across four Python modules. `weights.py` declares all static configuration: the ordered list of 14 sources (`SOURCES`), the source–type reliability matrix (`RELIABILITY`), the full variable framework (`FRAMEWORK`), category labels and weights, and the secondary source bonus coefficient. `loader.py` reads a filled-in Excel workbook and maps each non-empty cell to the corresponding variable–source entry in `FRAMEWORK`, producing a structured case dictionary. `scorer.py` consumes that dictionary and applies the formulas described in Section 2.4.1 to produce a `ScoreResult` object containing variable-level, category-level, and composite scores, together with diagnostic fields (minimum category score, Tier-S and Tier-E missing variable counts). The fourth module, `dashboard.py`, implements an interactive dashboard interface and wraps the scoring core.

3.3.1 Variable framework

3.3.1.1 Source columns

The 14 source columns span six methodological families (Table 3.2). Each source is associated with a reliability profile that depends on the type of variable being supplied, as detailed in Section 2.4.3.

Table 3.2: Source columns of the reconstruction variable framework, in order of appearance.

Source	**Source family**
Witness on-site	Statements
Witness later (questionnaire / claim)	Statements
CCTV / on-site surveillance	On-site footage
Dashcam / body-cam	On-site footage
Vehicle registration authority	Official records
Police / forensics	Official records
External sources (weather, road DB)	Official records
EDR log	Vehicle on-board systems
CAN log	Vehicle on-board systems
Telematics	Vehicle on-board systems
Scene measurements / survey	Scene documentation
Photographs	Scene documentation
Map services (retrospective)	Scene documentation
Medical journals	Medical records

3.3.1.2 Variable categories and counts

The 75 variables are distributed across ten categories (Table 3.3). Variables associated with motor-vehicle participants are marked V; variables applicable only when an active traveler is involved are marked A; variables applicable to both participant groups or crash configurations are marked B. The applicability flag determines whether a variable is in scope for a given case (Section 2.4.1.2).

Table 3.3: Variable counts by category and applicability.

Category	**V**	**A**	**B**	**Total**
Pre-crash motion	4	2	1	7
Road & environment	0	0	9	9
Physical scene & kinematics	2	5	5	12
Vehicle properties	7	0	0	7
CDC/TDC primary collision	5	0	2	7
CDC/TDC secondary collision	5	0	2	7
VRU properties	0	6	0	6
Occupant properties	15	0	0	15
Injury severity	0	0	3	3
Human factors	1	0	1	2
Total	**39**	**13**	**23**	**75**

3.3.1.3 Crash type detection

`loader.py` automatically infers the crash type of each case from the filled Excel template. After reading all presence values, it checks whether any variable whose applicability flag is A has at least one source marked as present. If so, in *Case* mode, the case is classified as vehicle-to-active-traveler (V-A); otherwise, it is classified as vehicle-to-vehicle (V-V). The dashboard additionally offers a manual override control that allows the user to force a specific crash type, or, in *Dataset* mode, to designate the submission as *Mixed*—indicating that the aggregated template covers both V-V and V-A events, so that all variables including VRU-specific ones are treated as in scope.

3.3.2 Complete variable list

WCRS framework is constructed through the implementation of the proposed and explored methodology introduced in Section 2.4. As the main basis of the scoring scheme, Table 3.4 lists all 75 variables in the framework with their applicable crash type, importance tier, and variable type.

Table 3.4: Complete variable list with tier and type assignments. Applicable: V = motor vehicle, A = active traveler, B = both.

#	Variable	Applic.	W_v	Type
Pre-Crash Motion				
1	Vehicle initial speed (if any reference or estimation)	V	4	K
2	Vehicle crash speed (if any reference or estimation)	V	4	K
3	Vehicle brake/throttle pedal & steering input state	V	2	K
4	VRU crash speed (if any reference or estimation)	A	4	K
5	VRU posture at impact (upright/running/bent)	A	2	O
6	Vehicle evasive action taken (steer, brake, none)	V	2	K
7	Pre-crash scenario type	B	3	E
Road & Environment				
8	Road type (motorway, urban, rural…)	B	2	E
9	Road surface type (asphalt, gravel…)	B	2	E
10	Road condition (dry, wet, snow, ice)	B	2	E
11	Light condition (daylight, dusk, night, street lamp)	B	1	E
12	Weather (rain, fog, snow, clear)	B	1	E
13	Speed limit at location	B	2	E
14	Lane separation type (barrier, markings, none)	B	1	E
15	Location type (intersection, mid-block, roundabout)	B	2	E
16	GPS coordinates of crash location	B	3	E
Physical Scene & Kinematics				
17	Road surface friction coefficient μ	B	2	D
18	Longitudinal gradient / slope	B	1	E
19	Sight-line / available visibility distance to POI	B	2	E
20	Skid / braking mark length and position on road	B	3	D
21	VRU pre-crash marks on road	A	2	D
22	Point of impact (POI) location on road	B	3	D
23	VRU final rest position on road	A	3	D
24	VRU first ground contact point	A	2	D
25	VRU contact location on vehicle	A	3	D

continued on next page

Table 3.4 continued.

#	Variable	Applic.	W_v	Type
26	VRU transportation tool final rest position	A	2	D
27	Vehicle post-impact travel distance and rest position	V	3	D
28	Vehicle post-impact heading (direction change)	V	2	K
Vehicle Properties				
29	Vehicle make / model / year	V	3	V
30	Vehicle curb weight	V	2	V
31	Gross vehicle weight	V	2	V
32	Vehicle length, width, height	V	2	V
33	ABS / ESC system activation	V	2	V
34	Active safety system activation	V	2	V
35	Tire type & condition	V	1	V
CDC/TDC — Primary Collision				
36	Principal direction of force (PDOF) — clock position	B	3	D
37	Area of deformation on vehicle body	B	3	D
38	Specific longitudinal/lateral location of damage	V	2	D
39	Specific vertical location of damage (belt-line, undercarriage)	V	2	D
40	Type of damage distribution (wide/narrow/sideswipe/rollover)	V	2	D
41	Maximum extent of penetration — zone number	V	2	D
42	Maximum extent of penetration (% of vehicle dimension)	V	2	D
CDC/TDC — Secondary Collision				
43	PDOF secondary collision	B	2	D
44	Area of deformation secondary collision	B	2	D
45	Longitudinal/lateral location secondary	V	2	D
46	Vertical location secondary	V	2	D
47	Type of damage distribution secondary	V	2	D
48	Maximum extent of penetration secondary — zone	V	2	D
49	Maximum extent of penetration secondary (%)	V	2	D
VRU Properties				
50	VRU body height	A	2	O
51	VRU body mass	A	2	O
52	VRU type (pedestrian/cyclist/wheelchair/e-scooter...)	A	3	O
53	VRU age	A	2	O
54	VRU gender	A	1	O
55	Helmet use (active traveler)	A	2	O
Occupant Properties (motor vehicle)				
56	Occupant type (driver, front passenger, rear passenger)	V	3	O
57	Occupant age	V	2	O
58	Occupant gender	V	1	O
59	Occupant body mass (kg)	V	2	O
60	Occupant height (m)	V	2	O
61	Seat belt fitted (yes/no/type)	V	2	V
62	Seat belt in use at time of crash	V	3	O
63	Front airbag fitted	V	2	V
64	Front airbag deployed	V	2	V
65	Side/curtain airbag fitted	V	2	V
66	Side/curtain airbag deployed	V	2	V
67	Knee airbag fitted	V	1	V
68	Knee airbag deployed	V	1	V

continued on next page

Table 3.4 continued.

#	Variable	Applic.	W_v	Type
69	Child seat present	V	2	V
70	Helmet use (motorcycle occupant)	V	3	O
Injury Severity				
71	Maximum AIS score (MAIS)	B	3	I
72	AIS regional coding	B	2	I
73	Police injury severity classification	B	1	I
Human Factors				
74	Driver perception–reaction time (PRT)	V	1	K
75	Time-to-collision (TTC)	B	2	K

3.3.3 Designed components

A browser-based interactive application is built with the Plotly Dash [71] framework (`dashboard.py`). Users upload one or more filled Excel templates via a drag-and-drop interface; upon upload, a modal dialogue asks whether the submission represents a single *case* (one crash event) or an aggregated *dataset* (a template whose cells summarize variable coverage across multiple events). For dataset submissions the user additionally selects the crash-type composition from three options: V-V only, V-A only, or Mixed; for case submissions, crash type is auto-detected from the uploaded template (Section 3.3.1.3) with an optional manual override (Figure 3.7).

Same as in Section 3.2.1, all illustrations of the display page designs in this section are based on the PAV data as an example. Figures demonstrating whole-page layouts for this section are collected in Appendix B. Figure B.1 presents the layout of the complete initial web page of the WCRS dashboard.

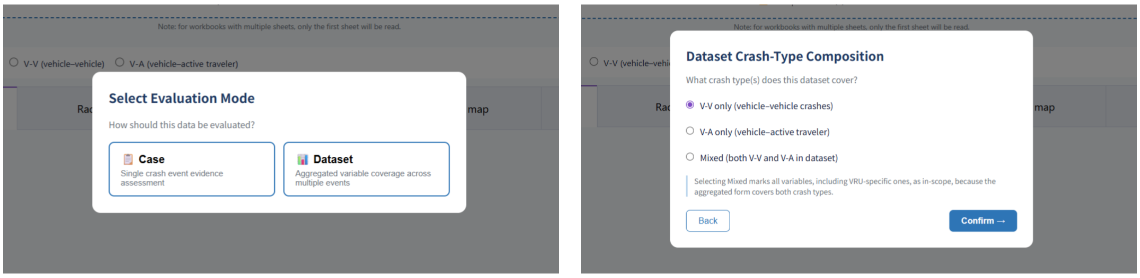
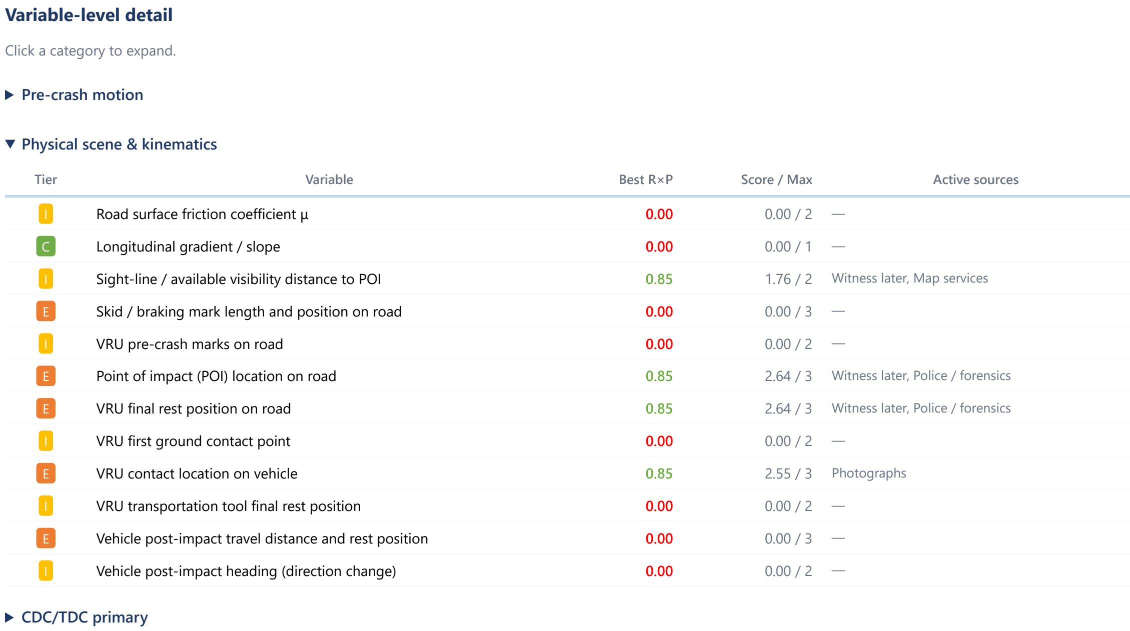


Figure 3.7: Web pop-up windows for evaluation mode and crash type selection (Dataset mode) of IGLAD Compatibility Assessment Dashboard.

The WCRS dashboard consists of the following components:

- Scorecard
- Radar chart
- Summary table
- Coverage map
- Source drill

The Scorecard tab (Figure B.2) presents the headline composite score together with the crash type and minimum category score displayed beneath it, and two diagnostic pills indicating the count of uncovered Tier-S and Tier-E variables, highlighted in red when non-zero. A *Category scores* panel below lists all ten categories, each with a color-coded progress bar (green at $\geq 80\,\%$, amber at $\geq 50\,\%$, red below $50\,\%$) and the category weight. A collapsible *Variable-level detail* panel organizes variables by category; expanding a category reveals a table with four columns: Tier, Variable, Best $R{\times}P$, Score / Max, and Active sources, where Active sources lists the names of all sources for which the variable is marked present and the source–type reliability is applicable (Figure 3.8).


Variable-level detail

Click a category to expand.

