UMTRI-2014-32 APRIL 2014

MUTUAL RECOGNITION METHODOLOGY DEVELOPMENT

CAROL A. C. FLANNAGAN
PAUL E. GREEN
KATHLEEN D. KLINICH
MIRIAM A. MANARY

UMTRI
UNIVERSITY OF MICHIGAN
TRANSPORTATION RESEARCH INSTITUTE

ANDRÁS BÁLINT
ULRICH SANDER
BO SUI
PETER SANDQVIST
SELPI
CHRISTIAN HOWARD

SAFER
VEHICLE AND TRAFFIC SAFETY CENTRE AT CHALMERS

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UMTRI-2014-32		

4. Title and Subtitle Mutual Recognition Methodology Development	5. Report Date April 2014
	6. Performing Organization Code
7. Author(s) Carol A.C. Flannagan, Paul E. Green, Kathleen D. Klinich, Miriam A. Manary, András Bálint, Ulrich Sander, Bo Sui, Peter Sandqvist, Selpi, Christian Howard	8. Performing Organization Report No.
9. Performing Organization Name and Address University of Michigan Transportation Research Institute 2901 Baxter Rd. Ann Arbor MI 48109 SAFER, Gothenburg, Sweden	10. Work Unit No. (TRAIS)
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address Alliance of Automotive Manufacturers	13. Type of Report and Period Covered Phase 1 Final Report
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Phase 1 of the Mutual Recognition Methodology Development (MRMD) project developed an approach to statistical modeling and analysis of field data to address the state of evidence relevant to mutual recognition of automotive safety regulations. Specifically, the report describes a methodology that can be used to measure evidence for the hypothesis that vehicles meeting EU safety standards would perform similarly to US-regulated vehicles in the US driving environment, and that vehicles meeting US safety standards would perform similarly to EU-regulated vehicles in the EU driving environment. As part of the project, we assessed the availability and contents of crash datasets from the US and the EU, as well as their collective ability to support the proposed statistical methodology.

The report describes a set of three statistical approaches to "triangulate" evidence regarding similarity or differences in crash and injury risk associated with EU- and US-regulated vehicles. Approach 1, *Seemingly Unrelated Regression*, tests whether the models are identical and will also assess the capability of the data analysis to detect differences in the models, if differences exist. Approach 2, *Consequences of Best Models*, uses logistic regression to develop two separate models, one for EU risk and one for US risk, as a function of a set of predictors, (i.e., crash, vehicle, and occupant conditions). The two models will then be exercised on a standard population for the EU and a standard population for the US. Approach 3, *Evidence for Consequences*, turns the question around to measures the overall evidence for each of a set of possible conclusions. Each conclusion is characterized by a range of relative risk on a single population. Evidence is measured using a weighted average of likelihoods for a large group of models that produce the same outcome. That evidence is then compared using Bayes Factors.

17. Key Word	18. Distribution Statement
Injury risk, safety regulations, statistical methods	

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

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This work was carried out by University of Michigan Transportation Research Institute (UMTRI) in association with SAFER – Vehicle and Traffic Safety Centre at Chalmers, Sweden. UMTRI and three SAFER partner organizations were involved in this project: Chalmers University of Technology (András Bálint and Selpi), Autoliv Research (Ulrich Sander and Bo Sui) and VTI, the Swedish National Road and Transport Research Institute (Peter Sandqvist and Christian Howard). This project was funded by the Alliance for Automotive Manufacturers. The opinions expressed herein are those of the authors and do not necessarily represent those of the funding agency.

EXECUTIVE SUMMARY

The United States and the European Union have begun negotiation of the Transatlantic Trade and Investment Partnership (TTIP). This agreement is designed to reduce barriers to trade between the two economic units. One significant barrier to trade is the differing safety standards testing and requirements for vehicles sold in the EU and the US. Testing the same vehicle make/model under both regimens could be expensive, and negotiation of common standards may be difficult and time-consuming.

An alternative to item-by-item harmonization is mutual recognition. Under this solution, vehicles that meet EU regulations would be recognized for sale in the US, and vehicles that meet US regulations would be recognized for sale in the EU. To justify mutual recognition, it would be helpful (and possibly even necessary) to demonstrate that safety in EU- and US-regulated vehicles is essentially equivalent.

This report describes a proposed methodology that could be used to measure evidence for the hypothesis that vehicles meeting EU safety standards would perform similarly to US-regulated vehicles in the US driving environment, and that vehicles meeting US safety standards would perform similarly to EU-regulated vehicles in the EU driving environment. If an acceptable level of similarity could be demonstrated, then a test-once approach may be supported; that is, passing regulations from one entity, US or EU, would certify a vehicle to be sold in either region.

The current report presents results of Phase 1 of the Mutual Recognition Methodology Development (MRMD) project. Phase 1 was focused on developing an approach to statistical modeling and analysis of field data to address the state of evidence relevant to mutual recognition. As part of the methodology development, we assessed the availability and contents of a variety of crash datasets from the US and the EU, as well as their collective ability to support the proposed statistical methodology.

As a starting point, we compared regulations at a high level and propose a scope of analysis as follows. First, we recommend separating analysis of crashworthiness and crash avoidance. Regulations addressing these are different, as are datasets and appropriate analysis methods. Second, we propose to limit the scope of crashworthiness analysis to US vehicles weighing less than 3.85 tonne (8500 lb) and EU vehicles classified as M1; vehicle model years ranging from 2003-2012; and crash years from 2003-2012. For occupants, we will include drivers and outboard front-row passengers aged 13 years or more. All restraint usage modes (e.g. belted, unbelted), except for child restraint systems, will be included. Planar crashes in all directions (frontal, lateral and rear), rollovers, and pedestrian crashes will be considered in separate analyses, although the results will be merged at the end. Note that there may be insufficient data on pedestrian injury outcome in US datasets to support that element of analysis. For crash avoidance, we will focus on four main areas of regulation: headlamps (affecting pedestrian and nighttime crashes), mirrors (affecting lane-change and merging crashes), brakes (if supported by crash data), and Electronic Stability Control (ESC; affecting rollover and run-off road crashes).

To support any analysis, datasets must be available, comparable, representative of their regions, and sufficiently large. Datasets in the US and EU were identified and catalogued. In-depth databases are those where crash investigation and reconstruction has been done, including recording of medically evaluated injury outcome. These are required for crashworthiness analyses. National databases generally contain only the contents of police reports and can contribute to crash avoidance analysis. The USDOT provides nationally representative in-depth and police-reported databases. Access is not restricted and sample sizes are relatively large. In the EU, in-depth databases are collected in only a few countries, notably Germany, Sweden, France, and Great Britain. The combined sample size is judged sufficient for analysis, but we have some concerns about representation. In general, these countries are among the wealthier EU countries and vehicle purchase choices may reflect that, leading the results to have limited generalization to southern and eastern European countries.

Data harmonization is required among all databases to be used for the analysis. This means that key variables are present in all datasets to be used and that their values are or can be made comparable (i.e., made to mean the same thing). The variable of greatest concern is Delta-V, which represents crash severity. This variable is reconstructed using different methods in different datasets, and there is some evidence that those methods may produce biased results relative to each other. However, we have identified a set of at least 67 cases that have sufficient detail to implement all reconstruction methods and will allow side-by-side comparison of the resulting Delta-Vs. This will allow adjustments to account for reconstruction method, if necessary.

We propose a set of three statistical approaches to “triangulate” evidence regarding similarity or differences in crash and injury risk associated with EU- and US-regulated vehicles. The main difficulty is that we must consider how to compare risk for each vehicle group (defined by the regulatory environment; i.e., “EU-regulated vehicles” vs. “US-regulated vehicles”) when the available data come from very different driving and crashing environments. We will describe a proposed approach below in terms of injury risk given a crash; a similar approach is planned to address crash-involvement risk.

To address the essential problem described in the previous paragraph, we propose to use observed data to produce statistical models that represent injury risk given a crash for US-regulated vehicles and for EU-regulated vehicles separately. Predicted risk in these models is dependent on the various conditions of the crash, vehicle and occupant involved. Each of these two models can then be used to model the resulting injury outcomes that would have occurred in any given population of crashes defined in terms of crash, vehicle and occupant characteristics. For example, the injury outcomes for an EU-regulated vehicle can be modeled in a population that represents US conditions in terms of crash, vehicle and occupant characteristics.

To explore evidence related to the hypothesis of equivalent field performance, we need seven components:

a. A statistical model of injury risk to an occupant of an EU-regulated vehicle, *given* the conditions of any crash/occupant/vehicle combination;
b. A statistical model of injury risk to an occupant of an US-regulated vehicle, *given* the conditions of any crash/occupant/vehicle combination;
c. A standard population of crashes in the EU, described by crash/vehicle/occupant characteristics; this population is thought to represent a likely near-future crash population for the EU;

d. A standard population of crashes in the US, described by crash/vehicle/occupant characteristics; this population is thought to represent a likely near-future crash population for the US;
e. One or more models of how US-regulated vehicles might enter the EU market;
f. One or more models of how EU-regulated vehicles might enter the US market;
g. A means of measuring the evidence for how injury risk in EU- and US-regulated vehicles is likely to differ (or not differ) in a particular crash population.

The seven components described above represent the pieces of two parallel analyses—one that predicts the consequences of allowing EU-regulated vehicles into the US and one that predicts the consequences of allowing US-regulated vehicles into the EU.

Items (a) and (b) are constructed separately from field data within the EU (for (a)) and the US (for (b)). Items (c) and (d) are test beds on which to compare the two risk models side-by-side. They represent the relative distribution of different crash, vehicle, or occupant characteristics involved in crashes in each region. If the two risk models are identical, then the crash distribution is irrelevant. However, if they are different in any way, then the overall relative outcome can be affected by the population of crashes. Thus, (a) and (b) must both be assessed side-by-side in both the populations represented by (c) and (d).

The three statistical approaches use (a)-(d) somewhat differently, in an effort to look at the evidence from different angles. Approach 1, *Seemingly Unrelated Regression*, tests whether the models are identical. The approach will also assess the capability of the data and analysis to detect differences in the models if differences exist. The end result will be a functional description of the tradeoff between two types of errors: Type 1 (decide "different" when "same" is true), and Type 2 (decide "same" when "different" is true). This will allow selection of a decision cutoff that chooses between the two conclusions ("same" and "different") while maintaining a maximum tolerated probability of error of a given type. Finally, there will be a single result to compare to the decision cutoff. The combination of these elements both describes the state of the evidence (i.e., the ability of the data to determine an answer) and the resulting answer (given a particular choice of cutoff).