▶ Pre-crash motion

▼ Physical scene & kinematics

Tier	Variable	Best R×P	Score / Max	Active sources
I	Road surface friction coefficient μ	0.00	0.00 / 2	—
C	Longitudinal gradient / slope	0.00	0.00 / 1	—
I	Sight-line / available visibility distance to POI	0.85	1.76 / 2	Witness later, Map services
E	Skid / braking mark length and position on road	0.00	0.00 / 3	—
I	VRU pre-crash marks on road	0.00	0.00 / 2	—
E	Point of impact (POI) location on road	0.85	2.64 / 3	Witness later, Police / forensics
E	VRU final rest position on road	0.85	2.64 / 3	Witness later, Police / forensics
I	VRU first ground contact point	0.00	0.00 / 2	—
E	VRU contact location on vehicle	0.85	2.55 / 3	Photographs
I	VRU transportation tool final rest position	0.00	0.00 / 2	—
E	Vehicle post-impact travel distance and rest position	0.00	0.00 / 3	—
I	Vehicle post-impact heading (direction change)	0.00	0.00 / 2	—

▶ CDC/TDC primary

Figure 3.8: Click-to-expand in Variable-level detail section of Source drill interface of IGLAD Compatibility Assessment Dashboard (result of PAV dataset presented).

The Radar chart tab (Figure B.3) renders a filled spider chart in which the ten radial axes correspond to the ten category percentage scores. A drop-down labeled *Display cases* allows the user to select one or more loaded submissions for overlay; multiple traces are drawn on the same chart for side-by-side visual comparison of coverage profiles. In addition, when the crash data type is selected as "V-V", the category "VRU properties" is not applicable for evaluation and is excluded from the chart; when the user views radar charts of data with different coverage ranges simultaneously, two separate charts will be displayed.

The Summary table tab (Figure B.4) provides a compact tabular view that lists all loaded submissions. Each row shows a color-coded indicator dot, the file name, the detected crash type, the composite rating, and the ten individual category scores. When both Case and Dataset submissions are loaded, two separate tables are shown: one titled *Summary table - Case* and one titled *Summary table - Dataset*.

The Coverage map tab (Figure B.5) presents a heatmap with variables on the vertical axis and sources on the horizontal axis; each cell displays the effective contribution $R \times P$, with darker shading indicating more trustworthy evidence. Cells are interactive: hovering reveals the variable name, the full source name, the raw presence flag P, the reliability value R, the computed $R \times P$, and the in-scope status of the variable.

The Source drill tab (Figure B.6) displays a stacked bar chart showing, for each category, each source's independent contribution. The bars make it immediately visible which sources provide evidential coverage in each area and where coverage gaps lie. Each stacked segment represents the raw contribution of one source to the category's total variable weight, computed as tier $\times R \times P$, summed over all in-scope variables in the category and normalized by the sum of variable maxima:

$$\text{contrib}(s, c) = \frac{\sum_{v \in c,\, \text{in-scope}} W_v \cdot R(s,\, \text{type}(v)) \cdot P(v, s)}{\sum_{v \in c,\, \text{in-scope}} W_v} \tag{3.1}$$

It is worth noting here that because the composite score in Eq. (2.5) uses only the best and second-best source per variable, the per-source bars are not additive decompositions of the category score and may sum to more than the actual category percentage when multiple sources cover the same variable; this tab is therefore a per-source breakdown for diagnostic purposes and does not replicate the final scoring logic.

3.4 Sample assessments

In this section, the WCRS dashboard introduced in Section 3.3 is applied to multiple samples. Both results of assessing the readiness of crash data sources to harmonize with IGLAD (Section 3.4.1) and the variable quality of individual cases on supporting IGLAD-compatible reconstruction (Section 3.4.2) are presented.

3.4.1 PAV and existing IGLAD member dataset samples

For dataset samples, scoring is conducted using *Dataset* mode of the WCRS dashboard. As introduced in Section 2.1, PAV only includes vehicle-to-active traveler accidents. The "V-A" dataset type is selected when uploading. The overall composite and category score results of the PAV dataset using *Dataset* mode are shown in Figure 3.9.

It is worth noting that, in the Occupant properties variables centered on vehicle drivers and passengers, many of them show not applicable results. Specifically, in terms of the severity of "V-A" type accidents, it is extremely difficult for these accidents to cause injury or death to vehicle occupants. However, such accident variables hold a certain number of mandatory variables within the IGLAD framework, and that WCRS is also designed around the IGLAD compatibility assessment. Moreover, some driver-related information is also crucial and can be found in the PAV. Considering these factors, Occupant properties are included in PAV scoring.

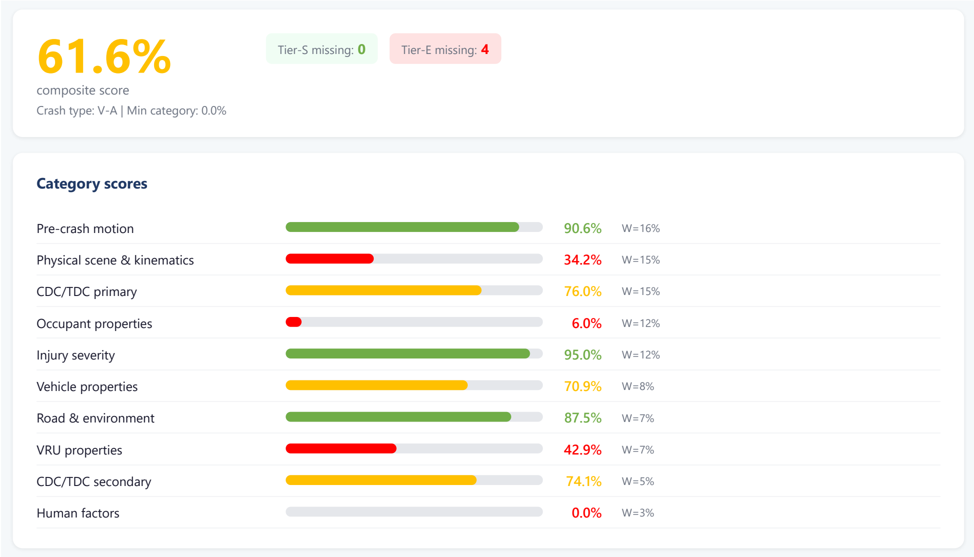


Figure 3.9: Overall composite and category score of PAV dataset.

By distributing the survey documents (questionnaire in Figure C.1 and reconstruction relevant variable grid in Figure C.2), information on the original datasets provided by multiple existing IGLAD members is collected.

The University of Florence in Italy provides information on the dataset it offers to IGLAD. During harmonization with IGLAD, reconstructions are conducted with simulation softwares. The dataset includes vehicle-to-VRU accidents; therefore, the "Mixed" dataset type is selected when uploading. The overall composite and category score results of the dataset are shown in Figure 3.10.

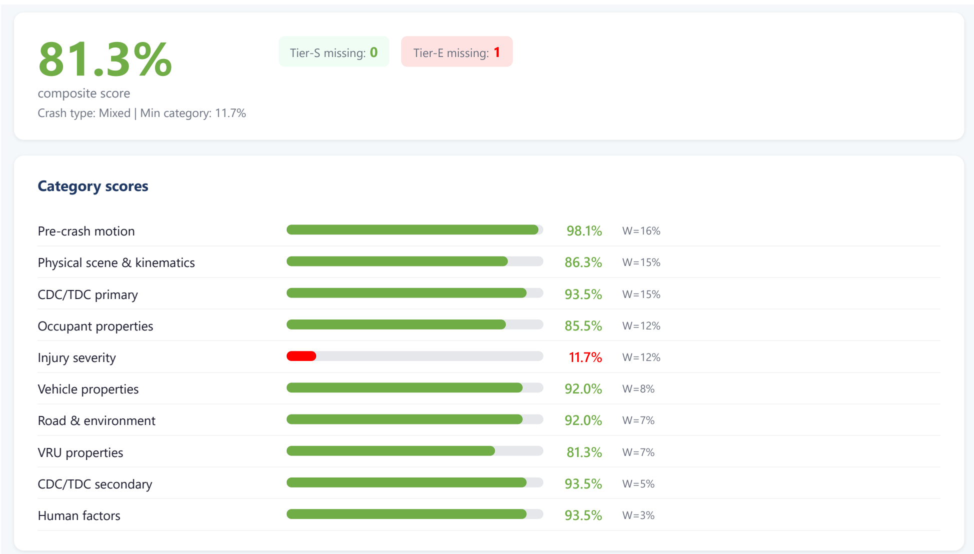


Figure 3.10: Overall composite and category score of IGLAD-IT dataset.

Asymptotic AB Sweden and NHTSA (National Highway Safety Administration) in the US provides information on the dataset they offer to IGLAD. The original dataset is CISS (Crash Investigation Sampling System). During harmonization with IGLAD, no reconstruction on any case is conducted. Currently, the dataset includes vehicle-to-VRU accidents, and VRU related accidents will be available from IGLAD release year 2027. Therefore, the “Mixed” dataset type is selected when uploading. The overall composite and category score results of the dataset are shown in Figure 3.11.

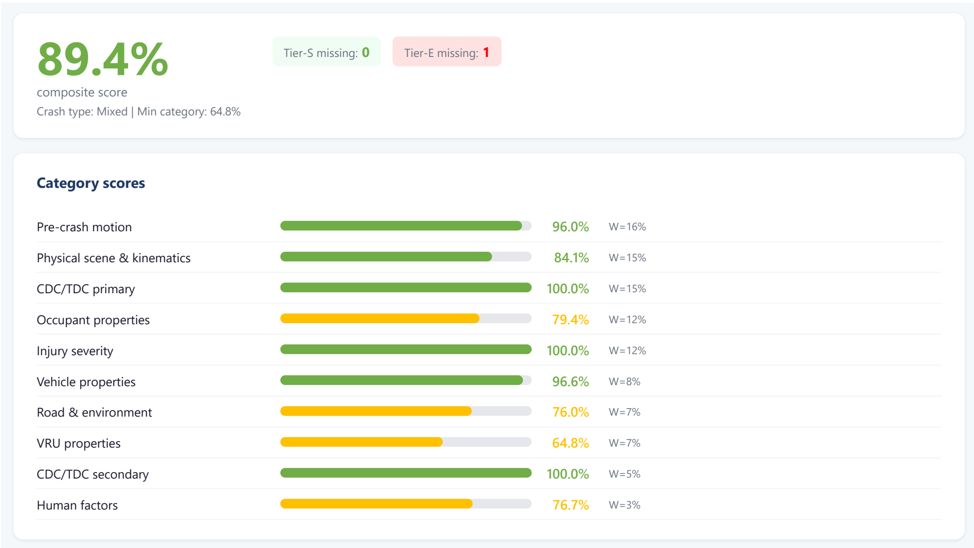


Figure 3.11: Overall composite and category score of IGLAD-US dataset.

Information of data subset provided from National Automobile Accident In-Depth Investigation System from China is collected. Data providers include multiple universities and traffic accident forensic centers. During harmonization with IGLAD, reconstructions are conducted with simulation softwares. The dataset includes vehicle-to-VRU accidents; therefore, the “Mixed” dataset type is selected. The overall composite and category score results of the dataset are shown in Figure 3.12.

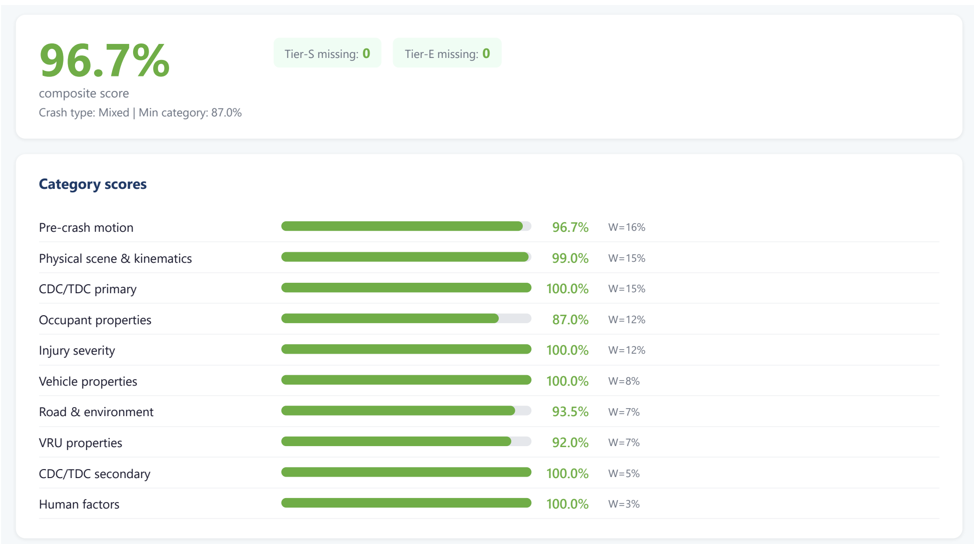


Figure 3.12: Overall composite and category score of IGLAD-CN dataset.

The category score distributions of the three datasets are further visualized as radar charts in Figure 3.13, where the shape and area of each trace reflect the overall coverage profile and allow category-level strengths and weaknesses to be identified.

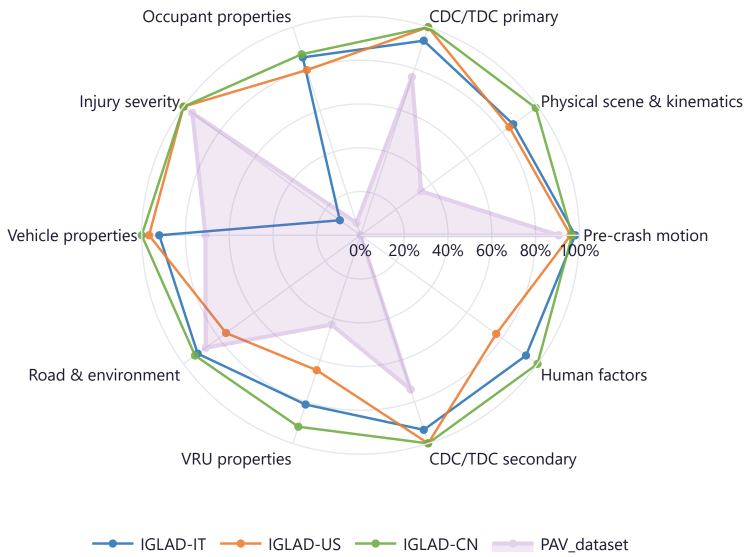


Figure 3.13: Assessment scoring result of the IGLAD member datasets and PAV in form of radar chart.