Approach 2, *Consequences of Best Models*, uses logistic regression to develop two separate models, one for EU risk and one for US risk, as a function of a set of predictors, (i.e., crash, vehicle, and occupant conditions). These correspond to (a) and (b). The two models will then be exercised on a standard population for the EU (i.e., (c)) and a standard population for the US (i.e., (d)). In both cases, the model will include a measurement of uncertainty, or error, and this will be accounted for in the evaluation. The approach will produce a distribution of relative risk (injury risk in EU-certified vehicles relative to injury risk in US-certified vehicles) across the EU population; the distribution values represent the probability that each possible relative risk is the true relative risk. The process will be repeated for the US standard population and that distribution may differ from the first one. The results will identify both the best guesses as to the true relative risks and measure the conclusiveness of the evidence for each.

Approach 3, *Evidence for Consequences*, turns the question around and seeks to measure the overall evidence for each of a set of possible conclusions. Each conclusion is characterized by a range of relative risk on a single population (e.g., injury risk in EU-regulated vehicles/injury risk in US-regulated vehicles in the EU crash environment is between 0.95 and 1.05). Evidence is measured using a weighted average of likelihoods for a large group of models that produce the same outcome. That evidence is then compared using Bayes Factors, the magnitude of which indicates the relative strength of evidence for two groups of hypotheses. Broadly, this approach, like the others, will both provide a best-guess answer and an assessment of the narrowness, or conclusiveness, of the evidence.

One of the major challenges for this methodology is to identify the size of difference that matters. In other words, how close is “close enough”? Items (e) and (f) are required to address this issue, but these are not generated from data. Instead, they are agreed upon. The methodology can be implemented with any fleet penetration model, but different models will result in different injury consequences. A sensitivity analysis is recommended to assist decision-makers in selecting models.

The final section of the Phase 1 report presents our judgment on the feasibility of implementation in Phase 2. In general, we believe that sufficient data are available to conduct the analysis, and datasets are sufficiently harmonized (or amenable to harmonization, in the case of Delta-V) to support analysis. The unknown factor is the level of uncertainty in the models, which in turn contributes to uncertainty in the comparisons. There is a reasonable chance that results will be inconclusive due to model uncertainty. However, all three approaches produce “best guesses” in the form of point estimates. In addition, Approach 2 provides a fairly rich description of patterns of relative risk across different types of crashes, vehicles, and occupants. At a minimum, this could provide some data-driven support for areas in which an item-by-item harmonization effort is least likely to have major consequences for risk.

BACKGROUND

Overview

Negotiations are underway between the United States (US) and the European Union (EU) to establish a Transatlantic Trade and Investment Partnership (TTIP). A primary objective of TTIP is to reduce or eliminate trade barriers between the two economic entities. The differences in motor vehicle regulations between the US and the EU are a major barrier to trade in the automotive sector. Currently, vehicle manufacturers must test to both separate sets of regulations to sell the same vehicle make/model in both places. Testing twice is more expensive and a comprehensive harmonization of the two sets standards would be an arduous and lengthy process.

One way to eliminate this barrier is global or regulation-specific mutual recognition (see next section). Under a global mutual recognition solution, vehicles that meet EU regulations would be permitted for sale in the US, and vehicles that meet US regulations would be permitted for sale in the EU. To justify such mutual recognition, it would be helpful (possibly even necessary) to demonstrate that safety in EU- and US-regulated vehicles is essentially equivalent, using field data.

The methodology presented in this report could be used to measure evidence for the hypothesis that vehicles meeting EU standards would perform similarly to US-regulated vehicles in the US driving environment and that vehicles meeting US standards would perform similarly to EU-regulated vehicles in the EU driving environment. If an acceptable level of similarity could be demonstrated, then a test-once approach may be supported; that is, passing regulations from one entity, US or EU, would certify a vehicle to be sold in either region.

The US-EU Bilateral Air Safety Agreement (BASA)

The US and the EU have prepared agreements in several areas before the TTIP negotiations concerning the harmonization and/or mutual acceptance of each other's regulations. A specific group is called mutual recognition agreements, which mean that officials on each side agree to accept products or services from the other side based on a "test once" criterion. It was briefly investigated whether information about such agreements related to traffic safety could be relevant in the present research project. The investigation was focused on the largest agreement of this kind up to date related to traffic safety, namely the US-EU Bilateral Air Safety Agreement on the regulation of civilian aviation safety[1] (BASA) from 2011, together with a description of the technical implementation[2] of the BASA agreement (abbreviated here as TIPAEC).

Although BASA is focused on avoiding doubled full-scale certification efforts, it is *not* a mutual recognition agreement in the global sense described above. Instead, it enables and encourages mutual acceptance of *judgments and tests* performed by the appropriate authorities from the other side to the maximum extent possible. This means that the findings made by one party may only need to be

[1] Agreement between the United States of America and the European Community on cooperation in the regulation of civil aviation safety, 2011-0088. Available at http://www.state.gov/documents/organization/169475.pdf.

[2] *Technical Implementation Procedures for Airworthiness and Environmental Certification*, Revision 3, April 23, 2013

reinforced by a validation process by the other party which, according to the vision described in Section 1.4 of part C-3 in TIPAEC, is a “simple process based on mutual authority trust, which leads to design acceptance in compliance with the Validating Authority’s airworthiness standards.” At the same time, each party retains its regulatory power and the authorities retain the responsibility to “make findings of compliance” with regulations in their respective regions.

Furthermore, a set of regulations have been identified that are considered equivalent under the BASA agreement. The principles for the identification of equivalent regulations are described in Section 3 of part C-6 in TIPAEC. Importantly, the agreement allows for judgment of equivalence even in those cases when the regulations are formally different. It is specified in Section 3.1.1 of part C-6 in TIPAEC that: “A literal comparison of the airworthiness standards developed by the US Federal Aviation Administration (FAA) and European Aviation Safety Agency (EASA) indicates that there are instances where the standards text differs extensively. In some cases, the FAA and EASA airworthiness standards may be determined to be equivalent despite such text differences.” A list of standards that are deemed equivalent can be found in Appendix B of TIPAEC.

In Article 5, Annex A of BASA, it is stated that “the Parties agree that each Party's civil aviation standards, rules, practices and procedures are sufficiently compatible to permit reciprocal acceptance of approvals and findings of compliance with agreed upon standards made by one Party on behalf of the other as specified in the Annexes.” However, there is no reference or description in the document of how the conclusion on “sufficient compatibility” has been achieved. Therefore, it appears that the BASA agreement was not attained through specific research aimed at evaluating functional equivalence of aircrafts certified in the US and the EU with respect to real-world risk, but rather followed a different path based on mutual trust and a direct comparison of aviation safety regulations.

Although the stated goal of this project is to address evidence regarding global mutual recognition, elements of the analysis could address specific regulations or groups of regulations as treated in the BASA agreement. Within BASA, the “mutual trust” approach does not seem to require data or analysis; but evidence for “sufficient compatibility” of subgroups of regulations might be obtained through analyses described in the methodology below, whether or not global mutual recognition is supported.

PROBLEM STATEMENT

The objectives of this project are to:

1) Develop a methodology to test the hypothesis that vehicles regulated in the US and EU would have essentially equivalent real-world safety performance when driven in a common driving environment. In other words, vehicles meeting EU standards would perform similarly to US-regulated vehicles in the US driving environment and vehicles meeting US standards would perform similarly to EU-regulated vehicles in the EU driving environment.
2) Identify appropriate datasets, evaluate their preliminary usage in the analysis, and identify potential issues.

The idealized approach to this question would be to have a set of similar EU-regulated and US-regulated vehicles that are driven in both the EU and the US by similar drivers. Crashes and injury outcome in such a dataset would provide a side-by-side comparison of injury risk for the two vehicle groups. However, since the point of the analysis is to address regulatory restrictions to vehicles that can be driven in each region, this is essentially a logical impossibility.

Instead, we must consider how to compare risk for each vehicle group (defined by the regulatory environment; i.e., “EU-regulated vehicles” vs. “US-regulated vehicles”) when the available data come from very different driving and crashing environments. A further difficulty is that the most relevant driving/crashing environment is one that is in the future (i.e., when mutual recognition would hypothetically allow both vehicle groups to be driven in the US and the EU), and therefore not observed yet. We will describe a proposed approach below in terms of injury risk given a crash; a similar approach is planned to address crash-involvement risk.

To address the essential problem described in the previous paragraph, we use observed data to produce statistical models that represent injury risk given a crash for US-regulated vehicles and for EU-regulated vehicles separately. Predicted risk in these models is dependent on the various conditions of the crash, vehicle and occupant involved, but is not dependent on the *distribution* of such conditions in the crash population. These models (one for US-regulated vehicles and one for EU-regulated vehicles) can then be used to estimate the resulting injury outcomes that would have occurred in any given population of crashes if the vehicle involved had been regulated in the US or if it had been regulated in the EU. In particular, it can be estimated how the model that corresponds to EU-regulated vehicles would perform in a population representing US conditions and vice versa.

To conduct such an analysis, databases must be found to support it. The sections that follow start with an assessment of the scope of the problem that we propose to address (which subset of regulation, vehicles, crashes, and occupants are considered). This is followed by a description of available databases, and then a discussion of data-related issues. The following section covers statistical models and methods in detail, and finally, the last section discusses whether Phase 2 could go forward.

SCOPE

Regulation comparison

This section provides an overview comparing EU and US vehicle safety standards. The purpose of this review comparing regulations is to define a scope of data analysis that applies to relevant and comparable crash types.

Vehicle Types

In the EU, the first method of vehicle type classification is by the number of wheels and the vehicle’s purpose. Category L consists of mopeds, motorcycles, motor tricycles, and quadricycles. Category M includes power-driven vehicles having at least four wheels and used to carry passengers. Category N includes power-driven vehicles having at least four wheels and used to carry goods. Category O includes trailers and semi-trailers, T includes agricultural and forestry tractors, and category G includes off-road vehicles. Other categories include special purpose vehicles and non-road mobile machinery. The second way of classifying vehicles in the EU depends on either the engine capacity, mass, number of passengers, and capacity for standing passengers. Figure 1 shows the EU vehicle category M1 relevant to this analysis; vehicles in this category are intended to carry 8 occupants in addition to the driver or less and weigh less than 3.5 t.

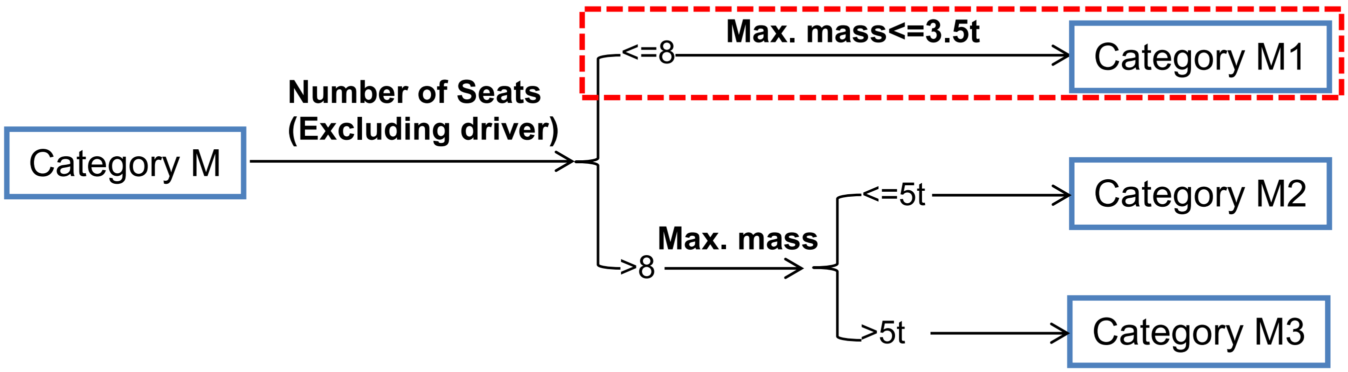


Figure 1. Relevant vehicle classes as defined by EU regulations.

The US Department of Transportation also has many ways of categorizing vehicle types, defined one way in FMVSS 571.3. Motorcycles are defined as powered motor vehicles having less than or equal to three wheels equipped with a seat or saddle for use by the rider. Those with less than 5-brake horsepower are considered to be motor-driven cycles. Passenger vehicles are powered vehicles carrying less than or equal to 10 persons except for other categories described here. A multipurpose passenger vehicle is also a powered vehicle carrying less than or equal to 10 persons, but is either constructed on a truck chassis or with special features for off-road operation. A subcategory is motor homes, which include provisions for temporary residential accommodations.

Buses are powered motor vehicles carrying more than ten persons. Subcategories are school buses or multifunctional school activity buses. Trailers are motor vehicles, with or without power, designed for carrying persons or cargo and for being drawn by another motor vehicle. Subcategories include boat, semi, pole, and full trailers. A truck is a powered motor vehicle, except a trailer, designed for transporting property or equipment. A truck tractor is a truck designed to draw other motor vehicles. Firefighting vehicles are in a separate class. Low-speed vehicles are 4-wheeled powered motor vehicles with a with a gross vehicle weight rating (GVWR) of $\leq$ 1,361 kg (3,000 lb) whose maximum attainable speed in 1.6 km (1 mile) ranges from >32 km/hr (20 miles/h) to $\leq$ 40 km/hr (25 miles/hr) on a paved level surface.

FMVSS 523.2 uses a different method of vehicle categorization. Passenger automobiles are those manufactured primarily for use in the transportation of <=10 individuals. Non-passenger automobile are those that are not a passenger automobile or a work truck. Medium duty passenger vehicles are designed to mainly transport passengers, and either have a GVWR >3,856 kg (8,500 lb) or a vehicle curb weight (VCW) > 2,722 kg (6,000 lb) or has a basic vehicle frontal area in excess of 4.18 m^2 (45 square feet). Heavy-duty vehicles are defined to include any commercial medium- and heavy-duty on highway vehicle or a work truck. Weight ranges are as follows: 1) Heavy-duty pickup trucks and vans: 3,856 kg <= GVWR <=6,350 kg (8501<=GVWR <= 14,000 lb), 2) Heavy-duty vocational vehicles: GVWR> 3,856 kg (8,500 lb), 3) Truck tractors with a GVWR > 11,793 kg (26,000 lb).

Table 1 compares the vehicle classifications used in the US and EU. The categories in gray shaded cells are considered to be most relevant to this project.

Table 1.
Comparison of US(DOT) and EU vehicle classifications.

		US		EU	
Mass	**#**	**Category**	**Category**	**#**	**Mass**
	≤ 3 wh	Motorcycles	Motorcycles & LSV	≤ 4 wh	< 400 kg
< 1.3 t	4 wh	LSV			
< 4.5 t	≤ 10 p	Passenger vehicles	M1	≤ 9 p	
>3.9 t	≤ 10 p	Multi purpose passenger vehicle (truck chassis)	M2	>9 p	< 5 t
	>10	Bus	M3	> 9 p	> 5 t
			N1		< 3.5 t
> 3.9 t		Heavy duty vehicle	N2		3.5 – 12 t
> 11.8 t		Heavy trucks	N3		> 12 t
			O		.75 t, .35 t, 10 t

Crash Avoidance

The regulations in the US and EU that pertain mostly to "traditional" crash avoidance vehicle features, such as brakes and lighting, are summarized in Table 2. There are somewhat comparable regulations for lighting, brakes, tires, and mirrors. The pedestrian regulation in the US only applies to school buses. In the US, there are also several regulations for items not necessarily tied to crash avoidance that address windshields and wipers, transmissions, vehicle theft, rollaway, hood latch, and window latch.

Table 2.
Summary of EU and US regulation numbers pertaining to crash avoidance vehicle features.

US	TOPIC	EU
108	Lighting	1, 3, 7, 8, 19, 20, 31, 38, 87, 91, 98, 99, 104, 112, 123, 128
101, 123, 124, 125	Controls, steering, warning	28, 68, 79
109, 110, 117, 119, 120, 129, 138, 139	Tires, wheels	30, 54, 64, 117,
103, 104	Windshields/wipers	
105, 105, 116, 121, 122, 135	Brakes	13, 90
102	Transmission	
111	Mirrors/vision	46, 125
114	Theft/rollaway	
113, 118	Hood latch/ windows	
126	ESC	13H
131 (school bus only)	Pedestrian	

The US includes requirements for electronic stability control (ESC) in FMVSS 126, which was phased into vehicle model years 2008-2012 and is now required on all passenger vehicles. For the EU, ESC is regulated in the United Nations Economic Commission for Europe (UNECE)-R13H. There are no other specific regulations that set the minimum safety performance of advanced driver assistance and crash avoidance systems before they are allowed to be used on vehicles to be driven on public roads in the EU or in the US, other than the general safety regulations as part of the Whole Vehicle Type Approval in the EU and Federal Motor Vehicle Safety Standards (FMVSS) in the US. However, there are some ISO standards that potentially have been followed by the vehicle manufacturers when developing their driver assistance and crash avoidance systems. Examples of such International Standards Organization (ISO) standards are ISO26262 defined for functional safety for passenger vehicles up to 3.5 tonnes, ISO15623:2013 for forward vehicle collision warning systems and ISO 22179:2009 for full speed range adaptive cruise control (FSRA) systems; they are made for harmonization purposes but are not regulated. A review on existing (up to 2011) ISO, SAE and NHTSA test procedures for driver assistance and crash avoidance systems can be found in (Evgenikos, Papantoniou, Yannis, Stanzel, & Kohsiek, 2011).

Crashworthiness

Table 3 contains a summary of US and EU regulations pertaining to crashworthiness. For ten areas, there are comparable standards in each region, although the exact details of requirements and testing procedures vary. For occupant protection, FMVSS 208 covers items that are addressed in six different EU standards. Two main areas are not addressed by both regions. The US has a standard to address occupant protection in rollovers, while the EU has standards intended to reduce the injury severity of pedestrians struck by vehicles.

Table 3.
Summary of comparable crashworthiness regulations

US	EU
201 Interior impact	21 Interior fittings
202a Head Restraints	25 Head restraints
203/2-4 Steering column	12 Steering column
205a Glazing	43 Glazing
206 Door locks/retention	11 Door latches/hinges
207 Seating systems	17 Seating systems
209 Seat belt assemblies	16 Safety belts
210 Seat belt anchorages	14 Seat belt anchorages
214 Side impact	95 Lateral impact
216 Roof crush	none
208 Occupant protection	33 Frontal impact 100/110 Electric/CNG vehicles 42 Front and rear protection 34 Fire prevention 32 Rear impact 126 Partitioning
None	**26/61 Projections (pedestrian)** **127 Pedestrian (**EG/78/2009, EG/631/2009**)**

Figure 2 provides a graphical representation of how the specifications for full-vehicle impact testing vary between the US and EU. The US requires full-frontal barrier tests with belted Hybrid III midsize and small female anthropomorphic test devices (ATDs) at 56 km/h. There are also full frontal unbelted requirements using the small female and large male at 32-40 km/h, with varying ranges of impact angles. Both the US and EU specify a 40% offset frontal barrier test using the midsize male, but the US test is run at 40 km/h and the EU is at 56 km/h. Both the EU and US specify a side impact barrier test using the ES-2 ATD. The US uses an impact barrier angled at 27 degrees and a test speed of 48 km/h, while the EU uses pure lateral impact at 50 km/h. The US also has a lateral pole impact test requirement.