The composite and category scores of all three datasets are consolidated in Figure 3.14, providing a direct comparison of the overall rating and the ten individual category scores across PAV, IGLAD-IT, IGLAD-US and IGLAD-CN datasets.

Summary table

File	Type	Rating	Pre-crash	Physical s	CDC/TDC pr	Occupant p	Injury sev	Vehicle pr	Road & env	VRU proper	CDC/TDC se	Human fact
PAV_dataset	V-A	61.6%	91%	34%	76%	6%	95%	71%	87%	43%	74%	0%
IGLAD-IT	Mixed	81.3%	98%	86%	94%	85%	12%	92%	92%	81%	94%	94%
IGLAD-US	Mixed	89.4%	96%	84%	100%	79%	100%	97%	76%	65%	100%	77%
IGLAD-CN	Mixed	96.7%	97%	99%	100%	87%	100%	100%	94%	92%	100%	100%

Figure 3.14: Summary of assessment scoring result of the IGLAD member datasets and PAV.

3.4.2 PAV case samples

Seven case samples covering various vehicle-to-active traveler crash scenarios are provided by If to further demonstrate the variable quality and coding coverage of specific cases within the PAV dataset. For these cases, scoring is conducted using *Case* mode of the WCRS dashboard. The radar charts showing category score distributions of the seven cases in comparison with overall PAV dataset are in Figure 3.15; the overall rating and ten individual category scores across seven PAV cases are shown in Figure 3.16.

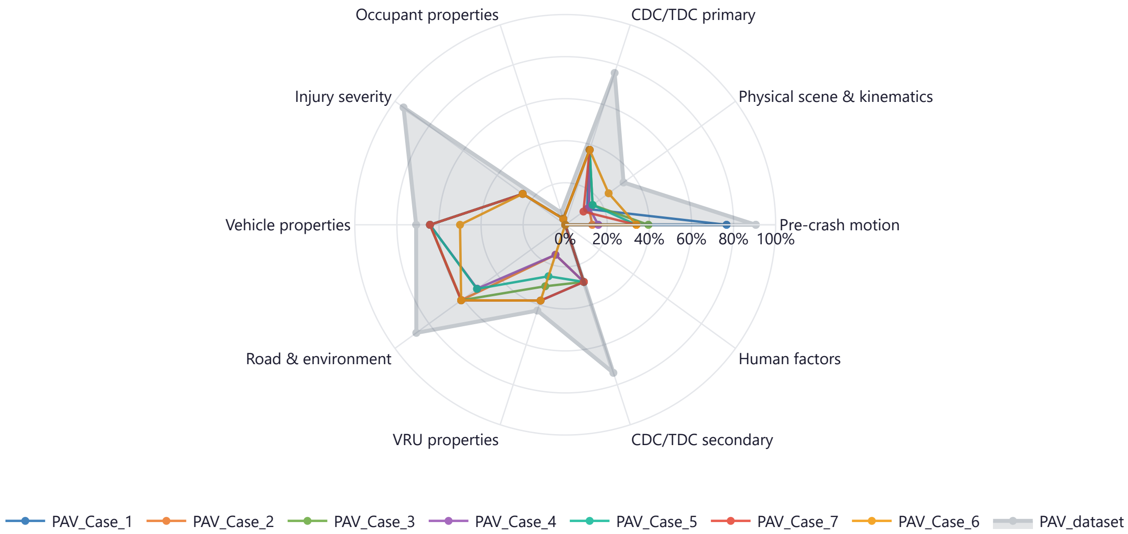


Figure 3.15: Assessment scoring result of the PAV cases in comparison with overall PAV dataset in form of radar chart.

Summary table - Case

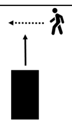
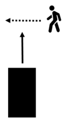
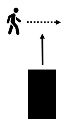
	File	Type	Rating	Pre-crash	Physical s	CDC/TDC pr	Occupant p	Injury sev	Vehicle pr	Road & env	VRU proper	CDC/TDC se	Human fact
●	PAV_Case_1	V-A	36.7%	77%	13%	38%	3%	25%	64%	61%	38%	29%	0%
●	PAV_Case_2	V-A	25.0%	13%	14%	38%	3%	25%	64%	61%	15%	29%	0%
●	PAV_Case_3	V-A	30.7%	40%	16%	38%	3%	25%	64%	61%	31%	29%	0%
●	PAV_Case_4	V-A	24.8%	16%	14%	38%	3%	25%	64%	52%	15%	29%	0%
●	PAV_Case_5	V-A	28.8%	34%	16%	38%	3%	25%	64%	52%	26%	29%	0%
●	PAV_Case_6	V-A	29.2%	34%	26%	38%	3%	25%	50%	61%	38%	0%	0%
●	PAV_Case_7	V-A	29.5%	34%	11%	38%	3%	25%	64%	61%	38%	29%	0%

Figure 3.16: Variable quality score of PAV Cases 1 to Case 7.

3.5 Reconstruction feasibility of PAV case samples

This section presents the assessed results of the seven PAV case samples against the input requirements of the reconstruction methods applicable to vehicle-to-VRU crashes. Each case is assigned a score on a five-point scale reflecting both whether a reconstruction can be attempted and how uncertain the resulting output would be. A score of 5 denotes an evidence set that supports a low-uncertainty reconstruction using the most demanding methods; a score of 1 denotes a case in which no quantitative reconstruction is feasible. The assigned scores roughly aligns with the ranking of evaluation results assessed through the WCRS system and displayed by the dashboard's *Case* mode above in Section 3.4.2. Table 3.5 provides a brief overview of the conflict situation of the crash events in each case sample.

Table 3.5: Overview of seven PAV case samples with conflict situation pictograms [53] and reconstruction feasibility scores.

Case No.	**VRU type**	**Conflict situation**	**Reconstruction feasibility score**
1	Pedestrian		4 / 5
2	Cyclist		2 / 5
3	Cyclist		3 / 5
4	Cyclist		3 / 5
5	Cyclist		3 / 5
6	E-scooter rider		3 / 5
7	Cyclist		3 / 5

Several conditions hold uniformly across the dataset and set the upper bound on achievable reconstruction quality. There is no variable for VRU body mass in the dataset. Since the dataset is retrospective, no case provides a measured skid mark length, a measured road surface friction coefficient μ, or a measured vehicle post-impact rest position. Furthermore, only rough hand-drawn sketches used for claims rather than dimensioned CAD files are available. The friction coefficient therefore needs to be assumed, and VRU body mass requires estimation from external statical anthropometric data.

3.5.1 Case 1: Score 4 / 5

Case 1 has the best information richness and quality, and is the only pedestrian-type VRU among all seven cases. Event Data Recorder (EDR) output is available, covering the vehicle initial speed, impact speed, and the brake, throttle and steering states at the time of the crash. Police records corroborate the pre-crash scenario type and the VRU type.

The available EDR data changes the reconstruction problem compared to the other six cases. Rather than seeking the vehicle impact speed as an unknown to be derived from the post-impact kinematics, the impact speed here is objectively instrument-derived. The reconstruction aims consequently shift: with the vehicle speed established, the central question becomes whether the photographed vehicle deformation pattern, the VRU contact areas, and the AIS regional injury distribution are mutually consistent with the known impact conditions.

The primary limitation of Case 1 is the absence of VRU status information. No source records the VRU movement speed or gait phase at the moment of impact, and no VRU body mass is available, though the latter is estimable from the recorded age. The expected uncertainty of reconstruction is the lowest of the seven cases, but the incomplete VRU status information prevents a score of 5.

3.5.2 Case 2: Score 2 / 5

Case 2 is the weakest in terms of information richness and quality. The vehicle initial speed and vehicle crash speed are missing. Police records are absent, and no VRU body status is available. What remains is a witness-stated point of impact, a witness-stated pre-crash scenario, vehicle parameters, road condition, VRU age and gender, and AIS-coded injuries.

Beyond missing inputs, the specific accident scenario (small-overlap rear impact with cyclist) in Case 2 produces vehicle deformation that is both ambiguous and relatively insensitive to vehicle speed, which means the photographed damage provides a limited constraint on impact speed even as a qualitative check. This is also the reason for the low feasibility score of this case, which is inconsistent with the WCRS evaluation results. The combination of a completely absent velocity prior and deformation evidence of low diagnostic value makes Case 2 hardly achievable. In such a situation, the estimated impact speed range from the reconstruction output could be unacceptable.

3.5.3 Case 3: Score 3 / 5

Case 3 has the most complete evidence profile among the non-EDR cases. Witness statements record both the vehicle initial speed and the vehicle crash speed, the point of impact, and uniquely the VRU body height. Police records corroborate the pre-crash scenario type and the principal direction of force.

However, two conditions substantially limit the achievable precision. Neither the VRU final rest position nor the vehicle rest position is recorded, so no possible throw distance can be established. The vehicle crash speed also depends on a single uncorroborated witness statement. Both conditions introduce uncertainty that cannot be reduced without additional data, and any velocity estimate must be presented as a wide interval rather than a point value. Additionally, for a cyclist-type VRU in this case, body height information can only be considered together with age and gender to estimate body weight, and it offers limited assistance in determining the center of gravity height.

3.5.4 Case 4: Score 3 / 5

Case 4 records a witness-stated vehicle crash speed and a witness-stated point of impact, but has no police records on any variable, no road condition, no vehicle initial speed, and no VRU anthropometry beyond age and gender.

The velocity prior rests on a single uncorroborated statement and should be treated as a hypothesis to be tested against the deformation and injury evidence rather than as a validated input. The absence of police records removes the only independent source that would otherwise have confirmed or contradicted the scenario description and the speed estimate. In addition, there are head injuries in this case, but the helmet information is missing, which is also a point deduction. The score of 3 is equal to that of Case 3 because the presence of both a stated crash speed and a point of impact compensates for the absence of police corroboration, leaving the overall reconstruction difficulty broadly comparable despite the different pattern of missing variables.

3.5.5 Case 5: Score 3 / 5

Case 5 has an evidence structure closely comparable to Case 3. Witness statements cover the vehicle initial speed, the vehicle crash speed and the point of impact. Police records corroborate the pre-crash scenario type and the principal direction of force.

The principal difference from Case 3 is the absence of a recorded VRU body height. The same constraints on the absence of rest positions and the uncorroborated velocity prior apply as in Case 3. The score of 3 reflects equal reconstruction feasibility, with a marginally wider uncertainty envelope relative to Case 3 due to the missing VRU anthropometry.

3.5.6 Case 6: Score 3 / 5

Case 6 has a distinct position in all seven cases: it is the only case for which the VRU final rest position is available, supported by both witness statements and police reports. The point of impact is also available from witness statement. Combined, these two variables allow the throw distance to be estimated as the distance between the stated impact location and the stated rest position — the only case outside of Case 1 where a throw distance can be approximated. A vehicle initial speed, road condition, and police corroboration of the pre-crash scenario, VRU type and principal direction of force are also present. Also, this is a minor crash which specifies that the crash occurred at low speeds. As the VRU in this case is a e-scooter rider, the helmet-use entry can be a plus.

Two factors are drawbacks to the score. Although relatively low impact speeds can reduce the absolute speed range and errors, the vehicle crash speed is not stated by any source, leaving the reconstruction without a velocity-reference initial input. Additionally, the available vehicle damage photographs are limited in quantity, and show only minor deformation without capturing any VRU contact traces on the vehicle surface (e.g. dust or grease from skin), which reduces their value as a constraint on the impact conditions.

3.5.7 Case 7: Score 3 / 5

Case 7 records a witness-stated vehicle initial speed but no vehicle crash speed, and no source provides a point of impact. Police records corroborate the pre-crash scenario, VRU type and principal direction of force. As with Case 6, helmet use is recorded, which can help match the impact of head injuries. AIS-coded injuries are available from medical records.

With neither a throw distance nor a velocity prior available, any quantitative reconstruction would require both to be assumed or inferred from the scenario description and the deformation photographs. The categorical evidence (deformation zone, VRU status, AIS injury pattern and scenario type) is complete and internally consistent, but every dynamic variable entering a reconstruction must be estimated.

3.5.8 Summary of evaluation of PAV cases

The scores range from 2 to 4, no case reaches a score of 5, which would require independently measured on-site scene geometry, an instrument-derived velocity reference, and fully documented VRU anthropometry — conditions characteristic of dedicated in-depth investigations but not of routine insurance claims data.

For crash cases in which the primary impacts can be classified as collisions with a wrap trajectory, established reconstruction methods are available ranging from simple analytical throw-distance models to three-dimensional multibody simulation [72]. However, the application of these methods presupposes that key inputs, above all the throw distance and a velocity reference, are grounded in physical evidence. Notably, even when such inputs are rigorously established from measured scene data, semi-analytical regression methods have yielded velocity intervals as wide as 19 km/h [72]; this suggests that for Cases 2 to Cases 7, where comparable inputs either rely on single uncorroborated witness statements or are absent entirely, any attempted reconstruction would be subject to considerably greater uncertainty. Any case scoring below 4, if subjected to reconstruction, would carry high uncertainty in the resulting velocity estimate, and this uncertainty increases as the score decreases.