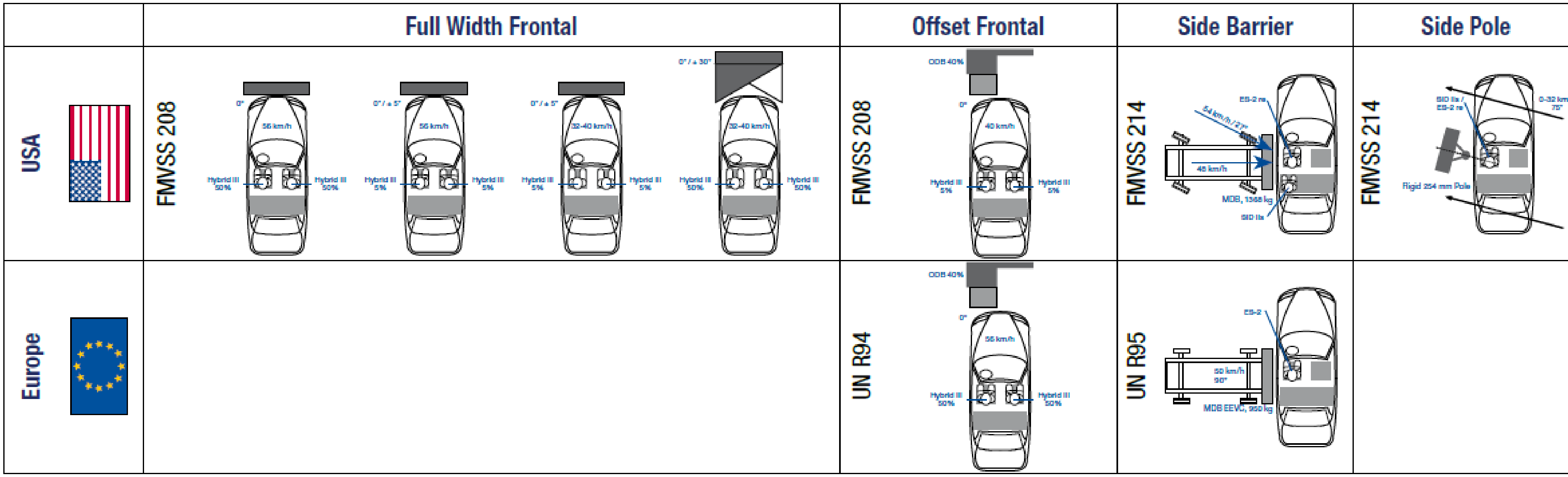


Figure 2. Graphical summary of regulated impact test requirements
(extracted from Crash Safety Wissen app http://www.carhs.de/en/).

Crashworthiness and Crash Avoidance

Analysis of crashworthiness and crash avoidance will be considered separately. In these areas, relevant datasets and outcomes are different, making it difficult to combine results. To analyze crashworthiness, we need to consider datasets containing injury outcome and details of the crash that generally require in-depth crash investigation. For crash avoidance, we need to analyze databases tabulating crash data as well as driving exposure. Because of differing regulations and data elements, rollovers and pedestrian crashes will be analyzed separately from other crash types.

Exposure

The methodology we propose for assessing essentially equivalent real-world performance of vehicle *crashworthiness* is based on the idea that although exposure to the driving environment may be different in the US and the EU (and even within the US and within the EU), vehicle testing is, in its essence, designed to assess risk given a crash and provide standards for maximum acceptable risk. Thus, the methodology should focus on measuring risk, independent of exposure, in the two regions and comparing the assessed risk. That is, when assessing crashworthiness it generally does not matter why the crash arises, whether it is related to driver error, road type, weather conditions or other factors of exposure that may differ between regions. What matters is the injury outcome given the experience of the vehicle (e.g., direction of force and crash severity), the characteristics of the occupant (e.g., age and gender), and the restraint systems used (e.g., seat belt use and airbag deployment). Many, if not all, of these factors are typically present in existing crash and injury datasets.

It is more challenging to use existing crash and injury datasets to assess essentially equivalent real-world performance pertaining to *crash avoidance*. To establish a vehicle's ability to aid in crash avoidance, one must estimate the crash risk by relating crash occurrences to the level of exposure (e.g., the number of miles travelled in a given environment). Crash datasets incorporating the latter are scarce. Thus, a common method of estimating risk is to use an external dataset (e.g., travel-based surveys or odometer readings from roadworthiness tests) for estimating the exposure. Such external datasets should be collected over the same geographical area as the crash dataset. For our purpose of comparing crash risk in the US versus the EU, there is a notable lack of harmonized exposure data. The *International Road Traffic and Accident database* (IRTAD), which is managed by the International Traffic Safety Data and Analysis Group under the Organization for Economic Co-operation and Development (OECD), contains estimates of total passenger vehicle kilometers travelled (VKT) per year as reported separately by each OECD member state. Most of the OECD member states are also members of the EU. Combined with crash datasets from the US and several individual EU states, this data could be used to estimate and compare crash risk on an aggregate level. However, this method may prove uncertain as individual OECD members most likely use different methods of calculating VKT. Furthermore, this approach would not allow for more detailed analysis to be performed because the VKT provided in the IRTAD database cannot be broken down into subcategories (such as vehicle occupant, make and model). Local availability of more detailed exposure data on a national level has also been investigated. Although some travel surveys and odometer datasets have been identified (e.g., in the UK, Germany and Sweden), our assessment is that a significant amount of data processing is needed to extract useful measures of exposure from these datasets.

An alternative method of estimated exposure is called Quasi-Induced Exposure (Cuthbert, 1994; Keall & Newstead, 2010). This method does not require an external exposure dataset. Conceptually, quasi-induced exposure is a method that estimates exposure from characteristics of vehicles that were judged to be non-culpable in crashes. For example, vehicles that were involved in rear-end crashes and were struck from behind are generally inferred to be not at fault in the crash, but information about them is recorded in crash databases at a level that is useful for calculating crash rates. The assumption is that these vehicles happened to be in the vehicle stream randomly and are similar to vehicles in the general population that were not involved in crashes. Under this assumption, counts of induced exposure vehicles serve as measures of exposure used in the calculation of rates.

Data Restrictions

Based on the review of pertinent regulations, this project will include US vehicles weighing less than 3.85 tonne (8500 lb) and EU vehicles classified as M1. Vehicle model years will range from 2003-2012, and crash years will be limited to those from 2003-2012.

For occupants, we will include drivers and outboard front-row passengers aged 13 years or more. All restraint usage modes (e.g. belted, unbelted) except for child restraint systems will be included.

Planar crashes in all directions (frontal, lateral and rear), rollovers, and pedestrian crashes will be considered in separate analyses, although the results will be merged at the end. There may be insufficient data on pedestrian injury outcome in US datasets.

For crash avoidance, we are focusing on four main areas of regulation where we believe that analysis can be carried out. Headlamps likely affect pedestrian and nighttime crashes. Mirrors are associated with lane-change and merging crashes. We hope to consider brakes, if brake failure information is available in the crash data. Stopping distance, also related to brake regulations, is not available. Electronic Stability Control (ESC), which primarily affects rollover and run-off road crashes, will also be evaluated. However, manufacturers generally use the same ESC technology in both regions. Therefore, it is expected to produce similar results.

DATASETS

This section provides details about the availability of databases in the EU and the US that are relevant for the analysis. A list of datasets is provided in Appendix A, including the following metadata about the databases:

- Name (often an abbreviation), full name, country/countries of data collection;
- Owner of the database, homepage, accessibility (public, consortium or private access);
- Number of crashes in the database, data years, inclusion criteria and other relevant information.

These details guide the identification of those datasets that have a potential for being used in Phase 2 of the project[3]. There are several aspects to consider in the selection of the most relevant databases. These include: 1) appropriate vehicle focus (passenger cars); 2) detailed information about crash severity and medical information (at a minimum, Delta-V[4] and Maximum Abbreviated Injury Score (MAIS[5]) values are necessary for the crashworthiness analysis–a list of potential key variables is provided in Appendix B); and 3) a sufficiently large sample size (primarily, databases containing at least 1000 crashes are considered). Since the scope of analysis is limited to vehicles manufactured in 2003 or later, it is a prerequisite for relevance that the database contains data years in this period. Furthermore, it is important that the inclusion criteria for the databases are comparable (or can be made comparable by applying data filters) with those in NASS-CDS and GIDAS[6] since unaccounted differences in inclusion criteria could bias the results.

To facilitate a proper understanding of Appendix A, some background information is provided in the rest of this section regarding crash databases in general, and additional details are given for the most relevant datasets. Crash databases in the US are reviewed first, then crash data in the EU (both multinational and country-specific databases), and then global databases are considered; finally, we summarize which datasets are under consideration for Phase 2 of the project and how data from Field Operational Tests (FOT) will be addressed.

Crash Data in the US

The US is a single country, and national crash datasets are made available for free. There are three major national datasets of crashes: 1) the *Fatality Analysis Reporting System (FARS)*; 2) the *National Automotive Sampling System—Crashworthiness Data System (NASS-CDS or CDS)*; and 3) the *National Automotive Sampling System—General Estimates System (NASS-GES or GES)*. The first, FARS, is a census of fatal crashes on public roads in the US. The second is an annual probability sample of approximately 3500-4500 tow-away crashes involving light vehicles. The CDS data collection includes in-depth crash investigation and estimation of Delta-V using the software WinSMASH, as well as details on injury outcome coded according to the Abbreviated Injury Scale (AIS). Finally, GES is an annual probability sample of approximately 50,000 police-reported crashes. The basis for the data in GES is information contained in state police crash reports, but the data elements are coded to a national standard. Therefore, in the second phase of the MRMD project, CDS will be used when in-depth data (including Delta-V and MAIS values) are required, while GES enables estimates on a national level and can give information about the standard crash-involved occupant population.

[3] For readability, it is assumed throughout this section that there will be a second phase in the MRMD project; i.e., the uncertainty regarding Phase 2 will generally not be expressed when writing about this phase.
[4] Change of velocity in a crash; this is the most commonly used measure for crash severity.
[5] Maximum Abbreviated Injury Scale value; this is a universal measure of overall injury severity.
[6] NASS-CDS: National Automotive Sampling System—Crashworthiness Data System; GIDAS: German In-Depth Accident Study. These are the largest in-depth crash databases in the US and the EU, respectively (detailed descriptions of these databases are given in this section), and therefore need to be included in the analysis.

Multinational in-depth databases in the EU

In-depth crash data in the EU have been collected in specialized crash investigation studies in EU projects focusing on, for example, the causation of road crashes, rollover crashes, passive safety of passenger cars and fatal crashes. The resulting databases each contain from hundreds to a few thousand crashes from 19 countries in the EU. The EU projects in which in-depth data collection was conducted are listed in Appendix A among multinational in-depth databases. A difficulty with using data from EU projects is that the special focus in these projects often results in incomparable inclusion criteria. A further condition restricting the set of relevant databases is that crash reconstruction was not conducted in some of these projects; hence Delta-V is unavailable. One exception is the PENDANT project, briefly described below. In this database, the inclusion criteria are comparable with those of GIDAS, and Delta-V reconstruction was conducted in the project.

PENDANT, the *Pan-European Co-ordinated Accident and Injury Database*, was developed between 2003 and 2005 in a project co-funded by the European Commission. The main objective of PENDANT was to support EU vehicle and road safety policy-making. The resulting database contains approximately 1100 crashes collected in eight EU countries (Austria, Germany, Spain, Finland, France, The Netherlands, Sweden and the United Kingdom). An inclusion criterion is that at least one vehicle occupant was injured in the crash; this is a stricter version of the criterion for GIDAS where all injury crashes are collected (including those in which, for example, a pedestrian is injured but all vehicle occupants are uninjured). A further criterion is that the crash includes a vehicle with model year 1998 or later; this does not mean any further restriction for the analysis in the MRMD project due to the limit of 2003 on vehicle model year in the analysis. Note, however, that in several crashes in PENDANT, all involved vehicles are of model years before 2003 hence the sample size for the analysis in Phase 2 will be substantially smaller than the total number of crashes.

Country-specific in-depth databases in the EU

Besides EU projects, an important source for in-depth crash data in the EU is provided by in-depth data collection projects in individual countries including Sweden, Germany, France, and the United Kingdom. Again, a complicating factor for the analysis is the variation of sampling criteria in these projects and differences in the data variables. Therefore, an analysis using multiple datasets needs to find a way to make up for differences in the sampling and identify comparable data elements. Access is limited for all these databases, which also means that metadata about the datasets can either be retrieved by literature review or via contact with the database owners/users; in the latter case, the availability of the information is strongly dependent on the willingness of the data owners to disclose information. The metadata is described in Appendix A under the heading “Country-specific in-depth databases in the EU” and additional information on the most relevant ones for the MRMD project, namely GIDAS, CCIS, VOIESUR, LAB and INTACT, is provided below.

The *German In-Depth Accident Study (GIDASe)* is the largest database of its kind in Europe. Data collection commenced in 1999 and was initiated by the German Federal Highway Research Institute (BASt) and the German Association for Research on Automotive Technology (FAT) which unites all

German passenger and commercial vehicle manufactures as well as numerous suppliers. Today GIDAS has 15 sponsors including the MRMD project partner Autoliv; the sponsors of GIDAS have exclusive access to the database. Crash data is collected by two teams, one at the Hannover Medical School (MHH) and one at the Traffic Accident Research Institute (VUFO) of Technische Universität Dresden (TU Dresden). After 15 years of continuous data collection the database includes over 22,000 injury crashes (i.e. crashes in which at least one person was injured) investigated in-depth. Delta-V values are reconstructed using a momentum method, primarily utilizing the software PC-Crash.

The *Cooperative Crash Injury Study (CCIS)* is a major crash database in the UK in which data collection, funded by the UK Department for Transport and industrial partners, started in 1983 and ended in 2010. The sponsors have exclusive access to the database, which contains more than 15,000 crashes. Crash events are collected according to a stratified sampling procedure, which favors cars containing fatal or seriously injured occupants (Richards and Cuerden 2009). More specifically, the inclusion criteria in CCIS require that at least one passenger car which is younger than seven years has been involved in the crash and towed from the scene and that at least one crash-involved occupant was injured, according to the police report. Data was collected retrospectively (up to several days after the crash) by teams of investigators from Birmingham Automotive Safety Centre (BASC) based at the University of Birmingham, Vehicle Safety Research Centre (VSRC) based at Loughborough University and Vehicle Operations and Standards Agency (VOSA) from various locations in England. Delta-V reconstruction in CCIS is damage-based, using the software AI-Damage.

VOIESUR (Véhicule Occupant Infrastructure Etudes de la Sécurité des Usagers de la Route - Vehicle Occupant Infrastructure and Road Users Safety Studies) is a project funded by the French National Research Agency and Foundation MAIF. In this project, a database of more than 9000 crashes is built from the in-depth analysis of police reports in France in 2011. More specifically, the database contains the following crashes in 2011: all fatal crashes in France, 5% of the injury crashes in France and finally every crash that year in the Rhône region. The data come from expert investigations of police reports, sketches and photos. However, police-coding of variables is not automatically accepted – instead, police information is used to understand what happened. Delta-V in the crash is reconstructed using a method based on the vehicle trajectories when there is sufficient data available to do so. A consortium of four French research organizations has developed the VOIESUR database: CEESAR[7], CETE NC[8], IFSTTAR[9] and LAB[10], and the agreement of all members is required for data access.

LAB (Laboratory of Accidentology and Biomechanics - Secondary Safety Database) is a French database for which crashes have been collected since 1993. The database is owned by the Laboratory of Accidentology, Biomechanics and human behavior [10] which is a shared laboratory between the two French car manufacturers, PSA (Peugeot-Citroën) and Renault. Around 400 crashes are collected annually in the database from Yvelines (west of Paris) and the whole of France for some targeted

[7] Centre Européen d'Etudes de Sécurité et d'Analyse des Risques
[8] Centre d'ÉtudesTechniques de l'Équipement Normandie-Centre
[9] Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux
[10] Laboratoire d'Accidentologie, de Biomécanique et d'Etudes du comportement humain

vehicles. The inclusion criterion for crashes is the presence of at least one injured occupant in a passenger car. The database contains detailed information on crash characteristics and injury characteristics.

The *INTACT* database was developed in consecutive research projects funded by the vehicle industry, the Swedish Governmental Agency for Innovation Systems (VINNOVA), the Intelligent Vehicle Safety Systems (IVSS) program, the European Commission (EC) and the Swedish Research Council (VR). These projects included both methodology development and data collection addressing different applications. The INTACT methodology, developed in the IVSS-funded project *Investigation Network and Traffic Accident Collection Techniques* during 2007-2010, was adapted by the EU project *Road Safety Data, Collection, Transfer and Analysis* (DaCoTA) in 2010 as the method to be used for in-depth crash investigation on a European level (Hill et al. 2012). Data collection using this methodology is ongoing in a VR-funded project; the resulting database currently contains approximately 250 crashes. Data collection is conducted in Gothenburg, Sweden and the six surrounding municipalities; the inclusion criteria are that at least one passenger car, bus or truck was involved in the crash and an ambulance was called to the crash scene. The software PC-Crash is used for Delta-V reconstruction.

Due to the relatively small number of cases in INTACT, the contribution of the database in estimating the overall injury risk for EU-registered vehicles will be minor. The INTACT data is, however, expected to be important in the MRMD project because it can address a crucial question related to data harmonization. This aspect is further elaborated in a subsequent section.

EU level crash data

The most comprehensive source for national crash data in the EU is the *Community Road Accident Database (CARE),* which contains national data from all 28 EU countries plus Iceland, Norway and Switzerland. CARE has no data collection activity of its own but the data come from the member states; such data are recoded according to uniformization protocols (CAREPLUS and CADaS) to obtain a standardized dataset. Only selected organizations from each participating country (at most three per country) have access to the database; these include the MRMD project partners Chalmers and VTI. However, CARE does not contain Delta-V or MAIS values since those are generally not included in national crash data; therefore, the main use of CARE in this project is the specification of the standard crash-involved occupant population in Europe.

Global datasets

There have been efforts to standardize datasets internationally, including the *International Road Traffic Accident Database (IRTAD),* already mentioned in the “Exposure” section. In this database, aggregated road safety data as well as relevant exposure data from OECD member countries are collected in a standardized format. Of the 29 OECD countries that participate, 19 are EU members. Besides the number of injury crashes, fatalities and hospitalizations, IRTAD has information on seat belt wearing rates by road network areas. Access to the database is restricted to IRTAD member organizations (which include the MRMD project partners UMTRI and VTI).

A recent effort for data standardization is the on-going project *Initiative for the Global Harmonization of Accident Data (iGLAD),* which was initiated in 2010 by European car manufacturers. The MRMD project partner Chalmers has been given the commission of trust to administer the project. The aim of iGLAD is to develop a global in-depth dataset to improve road and vehicle safety. In the first phase of iGLAD, funded by the European Automobile Manufacturers Association (ACEA), ten countries from four continents contribute with 100-200 cases from their own data repository. The data are recoded according to a common data scheme defined within the iGLAD project, and after merging the datasets the database will be delivered to the ACEA members and the data providers mid-2014. This first data delivery will contain 1580 crashes in total.

In the following years, data providers will contribute to the iGLAD database with at least 100 cases annually, but those data will be delivered beyond the time frame of the MRMD project. Note also that GIDAS, INTACT and NASS-CDS are data providers in iGLAD; hence, the corresponding data in iGLAD does not increase the total sample size for the analysis in Phase 2. Also, there are data from India and Australia in iGLAD, which may not be relevant for the MRMD project. This implies that the contribution of iGLAD to the sample size may be substantially smaller than the total number of crashes. A further difficulty is that each data provider may have their own set of inclusion criteria; therefore, any analysis of iGLAD data must account for such differences. Nevertheless, iGLAD may be an important data source in the project because it contains crash data from countries in Southern and Eastern Europe (namely Spain, Italy and the Czech Republic) in a standardized format.

New Zealand provides an interesting and potentially useful situation. Although it is not part of the US or the EU, it allows vehicles compliant with either set of regulations to be driven on public roads. Databases in New Zealand are available and in good condition. However, preliminary information from colleagues in New Zealand suggest that the number of US regulated vehicles driven is small and that the majority of them are either sports cars or heavy pickup trucks. Nonetheless, these data might still be useful for small-scale model testing. Models developed from EU and US datasets could be applied to the New Zealand data to test the success of the effort to separate modeling of injury risk from the population of crashes in the data. This activity would be useful, but is not critical to the project as a whole.

We also investigated Korea and Mexico, which similarly allow EU- and US-regulated vehicles. However, their laws do not lend themselves to this analysis. In Korea, manufacturers can only sell US-regulated vehicles if the total volume is low (<25,000 units). In Mexico, vehicles can be sold with safety systems removed and vehicles significantly altered. We did not further investigate the condition and access to databases in those countries.

Relevance of crash data to the crashworthiness and crash avoidance analysis

To summarize, the relevant datasets for crashworthiness analysis include detailed information regarding occupants in crashes, their crash circumstances, and their injury outcomes. The following in-depth databases have already been selected to be used in Phase 2 of the project, assuming that issues concerning access to the databases can be solved:

- NASS-CDS (US);
- GIDAS (Germany);
- INTACT (Sweden);
- iGLAD (Global);
- PENDANT (eight EU countries).

Other likely options, currently under consideration, are the following in-depth crash databases:

- CCIS (UK);
- LAB (France);
- VOIESUR (France).

For the crash avoidance analysis and for establishing standard populations of crash-involved occupants, we need national police-reported crash datasets; detailed crash severity and injury information are not required. The relevant datasets are as follows:

- NASS-GES (US);
- CARE (EU);
- National datasets from specific countries in the EU. National data from those countries where the relevant in-depth databases are from may be especially relevant; these include:
    a. DeStatis (Germany);
    b. STRADA (Swedish TRaffic Accident Data Acquisition; Sweden);
    c. STATS19 (UK);
    d. BAAC (Bulletin d'Analyse des Accidents Corporels de la Circulation; France).