4

Discussion

This study takes the PAV crash database together with a subset of it as examples to analyze the harmonization feasibility between insurance crash data and the IGLAD common data scheme. The research content covers aspects such as structural mapping coverage analysis, exploration of quantitative data quality assessment methods, and evaluation of datasets and case samples, aiming to identify the gaps between PAV and IGLAD at multiple levels. The research outcomes include not only the results of PAV compatibility assessments with IGLAD, but also the proposal and design of digital tools that support broader research. Through a review of the methods proposed in this project and the results obtained, several findings have emerged. This chapter discusses these findings, distinguishes between certain design concepts, and points out the limitations of the project and expectations for future work.

4.1 Gaps between PAV and IGLAD

Based on the results of sample assessments presented in Section 3.4.1, the gaps between the PAV crash database and the IGLAD common data scheme are identified. The evaluation conducted using WCRS showed that the PAV dataset scored 61.6 on a 100-point scale in the overall composite. For comparison, all three existing IGLAD member datasets scored above 80. Tier-S (Speed-critical, $W_v = 4$) variables are not missing in the structure of all four datasets, however, the absence of Tier-E (Essential, $W_v = 3$) variables in PAV is more severe than in the other three datasets. It can be inferred from Figure 3.13 that, except for the member dataset IGLAD-IT performing poorly on variables in the injury severity category, the three datasets have advantages in almost all other aspects. Therefore, as a new data source, the gap between PAV and IGLAD is transparent. The current readiness of harmonizing PAV with IGLAD is also relatively low.

As presented in Figure 3.9, although PAV is highly fine-grained in Road & environment, Pre-crash motion and VRU Injury severity, it is poor in variables related to occupant properties and human factors; since PAV only has a single source (photograph) and lacks the Maximum extent of penetration (% of vehicle dimension), its CDC/TDC category scores are at a moderate level. The structural limitations of PAV are evident: due to its focus on vehicle-active traveler accidents, many variables related to vehicle occupants, vehicle passive safety systems, and crash dynamics in IGLAD are naturally absent from PAV. This results in structural gaps in consistency for certain IGLAD categories (such as OCCUPANT and SAFETYSYSTEM).

These variables are excluded when analyzing mapping coverage and its ideal upper bound; however, since WCRS is a widely applicable compatibility assessment tool for identifying general gaps, they are not excluded from the scope of the assessment. This approach affects the compatibility score of PAV, but it does not indicate that its data quality is similarly deficient.

In addition, through analyzing the results in Section 3.4.2, some given sample PAV cases (such as Case 1) can achieve a relatively high IGLAD-compatible reconstruction feasibility score. Most cases show varying degrees of missing or inferred key input variables (such as information related to speed and VRU properties). There is also a considerable gap between the seven cases and the design of their original dataset's structure (Figure 3.15), indicating that that even some variables are designed in the PAV, they are not showing valid codes in the case samples. This variability indicates that data quality is not evenly distributed among cases, but largely depends on the completeness of the case narrative, the richness of evidence sources, and the availability of supplementary materials.

The differences between PAV and IGLAD in variable definition and coding logic are also revealed. In Section 2.3, the method demonstrates that multiple variables in PAV need to be combined, inferred, or mapped through rules to correspond to a single variable in IGLAD (such as ACCDESC, ACCTYPE, ROADTYPE). At the sample level, this structural difference could lead to information fragmentation: although the information exists in PAV, it is stored in a form that differs significantly from the form required by IGLAD, thereby reducing the consistency and feasibility of conducting direct mapping of the sample.

4.2 Recommendations for data collection practice

There are several areas for optimization in PAV data collection practices; therefore, this section proposes several recommendations for reference.

Although the structure of the PAV dataset includes corresponding variables containing VRU (active traveler) final rest position information, in the analysis of the case samples obtained for this project, it is one of the most critical missing variables. In crashes where the VRUs are thrown due to impacts, without this positional information, it is unable to determine the throwing distance, and consequently, the primary geometric inputs required for standard reconstruction methods cannot be obtained, greatly increasing the difficulty and uncertainty of reconstruction. Insurance accident investigations are retrospective in nature. Without introducing a separate workflow for new on-site scene records, case analysts or designated investigators should inquire about and record this location information as standard fields when interviewing the claimant shortly after the crash, or when police records are available. Ideally, fixed landmarks should be referenced using other available information (e.g. map services). In addition, dimensioned sketches are expected to be drawn.

Another critical missing variable arises from the absence of a key variable in terms of the dataset structure. In the PAV dataset, the information of body mass of VRU is unavailable. Although this can be estimated using external statistical anthropometric data when the body height is available, there are also frequent cases in the analyzed data where VRU body height information is coded unknown. Given that the mass and height of VRU are essential inputs for reconstruction, and that obtaining these two variables from the claimant or medical records during injury assessment can be not particularly difficult, it is therefore recommended to add VRU height and mass as mandatory fields in the claim application form for this specific type of accidents.

An outstanding feature of the PAV dataset is that all cases provide photographs of the damaged parts of the vehicles. However, none of the photographs include embedded measurement scales currently; some photographs cover only localized areas, limited to scratches or dents only, making the global positional information on larger vehicle components (e.g. hood or roof) relatively unclear. A simple recommendation would be to require that at least one photograph includes a measurement scale, and optimize photographing to systematically capture both broader areas and detailed views of vehicle collision marks, thereby transforming the recorded photographs from qualitative indicators into quantitative inputs which can be directly usable for in-depth investigations.

Case 1 demonstrates the transformative influence that available objective on-board vehicle data can have on the feasibility of reconstruction: once the vehicle kinematic data is determined through instrumentation, the reconstruction problem is simplified from speed output estimation to consistency verification. Modern vehicles are increasingly equipped with various on-board sensing devices, but data is only recorded under specific circumstances, for example, the EDR information available in Case 1 is most likely triggered by the activation of the pedestrian protection system. In intelligent connected vehicles, on-board logs could potentially be recorded continuously, but obtaining access to them would require permission from the manufacturer or traffic authorities. Currently, If in Norway has implemented a "Usage-based car insurance" program, which assesses driving behavior through a chip installed in the vehicle and a phone application, providing insurance pricing discounts and automatic accident notification services [73]. A feasible suggestion is to extend this on-board sensing telematics technology to information acquisition in in-depth accident investigations. Rich objective vehicle kinematic data would greatly increase the proportion of cases in the dataset available for reconstruction.

4.3 WCRS: case mode versus dataset mode

The dual evaluation mode options on the dashboard reflect two distinct practical use cases.

In *Case* mode, a single filled template represents one specific crash case; the WCRS score quantifies the evidential completeness of that specific case and shows which variables are missing, whether it is that case information is insufficient and therefore coded as unknown, or that variables are originally missing from the dataset it comes from. In this mode, the system is essentially performing reconstruction input variable assessment, which can be applied to prospective data quality checks at an early stage of an ongoing investigation. The results can be referenced as an aid for the feasibility analysis of conducting reconstruction, however, as mentioned in Section 3.5, due to the presence of case-specific characteristics that are difficult to quantify uniformly, this cannot replace an in-depth, detailed, and specific feasibility study.

In *Dataset* mode, the uploaded template is assumed to have been pre-aggregated by the submitting organization: a cell should be marked present if the variable appears in at least one case in the collection. The current implementation does not automatically aggregate multiple case-level templates into a dataset-level template. The Mixed crash type, available only in dataset mode, ensures that the full variable framework including VRU-specific rows is evaluated, which is appropriate when the dataset covers both V-V and V-A events. This distinction matters because the category weight for VRU Properties (7 %) is redistributed among active categories in V-V-only assessments. Declaring Mixed when the dataset genuinely contains both types prevents score inflation, therefore, a question is included in the survey questionnaire that requires to be completed by the organizations providing datasets.

4.4 Limitations

In this project, only partial detailed mapping and recoding work was completed for the framework-level variable harmonization of the PAV and IGLAD datasets. This project has detailed the feasibility of mapping PAV to IGLAD at the data framework-level and explored the methods; however, the complete matching and recoding work still requires further analysis and more effort.

In accident investigations, even if the information and sources of variables are both the same, there may still be differences in quality. For example, whether the variable information from a witness statement is certain (high quality) or vague (low quality). In the current WCRS design, the binary presence indicator cannot distinguish between them. Another limitation is that the current loader treats the Excel template primarily as a binary presence matrix. As a result, qualitative comments in the notes column, such as partial availability, conditional relevance, or explanations of why a variable is not applicable, do not directly affect the score. This design keeps the rating process transparent and reproducible, but it also means that some contextual judgment remains outside the automated calculation and is suggested to be considered during manual interpretation of the dashboard results.

The separation between primary and secondary CDC/TDC variables is retained in the WCRS mainly to reflect the coding structure used in the IGLAD Codebook, where vehicle damage information may be recorded separately for different collision events. In *Dataset* mode, this treatment is reasonable because the assessment focuses on whether the dataset structure is capable of representing such variables, including both primary and secondary collision fields. However, in *Case* mode, the current implementation is less case-sensitive: it does not include an explicit flag indicating whether a secondary collision actually occurred in a specific crash. Therefore, secondary-collision variables may remain part of the scoring denominator even when they are not meaningful for a particular case. This should be considered when interpreting single-case scores.

Currently, the variable information from sketches is classified according to its original source. For example, if a sketch draws the Point of Impact (POI) on road using information provided by a police report, and this information is ultimately obtained by reading the sketch, the source of the information is considered to be the police report. However, such logic does not take into account the availability of sketches in the evaluation, nor does it include a mechanism for assessing sketch quality, whereas sketch quality is important for harmonizing with IGLAD but is hard to be automatically detected and quantified.

Unfortunately, we cannot draw conclusions about how representative the seven analyzed case samples are of the entire PAV dataset. Although this study has proposed recommendations, the optimal data collection strategy for future practice remains to be explored. In addition, as more objective data are collected, whether reconstruction will remain a necessary option is beyond the outcomes of this project and require further consideration. For example, even improved into standard fields, the VRU final rest position on the road remains extremely difficult to capture when the crash is not observed; however, when more objective data from CCTV, dashcam, or body-cam footage become available, this crucial reconstruction input variable may become less critical.

4.5 Proposal for future work

This project has contributed to the field of crash data harmonization, particularly those involving accidents between vehicles and VRUs. The deliverables are expected to be provided to more potential new data providers of IGLAD, but only after review and refinement by professionals, subsequent researchers, and accident analysts. Future research work that can be carried out for this project is proposed here.

Rigorous empirical calibration that compares reconstruction results when different source subsets are available with a ground truth from high-quality in-depth investigations would provide a stronger foundation for the reliability weights. Future extensions could allow for a graded presence score, replacing the binary presence indicator with a continuous value reflecting data granularity. Future versions could include a case-level secondary-collision applicability flag to handle the distinction with greater precision when secondary-collision variables are not meaningful for a particular case. Structured note fields or controlled labels could also be introduced so that partially, uncertain, or conditionally applicable information can be reflected more formally in the scoring process. It is important and worthwhile to include sketches as a separate source in the evaluation system and attempt to encode their quality with variable settings for IGLAD. Finally, despite the uncertainty of case features, a method of assessing suitability for reconstruction can still be further explored, and be integrated into the *Dataset* mode of the WCRS dashboard. It is expected to assist the current manual, case-by-case analysis to a greater extent through automated and batch processing methods.

5

Conclusion

This project has examined the feasibility of harmonizing insurance crash data from the PAV database with the IGLAD common data scheme, through a methodology spanning variable-level mapping, structural harmonization, and case reconstruction readiness assessment. The mapping analysis demonstrated that effective coverage increases considerably once structurally inapplicable variables are excluded. The vast majority of inapplicable mapping results appear in the occupant-facing and passive safety system related variables; this reflects the inherent nature of vehicle-to-active-traveler accidents, rather than a fundamental deficiency of PAV. Several digital tools designed with broader applicability beyond PAV were also developed to support the harmonization process. The Weighted Compatibility Rating System (WCRS), as the vital methodological contribution, provides a transparent and reproducible evaluation process at both the dataset and case levels. Applied comparatively, the WCRS results position PAV below the level of existing IGLAD member datasets, with physical scene and kinematics, VRU properties, and human factors among the categories showing relatively large gaps. The case-level assessment further confirmed that insurance claim cases are generally constrained by the absence of measured scene geometry, objective velocity references, and VRU anthropometric data, with data quality unevenly distributed across cases. These findings indicate that the current gaps between PAV and IGLAD, while substantial, are not insurmountable. Targeted improvements in data collection practice have the potential to increase both mapping coverage and IGLAD-compatible reconstruction readiness, and to improve future in-depth accident data harmonization. This project supports the evaluation of potential member datasets beyond PAV, contributing to the ongoing efforts to expand the membership of IGLAD and diversify its data sources on a globally comparable basis.

Bibliography

[1] World Health Organization. (2023). *Global status report on road safety 2023.* World Health Organization, 2023. ISBN 978-92-4-008651-7.

[2] U.S. Department of Transportation. (2022). *The Safety Problem.* National Roadway Safety Strategy. transportation.gov/NRSS/SafetyProblem.

[3] U.S. Department of Transportation. (2023). *2023 Progress Report on the National Roadway Safety Strategy.* U.S. Department of Transportation.

[4] European Commission. (2025). *Road safety thematic report—Cyclists.* European Road Safety Observatory. Brussels: European Commission, Directorate-General for Mobility and Transport.

[5] Wegman, F., Aarts, L., & Bax, C. (2008). *Advancing sustainable safety: National road safety outlook for The Netherlands for 2005–2020.* Safety Science, 46(2), 323–343. doi: 10.1016/j.ssci.2007.06.013.