For either domain, we will need to apply the conjunction of the most restrictive inclusion criteria from each considered dataset to all of the datasets. For example, CDS samples on tow-away crashes, and GIDAS chooses crashes where at least one person is injured. The GIDAS dataset will need to be restricted to only tow-away crashes, and the CDS dataset will need to be limited to only crashes where one person is injured. This way, the analysis can be conducted on the comparable subsets of tow-away crashes in which at least one person was injured. In addition, all variables referring to a certain physical quantity must mean the same thing and be measured in the same way. Further discussion of data harmonization is included under the section “Issues”.

Field Operational Test (FOT) Data

Field Operational Tests (FOTs) using both prototype and production crash-avoidance systems have been conducted in both the EU and the US. No such dataset includes national representation of drivers in any country, but all have been used to estimate effectiveness of these systems in their respective countries. A full list of FOT activities together with detailed information on each FOT study is provided at http://wiki.fot-net.eu/index.php?title=FOT_Catalogue. In particular, UMTRI has conducted a series of FOTs, including the recent Integrated Vehicle-Based Safety System (IVBSS) FOT, which involved vehicles equipped with both Forward Collision Warning (FCW) and Lane-Departure Warning (LDW). The EuroFOT project, in which SAFER was a key partner, collected data on a number of production systems, also including FCW and LDW. The capability of FOT data to contribute to the issue of functional equivalence of EU-registered and US-registered vehicles has been addressed by a literature review of FOT results; see “Assessing Crash Avoidance Benefits”.

ISSUES

For the methodology development several important issues have been identified. A brief summary of these issues is presented here. More detailed explanations, together with suggestions on how to address these issues, are provided in the next section.

One difficulty related to using crash data from multiple datasets relates to differences in sampling criteria. This aspect has been discussed above. Further, some of the key variables used for the modeling approach were coded and computed in different ways in different datasets and thus are not directly comparable. Due to limitations in the sample size of the European datasets, the threshold for the binary classification of the injury severity variable has to be aligned to a level that may not correspond to the injury severity limits addressed in regulations.

Another issue is the effect of the consumer information programs on the vehicle safety performance. Consumer test requirements may exceed regulatory requirements and thus the vehicles involved in real-world crashes may not represent the safety level requested in regulatory standards. These consumer rating systems can influence vehicle safety design and sales patterns, particularly for affluent sectors of the population. In some cases, the rating system is so important to marketing and market share for the vehicle that they act as de facto regulations.

The last issue described in this section is the evaluation of crash avoidance systems. A significant impact on traffic safety is expected by the introduction of crash avoidance and mitigation systems, but this is not easily measured because the presence of equipped vehicles is quite low in the available accident datasets. Further it is questionable how data from field operational test can be used to identify a crash avoidance and mitigation benefit for a specific system.

Data harmonization

Method of Calculating Delta V

One of the most frequently used parameters to describe the severity of a crash is the change in velocity, commonly known as ‘Delta-V’. It describes the vector difference between the inbound and the outbound velocity immediately before and after a crash event and therefore has a magnitude and a direction. In crash databases the Delta-V generally states the magnitude, whereas the direction is expressed in the ‘principal direction of force’ (PDOF). While the use of these terms is universal, the methods of calculating them differ between databases as they use different data collected from the crash site and vehicle damage measurements.

In general, with on-site crash data collection, it is possible to identify the inbound directions, collision points, and end positions of the collision participants. Thus, a trajectory-based reconstruction is feasible. The Delta-V and the PDOF are then calculated from the difference of the run-out and run-in momentum vectors. Databases that use this reconstruction method include the German GIDAS study, the Austrian project ZEDATU, and the Swedish INTACT project. Most often, a program called PC Crash is utilized, which is a momentum-based crash reconstruction program.

PC-Crash was validated in 1996 on its capability to estimate immediate pre-impact speeds and post-crash trajectories (Cliff & Montgomery 1996). Unfortunately the accuracy of the estimation of Delta-V was not part of the validation. It is believed that reconstruction results with PC Crash do not show a systematic bias towards higher or lower Delta-V values.

Most often when the crash site is investigated retrospectively, there is less chance to identify the input parameters for a reliable trajectory-based reconstruction. Nevertheless, the energy absorbed during the crash can be estimated by measuring the residual crush of the vehicle and applying an estimate of the stiffness to the measured crush area. From the crush energies and masses of the crash participants a Delta-V can be calculated, either neglecting or considering the restitution of the deformed area (Nolan et al. 1998).

The NHTSA in the US developed an algorithm based on an incremental impact model called ‘CRASH3’ which is used in programs like WinSMASH or AI Damage. WinSMASH is used for the NASS-CDS study in the US and AI-Damage was used for the CCIS project in the UK. The main difference between the programs is the application of different vehicle structural stiffness values that allow for vehicle fleet differences between the US and UK, and the consideration of a coefficient of restitution in AI Damage, which is not available in WinSMASH. Earlier studies have shown that the application of WinSMASH to frontal crashes result in an average underestimation of Delta-V by 23% compared to event data recorders (EDR) (Niehoff & Gabler 2006). The underestimation was identified to be strongly dependent on the vehicle type. The use of vehicle-specific stiffness coefficients reduced the underestimation to about 11%. This study also showed that adding a restitution coefficient would reduce WinSmash's underestimation to about 1%.

WinSMASH itself had a major update in 2008 when a new library of specific stiffness values for passenger vehicles was introduced. In 2009, with the updated stiffness values, the underestimation of Delta-V in straight and angled frontal crashes in car-to-car collisions was assessed to be 16%, which was a significant improvement from the older WinSMASH version (Hampton & Gabler, 2009; 2010). Another study showed an underestimation of Delta-V for frontal car-to-car crashes by 11% in cases of large overlap and by 17% in cases of small overlap (Ireus & Lindquist 2014). For side crashes, an overestimation of Delta-V by 13% was identified when impacted by a car and by 2.4% when impacted by light trucks or vans (LTV) (Johnson & Gabler 2014).

Table 4. Published estimates of difference between updated WinSMASH 2008 DeltaV and EDR by crash mode.

Damage based Delta-V Estimation (WinSMASH 2008)		
Crash Types	Other Details	Error compared to EDR
Frontal Crashes	Average	- 16%
	Small overlap	- 17%
	Large overlap	- 11%
Side Crashes	Impacted by car	+ 13%
	Impacted by LTV	+ 2.4%

It is distinctly possible that there is a biased relationship between the two methods of Delta-V reconstruction. To resolve the issue, case data from the Swedish INTACT project will be used to develop a relationship between Delta-V calculated using WinSMASH, as is done in the NASS-CDS database (which is similar to the AI Damage protocol), and a trajectory-based reconstruction using PC Crash. This is possible because INTACT< unlike other in-depth databases, contains sufficiently detailed information to allow application of both damage-based and trajectory-based reconstruction methods. INTACT allows an export of relevant crush measurements for usage in WinSMASH (pre- and post-2008 version) and AI Damage, and thus a side-by-side comparison of Delta-V from PC Crash, WinSMASH pre-2008, WinSMASH post-2008, and AI Damage. In case the methods produce different results, a means for adjusting a bias among the three methods will be developed.

So far, 49 crashes involving 67 passenger cars have been reconstructed with PC-Crash in INTACT that have sufficient data for calculating Delta-V using WinSMASH. There may be other cases in the database that could be used for the comparison.

Appendix C contains a more detailed explanation of the different methods of Delta-V calculation.

Injury Definition

The preferred injury outcome is to examine injures of severity AIS3 or greater, as well as all fatalities (MAIS3+F), because these injuries most closely correspond to the severity addressed by regulatory test criteria. For example, the neck intercept criteria N_{ij} = 1 corresponds to a 22% probability for AIS3+ neck injury and the chest acceleration level of 60g corresponds to a 25% probability of AIS4+ chest injury. However, preliminary review indicates that the sample of MAIS3+F injured passenger car occupants will be too small in EU datasets if this injury level is used. Instead, we will use MAIS2+F as the injury outcome.

In doing so, we are implicitly assuming that ratios of more severe injuries are constant across crash/vehicle/occupant characteristics, and that results related to MAIS2+F apply to higher injury levels (e.g., MAIS3+F or fatality). This is known as the “proportional odds assumption” (UCLA 2014)and it will need to be tested. To the extent that the proportional odds assumption is not supported, we will need to develop a model of how MAIS3+F (and F only) are related to MAIS2+F. This can be done using the larger NASS-CDS dataset. The logic is that because bodies are physiologically the same in the EU and US, the relationship between risk of MAIS2 and risk of MAIS3 is simply a function of biomechanics and should not be different in the EU and US. However, some crash/vehicle/occupant characteristics might have different associations between risk of MAIS3+F and MAIS2+F. An example is given below that shows how CDS has been used for a similar purpose.

Figure 3 and Figure 4 are taken from a previous analysis, and use CDS data to estimate the risk of AIS2+ and AIS3+ injury by body region from 1999-2010. The head, thorax, and lower extremities have the highest risks at each injury level, but the order changes with severity. Upper extremity injuries are also among the body regions with higher risk when considering AIS2+ injuries, but not AIS3+ injuries. This initial look shows that trends with time for each body region tend to be similar for each injury level.