[6] Trafikverket. (2012). *Road Safety: Vision Zero on the Move.* Swedish Transportation Administration.

[7] U.S. Department of Transportation. (2022). *National Roadway Safety Strategy.* U.S. Department of Transportation.

[8] European Court of Auditors. (2024). *Reaching EU road safety objectives: Time to move up a gear* (Special Report 04/2024). European Court of Auditors.

[9] Olszewski, P., Szagała, P., Rabczenko, D., & Zielińska, A. (2019). *Investigating safety of vulnerable road users in selected EU countries.* Journal of Safety Research, 68, 49–57. doi: 10.1016/j.jsr.2018.12.001.

[10] Ptak, M. (2019). *Method to Assess and Enhance Vulnerable Road User Safety during Impact Loading.* Applied Sciences, 9, 1000. doi: 10.3390/app9051000.

[11] Organisation for Economic Co-operation and Development (OECD). (1998). *Safety of Vulnerable Road Users.* Scientific Expert Group on the Safety of Vulnerable Road Users (RS7), DSTI/DOT/RTR/RS7(98)1/FINAL.

[12] Constant, A., & Lagarde, E. (2010). *Protecting Vulnerable Road Users from Injury.* PLoS Medicine, 7(3), e1000228. doi: 10.1371/journal.pmed.1000228.

[13] Khayesi, M. (2020). *Vulnerable Road Users or Vulnerable Transport Planning?.* Frontiers in Sustainable Cities, 2, 25. doi: 10.3389/frsc.2020.00025.

[14] Silva, R. M., Azevedo, G. F., Berto, M. V. V., Rocha, J. R., Fidelis, E. C., Nogueira, M. V., Lisboa, P. H., & Almeida, T. A. (2025). *Vulnerable road user detection and safety enhancement: A comprehensive survey.* Expert Systems with Applications, 292, 128529. doi: 10.1016/j.eswa.2025.128529.

[15] Harper, W. W. (1958). *The physical circumstances of auto-pedestrian collisions.* Arizona Medicine, 15(4), 266–272.

[16] Du, B., Zhang, C., Sarkar, A., Shen, J., Telikani, A., & Hu, H. (2024). *Identifying factors related to pedestrian and cyclist crashes in ACT, Australia with an extended crash dataset.* Accident Analysis and Prevention, 207, 107742. doi: 10.1016/j.aap.2024.107742.

[17] Yuan, Q., Gao, Y., Zhu, J., Xiong, H., Xu, Q., & Wang, J. (2023). *Summarizing vehicle driving decision-making methods on vulnerable road user collision avoidance.* Digital Transportation and Safety, 2(1), 23–35. doi: 10.48130/DTS-2023-0003.

[18] Singh, P., Jami, M., Geller, J., Granger, C., Geaney, L., & Aiyer, A. (2022). *The impact of e-scooter injuries: A systematic review of 34 studies.* Bone & Joint Open, 3(9), 674–683. doi: 10.1302/2633-1462.39.BJO-2022-0096.R1.

[19] Zhou, N., Zeng, H., Xie, R., Yang, T., Kong, J., Song, Z., Zhang, F., Liao, X., Chen, X., Miao, Q., Lan, F., Zhao, W., Han, R., & Li, D. (2024). *Analysis of road traffic accidents and casualties associated with electric bikes and bicycles in Guangzhou, China: A retrospective descriptive analysis.* Heliyon, 10, e29961. doi: 10.1016/j.heliyon.2024.e29961.

[20] Cook, S., Stevenson, L., Aldred, R., Kendall, M., & Cohen, T. (2022). *More than walking and cycling: What is 'active travel'?* Transport Policy, 126, 151–161. doi: 10.1016/j.tranpol.2022.07.015.

[21] Pervaz, S., Bhowmik, T., & Eluru, N. (2022). *Integrating macro and micro level crash frequency models considering spatial heterogeneity and random effects.* Analytic Methods in Accident Research, 36, 100238. doi: 10.1016/j.amar.2022.100238.

[22] Kumar, P., Tiwari, G., & Paul, S. B. (2025). *Road safety studies at micro, meso, and macroscopic levels: A systematic review.* IATSS Research, 49, 10–26. doi: 10.1016/j.iatssr.2024.12.001.

[23] Ljung Aust, M. (2010). *Generalization of case studies in road traffic when defining pre-crash scenarios for active safety function evaluation.* Accident Analysis and Prevention, 42(4), 1172–1183. doi: 10.1016/j.aap.2010.01.006.

[24] Galer, M., Thomas, P., & Bradford, M. O. (1995). *Conducting in-depth accident studies.* Safety Science, 19(2–3), 119–124. doi: 10.1016/0925-7535(94)00013-S.

[25] Bakker, J., Jeppsson, H., Hannawald, L., Spitzhüttl, F., Longton, A., Tomasch, E. (2017). *IGLAD - International Harmonized In-Depth Accident Data.* Proceedings of the 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Paper No. 17-0248.

[26] Seeck, A., Gail, J., Sferco, R., Otte, D., Hannawald, L., Zwipp, H., & Bakker, J. (2009). *Development of the accident investigation and data handling methodology in the GIDAS project.*

[27] German In-Depth Accident Study (information page of the part commissioned by the FAT). *GIDAS – German In-Depth Accident Study.* gidas.org/start-en.html. Accessed March 24, 2026.

[28] Chen, Y., Buerger, C., Lin, M., Li, X., Labenski, V., Jin, H., Wang, H., Liu, Y., Ino, T., Feifel, H., Tan, T., & Chang, F. (2023). *Left-turn-across-path-from-opposite-direction accidents in China: CIDAS accident study.* Transportation Safety and Environment, 5(4), tdac070. doi: 10.1093/tse/tdac070.

[29] National Highway Traffic Safety Administration. (2026). *National Automotive Sampling System.* National Highway Traffic Safety Administration. nhtsa.gov/crash-data-systems/national-automotive-sampling-system. Accessed April 30, 2026.

[30] Zhang, F., Noh, E. Y., Subramanian, R., & Chen, C.-L. (2019, September). *Crash Investigation Sampling System: Sample Design and Weighting* (Report No. DOT HS 812 804). Washington, DC: National Highway Traffic Safety Administration.

[31] National Highway Traffic Safety Administration. (2026). *Crashworthiness Data System.* National Highway Traffic Safety Administration. nhtsa.gov/national-automotive-sampling-system/crashworthiness-data-system. Accessed April 30, 2026.

[32] National Highway Traffic Safety Administration. (2026). *Crash Investigation Sampling System.* National Highway Traffic Safety Administration. nhtsa.gov/crash-data-systems/crash-investigation-sampling-system. Accessed April 30, 2026.

[33] Herbert, G. (2024). *Overview of the 2023 Crash Investigation Sampling System* (Report No. DOT HS 813 667). Washington, DC: National Highway Traffic Safety Administration.

[34] Wisch, M., et al. (2016). *Accident analysis, NDS and project implications.* Deliverable D2.1, PROactive Safety for PEdestrians and CyclisTs, European Commission. cordis.europa.eu/project/id/634149. doi: 10.3030/634149.

[35] Initiative for the Global Harmonization of Accident Data. (2026). *IGLAD homepage.* iglad.org. Accessed January 31, 2026.

[36] IGLAD Technical Working Group. (2026). *Codebook (Member Year 2025, Phase V).* Initiative for the Global Harmonization of Accident Data (IGLAD).

[37] Liers, H., & Pett, L. (2024). *Road safety in a global perspective – Accidentology in Asia and Europe.* Verkehrsunfallforschung an der TU Dresden GmbH.

[38] Elawad, A., Murgovski, N., & Kovaceva, J. (2026). *Evaluating safety envelopes*

for bus-VRU interactions: Passenger and VRU perceived safety implications. Transportation Research Procedia, 95, 41–48. doi: 10.1016/j.trpro.2026.02.006.

[39] Youn, Y.-h., Lee, S., Kim, S. H., Lee, J. W., & Lee, K. (2015). *Establishment of Korean KIDAS under the limited accident related data.* api.semanticscholar.org/CorpusID:108869271.

[40] Lindman, M., & Kullgren, A. (2024). *Insurance data.* In *Traffic safety data* (Chapter 13, pp. 263–276). doi: 10.1049/PBTR028E_ch13

[41] Isaksson-Hellman, I., & Werneke, J. (2017). *Detailed description of bicycle and passenger car collisions based on insurance claims.* Safety Science, 92, 330–337. doi: 10.1016/j.ssci.2016.02.008.

[42] Insurance Institute for Highway Safety and Highway Loss Data Institute. (2026). *Insurance losses by make and model.* iihs.org/research-areas/auto-insurance/insurance-losses-by-make-and-model. Accessed May 15, 2026.

[43] Kuehn, M., Hummel, T., & Bende, J. (2009). *Benefit estimation of advanced driver assistance systems for cars derived from real-life accidents.* In *Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles (ESV)* (Paper No. 09-0317). Stuttgart, Germany: German Insurers Accident Research.

[44] SAE International. (2026). *Vehicle Crash Reconstruction: Principles and Technology.* SAE Mobilus, Course C1728, Instructor Led. saemobilus.sae.org/courses/vehicle-crash-reconstruction-principles-technology-c1728. Accessed May 1, 2026.

[45] Steffan, H. (2009). *Accident reconstruction methods.* Vehicle System Dynamics, 47(8), 1049–1073. doi: 10.1080/00423110903100440.

[46] Brach, R. M., & Brach, R. M. (1987). *A review of impact models for vehicle collision.* SAE Technical Paper 870048. doi: 10.4271/870048.

[47] Vangi, D., Begani, F., Spitzhüttl, F., & Gulino, M. S. (2019). *Vehicle accident reconstruction by a reduced order impact model.* Forensic Science International, 298, 426.e1–426.e11. doi: 10.1016/j.forsciint.2019.02.042.

[48] Duma, I., Burnete, N., Todoruţ, A., Cordoş, N., Danci, C.-C., & Terec, A. (2025). *Assessment of Road Vehicle Accident Approaches—A Review.* Vehicles, 7, 10. doi: 10.3390/vehicles7010010.

[49] Wang, J., Li, Z., Ying, F., Zou, D., & Chen, Y. (2022). *Reconstruction of a real-world car-to-pedestrian collision using geomatics techniques and numerical simulations.* Journal of Forensic and Legal Medicine, 91, 102433. doi: 10.1016/j.jflm.2022.102433.

[50] Sveriges riksdag. (1975). *Trafikskadelag (1975:1410).* Svensk författningssamling. Utfärdad December 15, 1975; amended through SFS 2024:1280. riksdagen.se. Accessed April 18, 2026.

[51] Isaksson-Hellman, I. (2012). *A study of bicycle and passenger car collisions based on insurance claims data.* Annals of Advances in Automotive Medicine, 56, 3–12.

[52] If.se. (2026). *Traffic research.* If P&C Insurance. if.se/en/about-if/sustainability/traffic-research. Accessed April 18, 2026.

[53] Jonsson, S., Lindman, M., Isaksson-Hellman, I. (2026). *If's People Around the Vehicle (PAV) crash database - Manual (English version).* Internal technical manual, If Skadeförsäkring AB (publ). (Original work established 2010; revised continuously 2010-2026).

[54] Association for the Advancement of Automotive Medicine. (2018). *Abbreviated Injury Scale: 2015 Revision* (6th ed.). Chicago, IL: Association for the Advancement of Automotive Medicine.

[55] SAE International Technical Standard. (2022). *Collision Deformation Classification.* SAE Standard J224_202205, Revised May 2022, Reaffirmed May 2011, Issued January 1971. doi: 10.4271/J224_202205.

[56] Trafikverket. (2026). *Enskilda vägar.* Trafikverket. trafikverket.se/resa-och-trafik/vag/Enskilda-vagar. Accessed April 27, 2026.

[57] Sveriges riksdag. (2007). *Vägmärkesförordning (2007:90).* Svensk författningssamling. Utfärdad March 8, 2007; amended through SFS 2025:268. riksdagen.se. Accessed April 27, 2026.

[58] Spitzhüttl, F., Petzold, M., & Liers, H. (2015). *PCM from iGLAD database.* FAT-Schriftenreihe, 280. Forschungsvereinigung Automobiltechnik e. V. (FAT).

[59] Gabauer, D. J., & Gabler, H. C. (2008). *Comparison of roadside crash injury metrics using event data recorders.* Accident Analysis and Prevention, 40(2), 548–558. doi: 10.1016/j.aap.2007.08.011.

[60] Vida, G., & Török, À. (2023). *Expected effects of accident data recording technology evolution on the identification of accident causes and liability.* European Transport Research Review, 15, 17. doi: 10.1186/s12544-023-00591-4.

[61] Vida, G., & Török, À. (2025). *Effects of developing data recording technologies on the reliability of accident reconstruction and liability determination.* European Transport Research Review, 17, 34. doi: 10.1186/s12544-025-00727-8.

[62] European New Car Assessment Programme. (2026). *Euro NCAP: The European New Car Assessment Programme.* euroncap.com. Accessed June 2, 2026.

[63] European New Car Assessment Programme. (2020). *Assessment protocol – Overall rating* (Version 9.0.1). Euro NCAP.

[64] Fryar, C. D., Gu, Q., Afful, J., Carroll, M. D., & Ogden, C. L. (2025). *Anthropometric reference data for children and adults: United States, August 2021–August 2023.* Vital and Health Statistics, Series 3(50), 1–28. doi: 10.15620/cdc/174595.

[65] Gabauer, D. J., & Gabler, H. C. (2006). *Comparison of delta-V and occupant impact velocity crash severity metrics using event data recorders.* Annual Proceedings of the Association for the Advancement of Automotive Medicine, 50, 57–71.