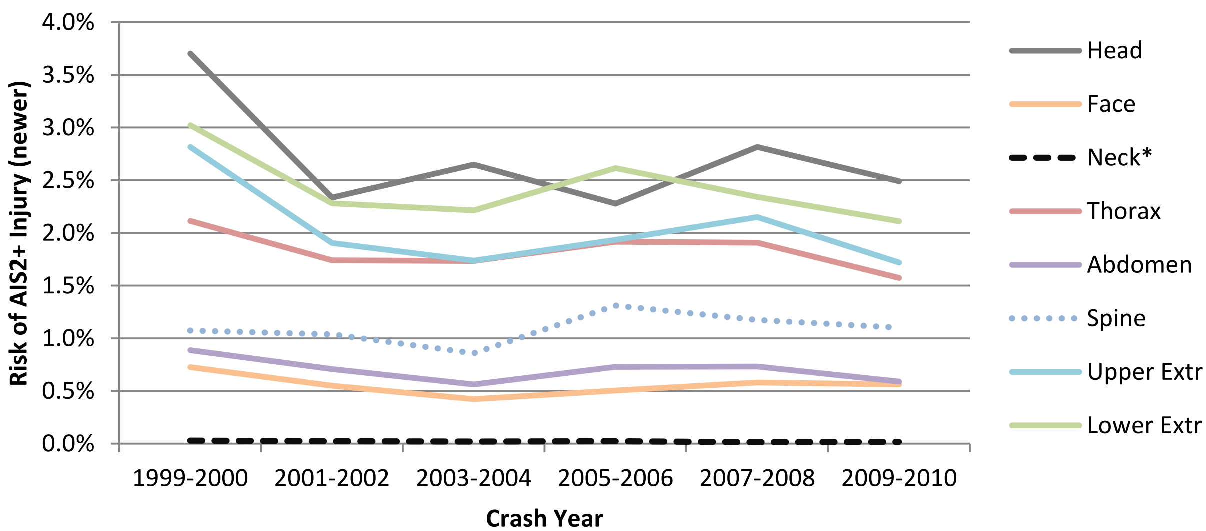


Figure 3. Risk of AIS2+ injury to each AIS body region for each crash year (vehicles aged 10 years or less)

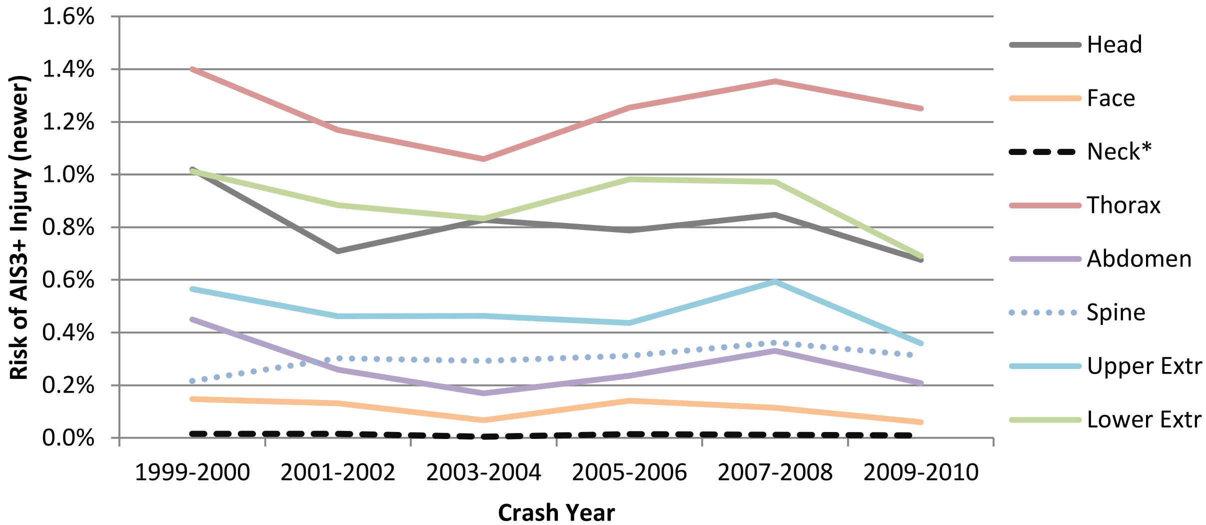


Figure 4. Risk of AIS3+ injury to each AIS body region for each crash year (vehicles aged 10 years or less)

The interpretation of the figures in this context would be an adjustment to the interpretation of MAIS2+F risk in different crash types. As a hypothetical example, if young occupants with MAIS2+F injuries have a preponderance of upper-extremity injuries and older occupants with MAIS2+F injuries have a preponderance of thorax injuries, then a much larger percentage of older occupants would be expected to also have MAIS3+F injuries, compared to young people. Thus, a 2:1 risk ratio for older vs. younger occupants with MAIS2+F injuries might translate to a 4:1 risk ratio for MAIS3+F injuries. These relationships can be measured using CDS data to allow for a more nuanced interpretation of results. We will work under the assumption that risk differences in MASI2+F injuries between EU- and US-regulated vehicles will translate to same-directional risk differences in MAIS3+F injuries as well. Under this assumption, the magnitude, but not the direction of differences, is the issue with using MAIS2+F as the dependent measure.

Consumer Information Programs

Both the US and the EU vehicle markets are influenced by consumer rating schemes and reward systems. NHTSA runs extra tests to provide more information to consumers through its NCAP process. These tests are run at a higher severity than regulatory tests, and results from frontal, side impact, and rollover tests are used to generate vehicle star ratings. The Insurance Institute for Highway Safety (IIHS) is an independent organization funded by the insurance industry, and it publishes safety ratings for vehicles based on several tests that differ from those used in NCAP evaluations.

From a crashworthiness standpoint, Euro NCAP includes several types of barrier tests run at higher levels than required by regulation. These include a 40% offset frontal, a pole side impact test, and a barrier side impact test. The program also evaluates vehicle features intended to mitigate whiplash injuries and pedestrian injuries. For child occupants, child restraint compatibility, performance in front and side impacts, and inclusion of other child-friendly safety features are rated.

Consumer information programs also provide information regarding crash avoidance technologies. In addition to the star-rating scheme, Euro NCAP also has a reward system (called Euro NCAP Advanced) for safety technologies or crash avoidance systems. In their current reward system, Euro NCAP has tested Blind Spot Monitoring, Lane Support Systems, Speed Alert Systems (ISA), Autonomous Emergency Braking, Attention Assist, Automatic Emergency Call (eCall), Pre-Crash Systems, Vision Enhancement Systems, and Multi Collision Brake. Compared to Euro NCAP, NHTSA only includes three active safety systems in their reward systems; these are Electronic Stability Control, Lane Departure Warning, and Forward Collision Warning. The reward systems help to increase public awareness of these rather new safety systems. It should be noted here that the availability of the crash avoidance technologies are not part of the 5-Star Ratings System by NHTSA. The IIHS recently added assessments of Frontal Collision Warning systems and automatic braking systems to their vehicle ratings. Appendix D specifically summarizes how different crash avoidance technologies are addressed by different consumer information programs.

The proposed analysis will primarily focus on regulations and not consumer information programs, but the Phase 2 analysis will explore the use of vehicle rating it as a predictor to estimate its effect. The goal is to remove any effect of purchase decisions that emphasize vehicles that perform above minimum (regulatory) standards. However, consumer information programs are not barriers to selling vehicles in the marketplace, and are thus not the focus of the analysis. Some of the rating systems change over time and only a subset of vehicles are tested and rated, so it may not be possible to associate every vehicle in the database with a rating. The high quality datasets are largely from countries that are wealthy and where vehicle purchasing may be driven by ratings.

Assessing Crash Avoidance Benefits

In the realm of crash avoidance, more and more vehicles are now equipped with advanced driver assistance systems, such as frontal collision warning (FCW) and lane departure warning (LDW). It is highly likely that such systems would be available in even more vehicles in the future, both in the EU and the US. Crash avoidance research efforts, particularly field operational tests (FOTs), were reviewed to determine if they could prove useful to identify whether such systems operate in the same way and whether they would give the same safety performance when being used in the EU and the US. Results were not clear with respect to how one could use information on the potential safety benefit of certain advanced driver assistance and crash avoidance systems for the purpose of this project. We illustrate the issues by taking one project in the EU and one project in the US as examples.

Though there have been many research projects conducted to evaluate the effectiveness of advanced driver assistance and crash avoidance systems, most such studies were done at a national level (i.e., per country) or even a smaller scale. To our knowledge, the EuroFOT project (Kessler, et al. 2012) was the first that tried to evaluate the effectiveness of such systems at a scale beyond a national level. Forward collision warning systems, lane departure warning systems, and adaptive cruise control systems from different vehicle manufacturers were tested in different European countries. EuroFOT also tried to estimate the potential safety benefit of these three advanced safety functions at EU level. However, an estimation of the potential safety benefit at the EU level could only be made for a combination of forward collision warning and adaptive cruise control (see Table 1 in (Malta et al., 2012)). Besides the three mentioned systems, the EuroFOT project also tested six other systems, but each of them was tested at a single-country level or smaller.

A recent example of FOT studies in the US is the Integrated Vehicle-Based Safety Systems (IVBSS) project (Sayer et al. 2011). The systems studied in IVBSS were forward crash warning, lane drift warning, lane-change/merge warning, and curve speed warning, tested on light and heavy vehicles. Name-wise, these systems are very similar to the ones tested in EuroFOT. However, the implementation of the systems are different; the ones tested in EuroFOT are systems that were already available on the market at the time of the FOT from the different vehicle manufacturers, while those tested in IVBSS were prototype versions, and did not represent particular vehicle brands. For the light vehicles, IVBSS provided an estimation of the potential safety benefit of the integrated system (i.e., a bundle of four systems) and forward crash warning as an individual system (see Table 25 in (Nodine, Lam, Stevens, Razo, & and Najm, 2011)).

The two projects used similar main ideas to estimate the safety impact on the EU level (for EuroFOT) and for the whole US (for IVBSS). That is, they first identified and calculated the target crash populations for the different systems, then identified changes related to safety by comparing FOT data with and without systems. Finally, they projected these changes in terms of reduction of crashes on the national level or beyond. While the main ideas on how to estimate the safety impact on the EU level and the whole US are similar, the results cannot and should not be compared directly. This is because the way the estimation task was implemented in the two projects is likely to be different, as it depends on the availability of crash databases in the area of interest, how detailed the available crash databases are, specific implementation of the systems (e.g., a system is designed to work only above certain speed), and the indicators and indicator levels were used to determine changes with and without systems. Furthermore, the effectiveness of the driver assistance and crash avoidance systems could be affected by the traffic environment and general driving style on the area of usage. For example, a majority of lane departure warning systems (including all the variants) depend on clear lane markers, so the system may not function well due to snow, mud, heavy rain, fog, road conditions, unusual/indistinct lane marker, etc. (see e.g. “What are the limitations” at http://www.euroncap.com/rewards/infiniti_ldp.aspx). This means that the effectiveness of lane-departure warning systems in a specific country may not reflect the effectiveness of the same system in another country with very different road conditions and weather conditions.

Besides having safety benefits, driver assistance and crash avoidance systems could potentially have risks too. For example, drivers might adapt to be less attentive ((Bayly, Fildes, Regan, & Young, 2006) and (Wege, Will, & and Victor, 2013)). However, further research is needed to be able to measure this aspect.