[66] Jurewicz, C., Sobhani, A., Woolley, J., Dutschke, J., & Corben, B. (2016). *Exploration of vehicle impact speed – injury severity relationships for application in safer road design.* Transportation Research Procedia, 14, 4247–4256. doi: 10.1016/j.trpro.2016.05.396.

[67] Lubbe, N., Wu, Y., & Jeppsson, H. (2022). *Safe speeds: fatality and injury risks of pedestrians, cyclists, motorcyclists, and car drivers impacting the front of another passenger car as a function of closing speed and age.* Traffic Safety Research, 2, 000006. doi: 10.55329/vfma7555.

[68] Moravcová, P., Bucsuházy, K., Zůvala, R., Semela, M., & Bradáč, A. (2024). *What should I use to calculate vehicle EES?* PLoS ONE, 19(2), e0297940. doi: 10.1371/journal.pone.0297940.

[69] Vangi, D., Cialdai, C., Gulino, M.-S., & Robbersmyr, K. G. (2018). *Vehicle Accident Databases: Correctness Checks for Accident Kinematic Data.* Designs, 2(1), 4. doi: 10.3390/designs2010004.

[70] Streamlit Inc. (2019). *Streamlit: A faster way to build and share data apps.* streamlit.io. Accessed May 16, 2026.

[71] Plotly Technologies Inc. (2015). *Collaborative data science.* Plotly Technologies Inc., Montreal, QC. plotly.com.

[72] Depriester, J.-P., Perrin, C., Serre, T., & Chalandon, S. (2005). *Comparison of several methods for real pedestrian accident reconstruction.* In *19th International Technical Conference on the Enhanced Safety of Vehicles (ESV) Proceedings* (Paper No. 05-0333).

[73] If.no. (2026). *Bruksbasert bilforsikring.* If. if.no/privat/forsikring/bilforsikring/bruksbasert-bilforsikring. Accessed May 12, 2026.

A

Appendix 1

A.1 IGLAD Variable Mapping Dashboard

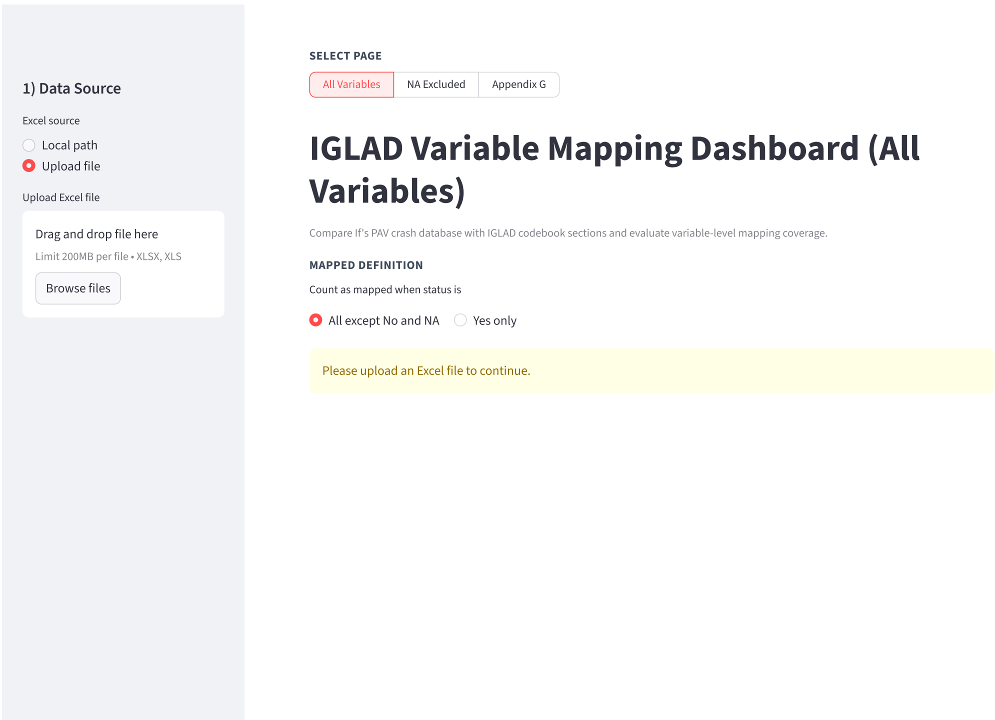


Figure A.1: Initial interface of the All Variables page of IGLAD Variable Mapping Dashboard with sidebar open.

SELECT PAGE

All Variables | NA Excluded | Appendix G

IGLAD Variable Mapping Dashboard (All Variables)

Compare If's PAV crash database with IGLAD codebook sections and evaluate variable-level mapping coverage.

MAPPED DEFINITION

Count as mapped when status is

◉ All except No and NA ○ Yes only

Overall KPIs

IGLAD total variables
133

Mapped (All except No and NA)
90

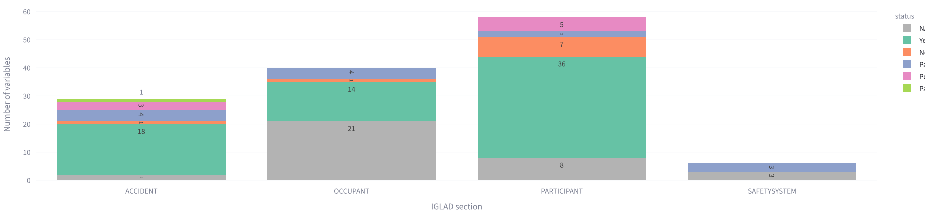
Mapped coverage
67.7%

Current mapped definition: All except No and NA

Coverage by IGLAD Section

Section	Target variables	Mapped (All except No and NA)	Mapped coverage	NA	Yes	No	Partly	Possibly	Partly/Possibly
ACCIDENT	29	26	89.7%	2	18	1	4	3	1
PARTICIPANT	58	43	74.1%	8	36	7	2	5	0
OCCUPANT	40	18	45.0%	21	14	1	4	0	0
SAFETYSYSTEM	6	3	50.0%	3	0	0	3	0	0

Status distribution by section



Status composition

Comments

Tips: Click a section to show/hide variable names; hover over a variable to view comment.

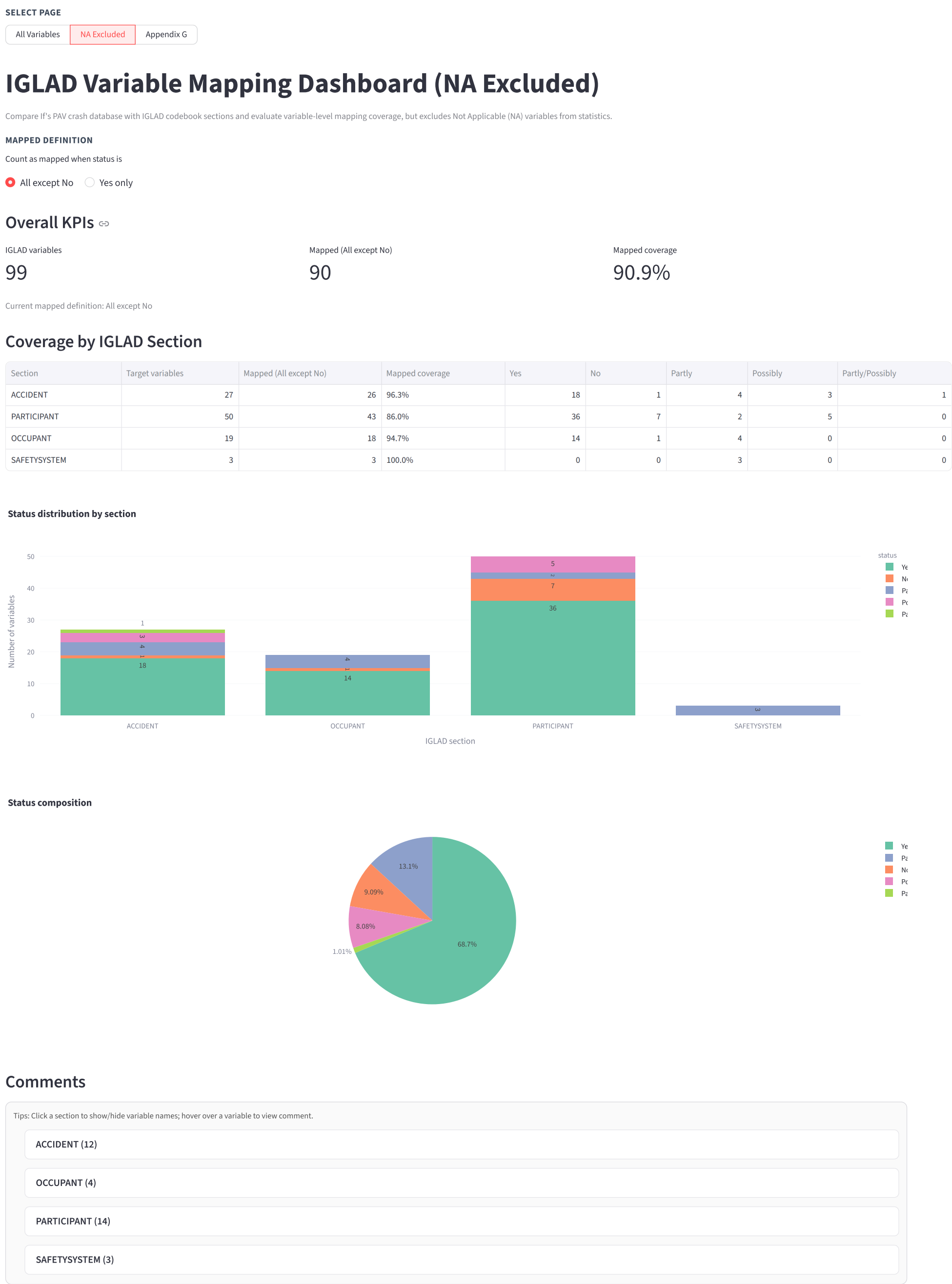
ACCIDENT (12)

OCCUPANT (4)

PARTICIPANT (14)

SAFETYSYSTEM (3)

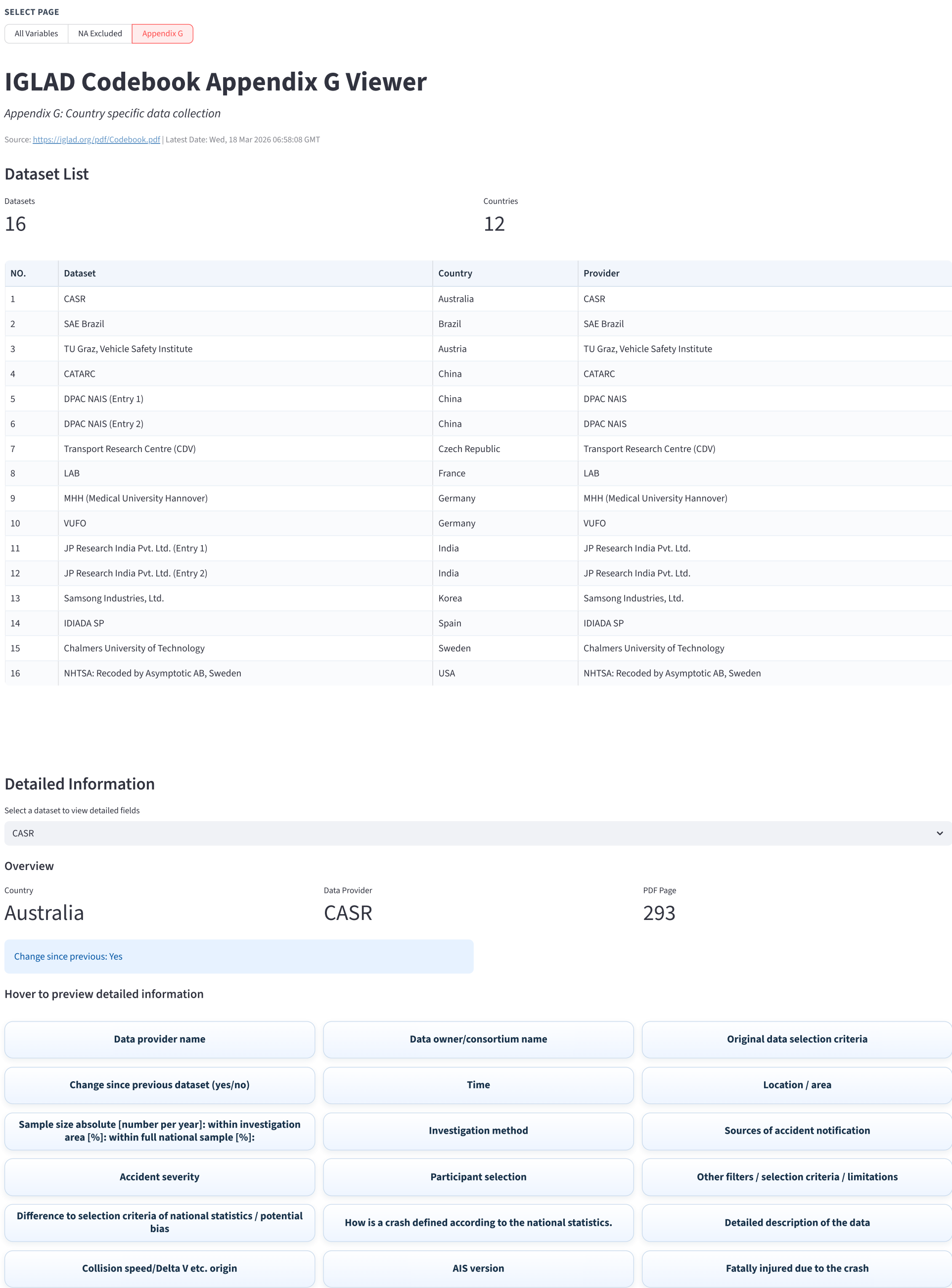
Figure A.2: All Variables page of IGLAD Variable Mapping Dashboard (result of PAV dataset presented).



Section	Target variables	Mapped (All except No)	Mapped coverage	Yes	No	Partly	Possibly	Partly/Possibly
ACCIDENT	27	26	96.3%	18	1	4	3	1
PARTICIPANT	50	43	86.0%	36	7	2	5	0
OCCUPANT	19	18	94.7%	14	1	4	0	0
SAFETYSYSTEM	3	3	100.0%	0	0	3	0	0

Figure A.3: NA Excluded page of IGLAD Variable Mapping Dashboard (result of PAV dataset presented).

A.2 IGLAD codebook Appendix G viewer



SELECT PAGE

All Variables | NA Excluded | Appendix G

IGLAD Codebook Appendix G Viewer

Appendix G: Country specific data collection

Source: https://iglad.org/pdf/Codebook.pdf | Latest Date: Wed, 18 Mar 2026 06:58:08 GMT

Dataset List

Datasets
16

Countries
12

NO.	Dataset	Country	Provider
1	CASR	Australia	CASR
2	SAE Brazil	Brazil	SAE Brazil
3	TU Graz, Vehicle Safety Institute	Austria	TU Graz, Vehicle Safety Institute
4	CATARC	China	CATARC
5	DPAC NAIS (Entry 1)	China	DPAC NAIS
6	DPAC NAIS (Entry 2)	China	DPAC NAIS
7	Transport Research Centre (CDV)	Czech Republic	Transport Research Centre (CDV)
8	LAB	France	LAB
9	MHH (Medical University Hannover)	Germany	MHH (Medical University Hannover)
10	VUFO	Germany	VUFO
11	JP Research India Pvt. Ltd. (Entry 1)	India	JP Research India Pvt. Ltd.
12	JP Research India Pvt. Ltd. (Entry 2)	India	JP Research India Pvt. Ltd.
13	Samsong Industries, Ltd.	Korea	Samsong Industries, Ltd.
14	IDIADA SP	Spain	IDIADA SP
15	Chalmers University of Technology	Sweden	Chalmers University of Technology
16	NHTSA: Recoded by Asymptotic AB, Sweden	USA	NHTSA: Recoded by Asymptotic AB, Sweden

Detailed Information

Select a dataset to view detailed fields

CASR

Overview

Country
Australia

Data Provider
CASR

PDF Page
293

Change since previous: Yes

Hover to preview detailed information

Data provider name | Data owner/consortium name | Original data selection criteria

Change since previous dataset (yes/no) | Time | Location / area

Sample size absolute [number per year]: within investigation area [%]: within full national sample [%]: | Investigation method | Sources of accident notification

Accident severity | Participant selection | Other filters / selection criteria / limitations

Difference to selection criteria of national statistics / potential bias | How is a crash defined according to the national statistics. | Detailed description of the data

Collision speed/Delta V etc. origin | AIS version | Fatally injured due to the crash

Figure A.4: IGLAD codebook Appendix G viewer.

B

Appendix 2

B.1 WCRS (IGLAD Compatibility Assessment) Dashboard

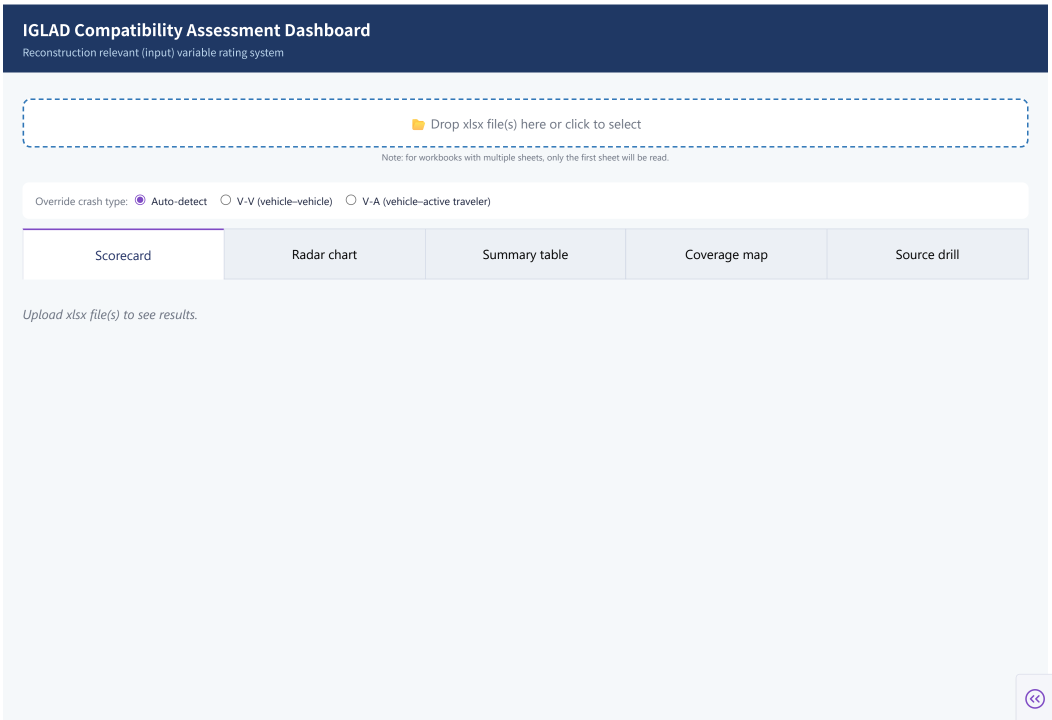


Figure B.1: Initial page status of IGLAD Compatibility Assessment Dashboard.

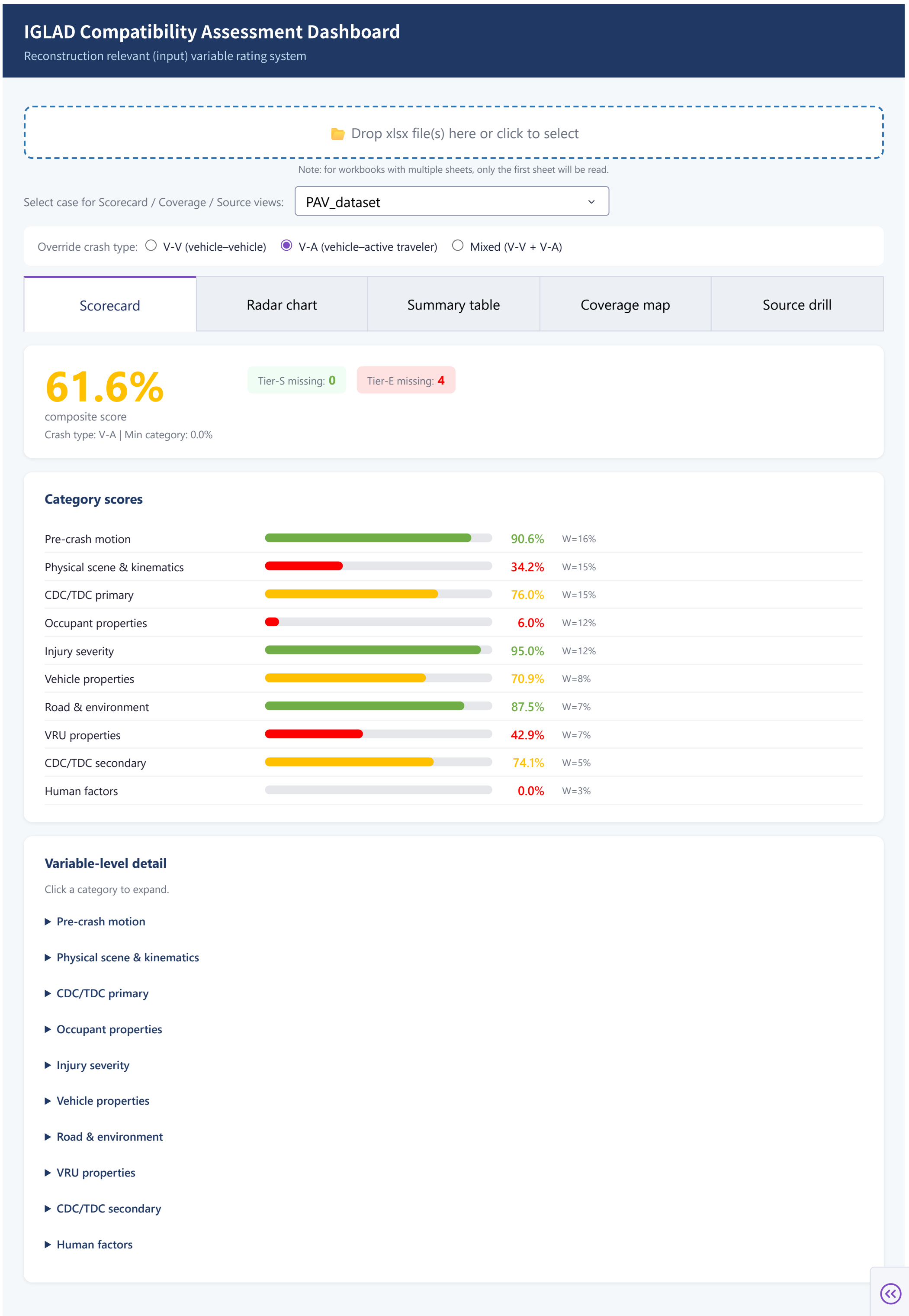


Figure B.2: Scorecard interface in IGLAD Compatibility Assessment Dashboard (result of PAV dataset presented).

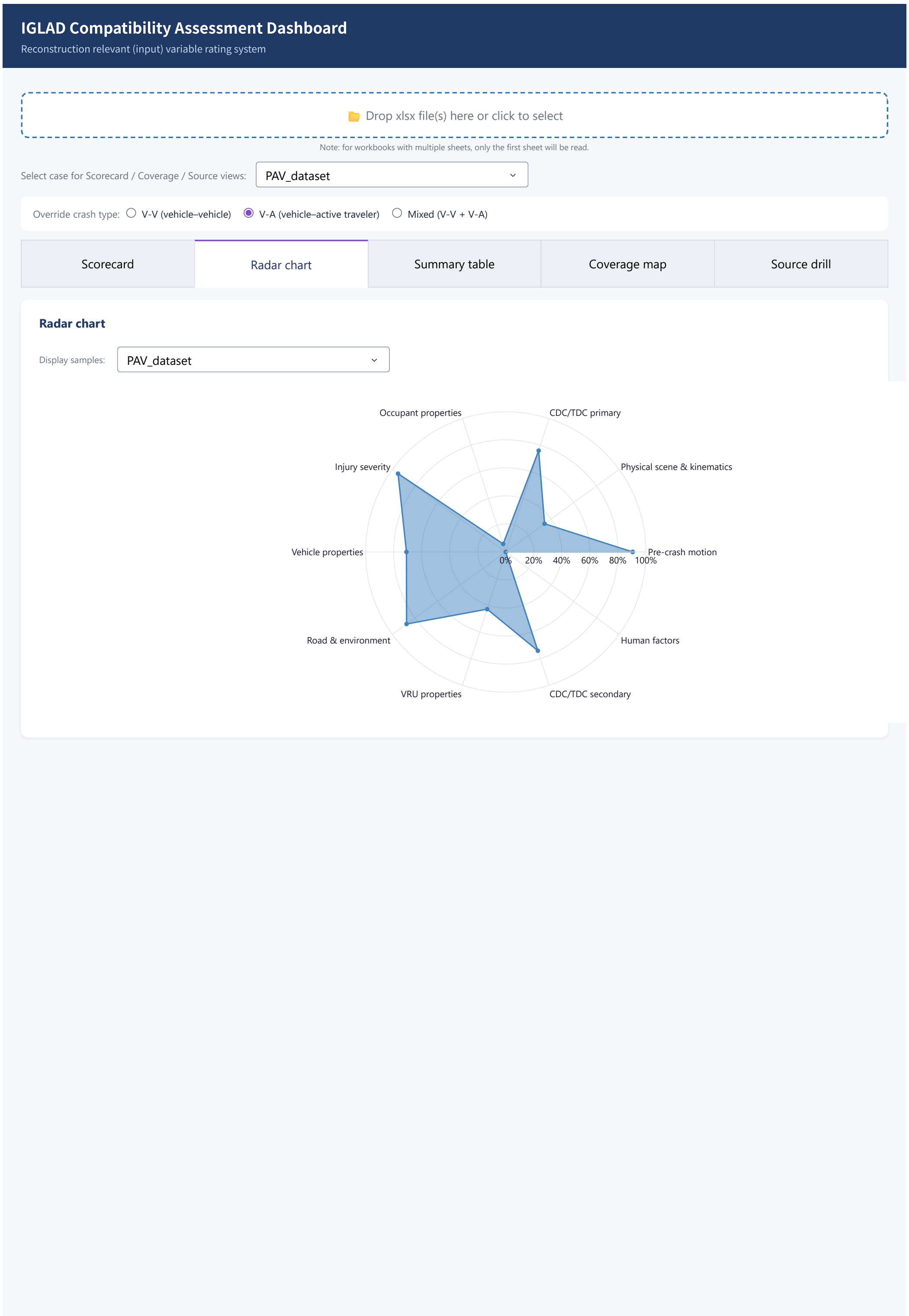


Figure B.3: Radar chart interface in IGLAD Compatibility Assessment Dashboard (result of PAV dataset presented).

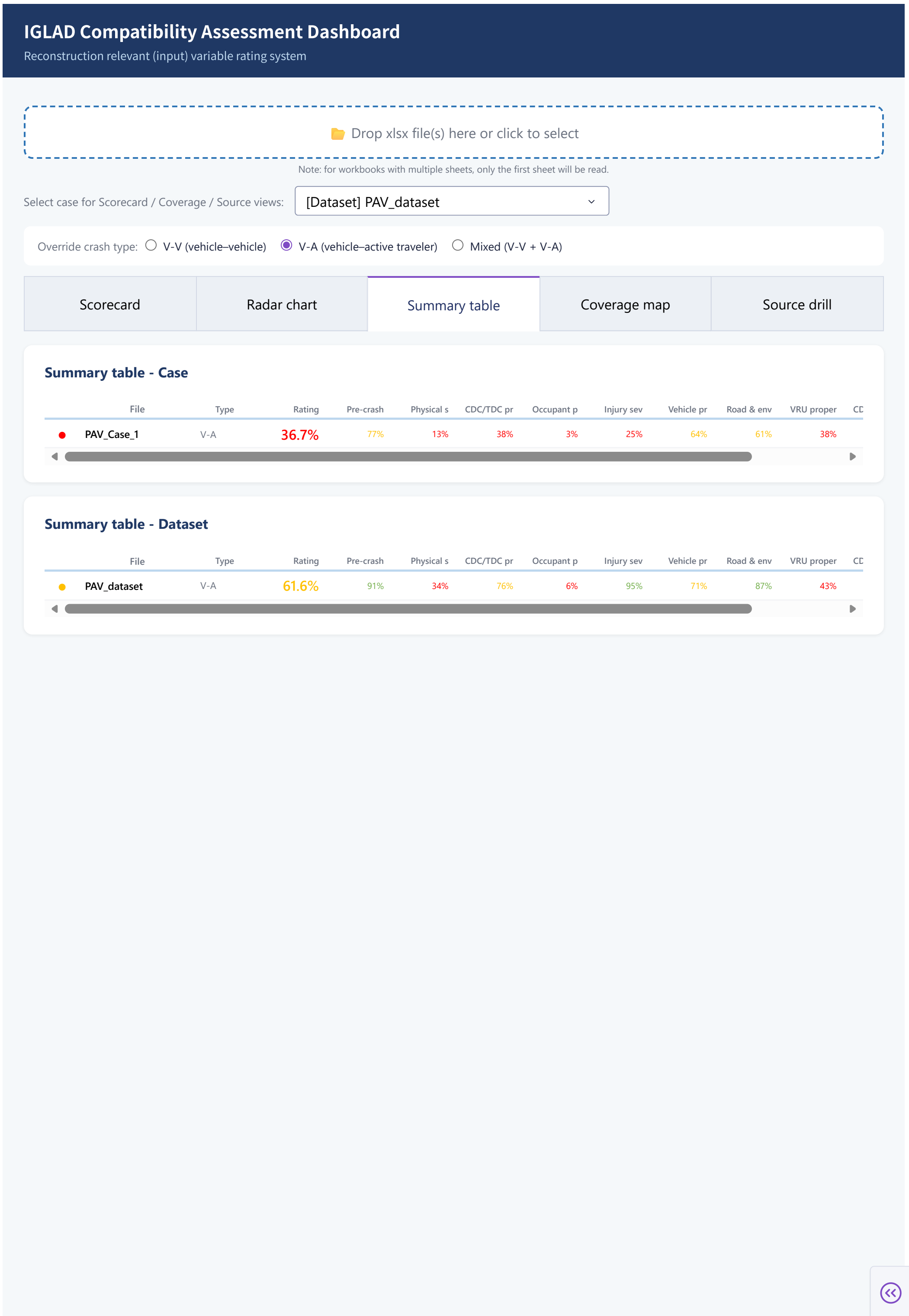


Figure B.4: Summary table interface in IGLAD Compatibility Assessment Dashboard (results of PAV case & dataset presented).

Figure B.5: Coverage map interface in IGLAD Compatibility Assessment Dashboard (result of PAV dataset presented).

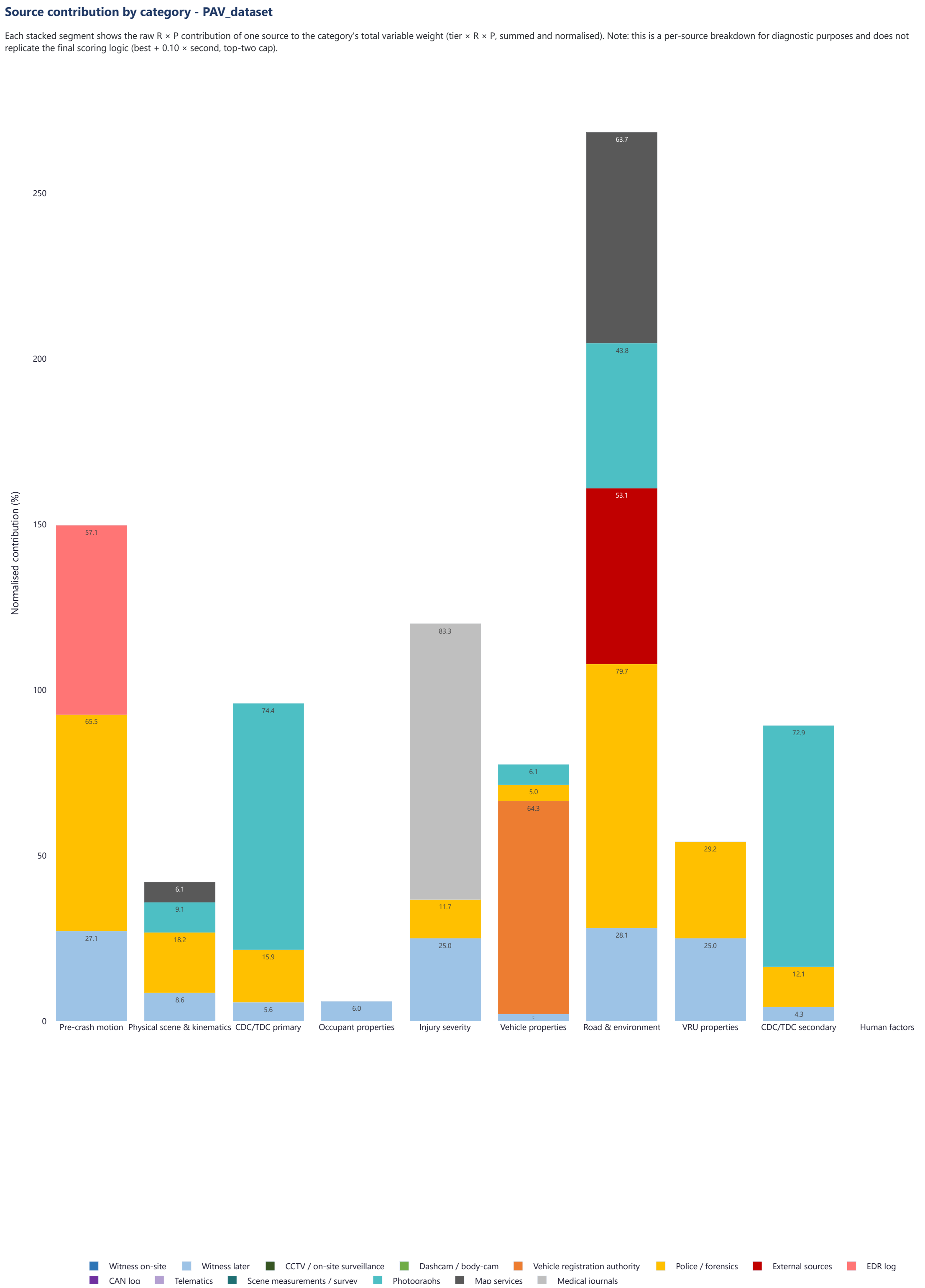


Figure B.6: Source drill interface in IGLAD Compatibility Assessment Dashboard (result of PAV dataset presented).

C

Appendix 3

C.1 Survey documents

Survey on Accident Dataset Characteristics and Reconstruction Practices

This survey aims to help ensure a more accurate assessment of dataset quality and contribute to obtaining more meaningful results.
Please fill in the following questions.

1. Dataset name:
____

2. Data provider / organization:
____

3. Country:
____

4. Does the dataset include vehicle-to-VRU accidents?
☐ Yes ☐ No

5. Was accident reconstruction conducted during harmonization with IGLAD?
☐ Yes ☐ No

6. If yes, what reconstruction methods were used?
☐ Manual estimation
☐ Simulation software (e.g., PC-Crash, MADYMO)
☐ Other: ____________________

Figure C.1: Questionnaire sent to existing IGLAD members.

*Note	Variable name	Applicable to (participant type)	Statements		On-site footage		Official records			Vehicle on-board systems			Scene documentation & physical evidence			Medical records
			Witness, incl participant (on-site, right after crash)	Witness, incl participant (later – questionnaires / claims filing)	On-site surveillance system / CCTV	Dashcam / body-cam footage	Vehicle registration authority	Police / forensics	External sources (e.g. weather services or transport administration road database)	On-board sensing / EDR log	CAN log	Telematics, ie - data collection using a telematics unit or retrieving data directly from vehicle on-board sensing	Scene measurements / survey	Photographs	Map services (retrospective)	Medical journals
V = motor vehicle participants (incl. vehicle, vehicle driver, vehicle occupant) | A = active travaler participants | Both = relevant to both participant types																
PRE-CRASH MOTION																
	Vehicle initial speed (if any reference or estimation)	V														
	Vehicle crash speed (if any reference or estimation)	V														
	Vehicle brake / throttle pedal & steering input state	V														
	VRU crash speed (if any reference or estimation)	A														
	VRU posture at impact (upright / running / bent)	A														
	Vehicle evasive action taken (steer, brake, none)	V														
	Pre-crash scenario type	Both														
ROAD & ENVIRONMENT																
	Road type (motorway, urban, rural...)	Both														
	Road surface type (asphalt, gravel...)	Both														
	Road condition (dry, wet, snow, ice)	Both														
	Light condition (daylight, dusk, night, street lamp)	Both														
	Weather (rain, fog, snow, clear)	Both														
	Speed limit at location	Both														
	Lane separation type (barrier, markings, none)	Both														
	Location type (intersection, mid-block, roundabout…)	Both														
	GPS coordinates of crash location	Both														
PHYSICAL SCENE & KINEMATICS																
	Road surface friction coefficient μ	Both														
	Longitudinal gradient / slope	Both														
	Sight-line / available visibility distance to point of impact	Both														
	Skid / braking mark length and position on road	Both														
	VRU pre-crash marks on road	A														
	Point of impact (POI) location on road	Both														
	VRU final rest position on road	A														
	VRU first ground contact point	A														
	VRU contact location on vehicle	A														
	VRU transportation tool final rest position (separate from rider)	A														
	Vehicle post-impact travel distance and rest position	V														
	Vehicle post-impact heading (direction change)	V														
VEHICLE PROPERTIES																
	Vehicle make / model / year	V														
	Vehicle curb weight	V														
	Gross vehicle weight	V														
	Vehicle length, width, height	V														
	ABS / ESC system activation	V														
	Active safety system activation	V														
	Tire type & condition	V														
CDC / TDC — VEHICLE DAMAGE (primary collision)																
	Principal direction of force (PDOF) – clock position	Both														
	Area of deformation on vehicle body	Both														
	Specific longitudinal / lateral location of damage	V														
	Specific vertical location of damage (belt-line, undercarriage…)	V														
	Type of damage distribution (wide / narrow / sideswipe / rollover)	V														
	Maximum extent of penetration – zone number	V														
	Maximum extent of penetration (% of vehicle dimension)	V														
CDC /TDC — VEHICLE DAMAGE (secondary collision)																
	PDOF secondary collision	Both														
	Area of deformation secondary collision	Both														
	Longitudinal / lateral location secondary	V														
	Vertical location secondary	V														
	Type of damage distribution secondary	V														
	Maximum extent of penetration secondary – zone	V														
	Maximum extent of penetration secondary (%)	V														
VRU PROPERTIES																
	VRU body height	A														
	VRU body mass	A														
	VRU type (pedestrian / cyclist / wheelchair / stroller / e-scooter / other)	A														
	VRU age	A														
	VRU gender	A														
	Helmet use (for avtive traveler)	A														
OCCUPANT PROPERTIES (motor vehicle)																
	Occupant type (driver, front passenger, rear passenger)	V														
	Occupant age	V														
	Occupant gender	V														
	Occupant body mass (kg)	V														
	Occupant height (m)	V														
	Seat belt fitted (yes/no/type)	V														
	Seat belt in use at time of crash	V														
	Front airbag fitted	V														
	Front airbag deployed	V														
	Side / curtain airbag fitted	V														
	Side / curtain airbag deployed	V														
	Knee airbag fitted	V														
	Knee airbag deployed	V														
	Child seat present	V														
	Helmet use (for motorcycle occupant)	V														
INJURY SEVERITY (AIS by body region)																
	Maximum AIS score (MAIS)	Both														
	AIS regional coding	Both														
	Police injury severity classification	Both														
HUMAN FACTORS / CONTRIBUTING FACTORS																
	Driver perception–reaction time (PRT)	V														
	Time-to-collision (TTC)	Both														

Figure C.2: Reconstruction relevant variable grid.

DEPARTMENT OF MECHANICS AND MARITIME SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se