To conclude, knowledge from current literature does not give enough evidence to say with sufficient confidence whether or not the vehicles equipped with advanced assistance and crash avoidance systems that have been tested in the EU would have the same safety performance in the US or the other way around. The main difficulties are the following:

- Lack of results on EU level;
- Results are often available concerning a combination of systems; such results could only be compared with the results from the same combination of systems;
- Even systems which are called the same name may actually be different (e.g., how they were implemented and their operation range);
- Potential side-effects of such systems (e.g., behavioral adaptation) have been indicated in a handful of cases, but have not been quantified.

STATISTICAL METHODS

Overview of Statistical Approach

A key challenge of this analysis is that because of current regulations (i.e., separate regulatory environments for the EU and US), there are no field data that speak directly to the question of whether US-regulated and EU-regulated vehicles perform similarly when driven in the other region[11]. To address this, we must use statistical models to represent expected real-world performance of the two vehicle groups as well as the expected driving/crashing environment in which they might be driven in the future. We also need models of how other-region vehicles are likely to enter the fleets, because this will influence the expected driving/crashing environment as well. Finally, we need a way to compare the performance of these models to estimate relative risk associated with mutual recognition.

To explore evidence for the stated hypothesis (of real-world equivalence), we require seven model components:

a. A statistical model of injury risk to an occupant of an EU-regulated vehicle, *given* the conditions of any crash/occupant/vehicle combination
b. A statistical model of injury risk to an occupant of a US-regulated vehicle, *given* the conditions of any crash/occupant/vehicle combination
c. A standard population of crashes in the EU, described by crash/vehicle/occupant characteristics; this population must arguably represent a likely near-future crash population for the EU
d. A standard population of crashes in the US, described by crash/vehicle/occupant characteristics; this population must arguably represent a likely near-future crash population for the US
e. One or more models of how US-regulated vehicles might enter the EU market
f. One or more models of how EU-regulated vehicles might enter the US market
g. A means of measuring the evidence for how injury risk in EU- and US-regulated vehicles is likely to differ (or not differ) in a particular crash population

[11] Note: Both EU- and US-regulated vehicles are driven in a common driving environment in New Zealand. Those data do not address the basic question of how EU and US vehicles would perform side-by-side in the US driving environment or EU driving environment.

In the overview that follows, we will use crashworthiness, or injury risk, as an example. To develop a statistical model of crashworthiness in EU-regulated vehicles, we can use field data from the EU. Our model will predict injury risk to passengers of EU-regulated vehicles involved in crashes, *as a function of characteristics of those crashes.* This can be represented as in Equation 1.

$$\textit{predicted probability of injury} = f(\textit{crash, vehicle, occupant characteristics}) \qquad (1)$$

In Equation 1, *f* is a function that takes crash, vehicle, and occupant characteristics and returns a probability value between 0 and 1. That probability value is our best guess as to injury risk for a specific occupant in a specific vehicle and crash. However, the model will also have associated uncertainty that results in a *distribution* of predicted injury risk for each case. This is shown in Equation 2 below.

$$\textit{predicted distribution of injury risk} = f(\textit{crash, vehicle, occupant characteristics}) + \textit{uncertainty} \qquad (2)$$

Equation 2 gives a more complete picture of the model of injury risk. Rather than predicting a single value of injury risk for each case, the model returns a probability distribution of predicted risk. Thus, for example, if a 50-year-old belted male driver were involved in a 30-km/hr Delta-V crash in an EU-regulated vehicle, the model might predict that his true value of injury risk is estimated according to the distribution in Figure 5. That is, the true value is unknown (and unknowable), but the model predicts that the true value is most likely to be 0.05, less likely (but still possible) to be 0.04 or 0.06, and extremely unlikely to be 0.02 or 0.08 (and so on).

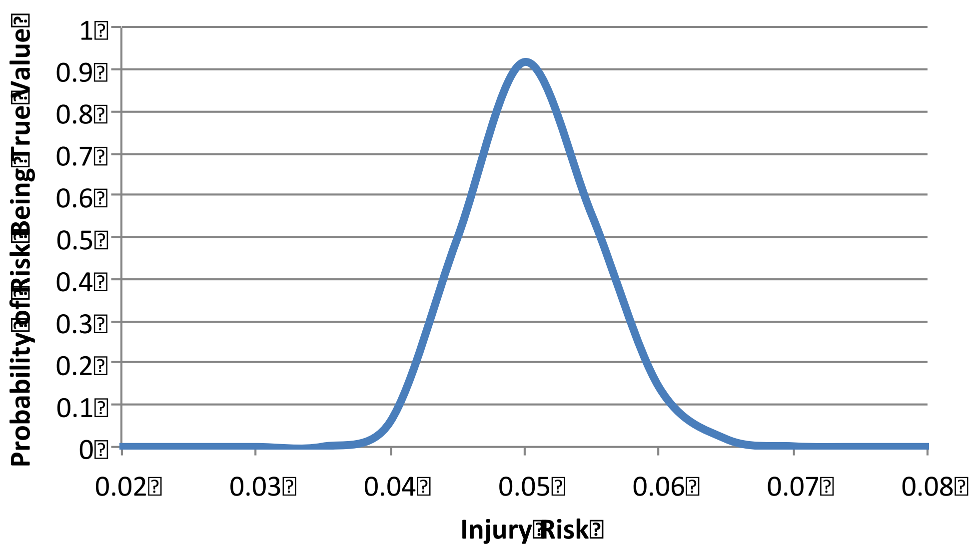


Figure 5. Illustration of risk model with uncertainty. In this example, the distribution represents the probability that each value is the true value of injury risk for a single combination of crash, vehicle, and occupant characteristics.

The same process can be used with US crash data to develop (b), the model of injury risk in US-regulated vehicles as a function of crash, vehicle, and occupant characteristics. Once again, we must account for the level of uncertainty in the model.

The key role of the injury risk models, (a) and (b), is to separate injury risk from the driving environment. This is critical because we are trying to envision how occupant risk will be affected by putting a set of vehicles into a *different* environment. For example, rollover crashes may be less common in the EU compared to the US. However, if an EU-regulated vehicle is sold in the US, it would be expected to experience a rollover at the US rate, not the EU rate. To predict how this might work in the future, we must estimate risk of injury *given* a rollover for the EU-regulated vehicles, independent of whether rollover is common or uncommon in the EU. Similarly, larger vehicles are more common in the US than in the EU; hence, small vehicles may experience lower-severity crashes in the EU, on average, compared to the US (because they are less likely to hit larger vehicles). However, the injury risk model for the small EU vehicle should predict injury risk *given* a specific Delta-V value. Thus, when sold in the US, the same vehicle would be expected to experience a larger number of high Delta-V crashes than it would have in the EU (because it will crash with larger cars on average), but we can predict the injury risk on the basis of the new expected Delta-Vs that that vehicle would experience in the US environment.

Once developed, the EU and US injury risk models must be compared. The simplest comparison would be to determine whether the models are identical. If the risk models are identical, then performance will be the same, regardless of the exposure of EU- and US-regulated vehicles to different types of crashes.

The first of our three proposed approaches, Seemingly Unrelated Regression, or Approach 1, tests this hypothesis. A critical challenge in testing this hypothesis is that because we are interested in the strength of evidence that the models are the same, traditional hypothesis testing is not appropriate. Traditional hypothesis testing sets the hypothesis of "same models" as the null, or default, hypothesis and only looks at evidence that the observed data *violate* that hypothesis. However, evidence against rejecting the null hypothesis is not the same as evidence *for* the null.

Thus, in Approach 1, we must measure the capability of the data and model to detect differences of various sizes when they do exist, in addition to measuring the evidence for a difference. Depending on the size of differences that are considered important, this method may or may not provide a conclusive result.

In the case where the two models are shown to be different or evidence is inconclusive, it is not necessarily the case that the overall consequences across the crash population will be different as well. Indeed, risk differences under different crash conditions are likely since the regulatory environments may seek to optimize performance for the crash conditions most relevant to the region. However, since populations of crash conditions are quite varied, different risk models can produce similar or different overall injury risk.

To take a simple example, suppose that there are two types of crashes, *A* and *B.* Table 5 shows a hypothetical scenario for type-specific injury risk and overall injury risk. In this example, EU-regulated vehicles have 1% injury risk in crash type A and 2% injury risk in crash type B. The EU crash population is made up of ¾ crash type A and ¼ crash type B. US-regulated vehicles have the reverse risk and the reverse population proportions. Thus, one could argue that regulation has been optimized for the different crashing environments. When a US-regulated vehicle is driven in the EU, it will be exposed to the EU crash environment and because injury risk is higher in the more common crash type A, overall injury risk across all crashes will be higher for US-regulated vehicles driven in the EU environment. The risk ratio (EU/US) is less than one, indicating that occupants of EU-regulated vehicles are at lower risk of injury in the EU crash environment. However, when EU-regulated vehicles are driven in the US crash environment, they are exposed to more of crash type B. The pattern reverses in this situation, and the risk ratio (EU/US) across all crashes in the US population is greater than one, indicating better performance for US-regulated vehicles in that driving environment.

Table 5.
Illustration Using Hypothetical Risk and Crash Type Scenarios

Region	Injury Risk		Proportion of Population		Overall Injury Risk Across All Crashes	
	Crash Type A	Crash Type B	Crash Type A	Crash Type B	EU Crash Environment	US Crash Environment
EU	0.01	0.02	0.75	0.25	0.01*0.75+ 0.02*0.25= 0.0125	0.01*0.25+ 0.02*0.75= 0.0175
US	0.02	0.01	0.25	0.75	0.02*0.75+ 0.01*0.25= 0.0175	0.02*0.25+ 0.01*0.75= 0.0125
Risk Ratio (EU/US)					0.714	1.40

As discussed earlier, Approach 1 tests the hypothesis that the risk models for EU- and US-regulated vehicles are the same. If they are, then the vehicles are functionally equivalent, even if the crash populations differ; this means that occupants of EU-regulated vehicles will have the same injury risk when driving in the US as the occupants of US-regulated vehicles, and vice versa. In contrast, Approach 2 assesses the consequences of models that are either shown to be different or for which evidence for "same" vs. "different" is inconclusive. In Approach 2, both models will be exercised on the EU crash/vehicle/occupant population as well as the US crash/vehicle/occupant population. Assessment of the overall risk for the two populations will be done in parallel and results cannot be merged. Thus, as in Table 5, there will be two results.

To exercise the risk models, we need a standard population of crashes that represents crashes in the EU (i.e., item (c)) and another standard population of crashes for the US (item (d)). In the crashworthiness case, this population will consist of a large number of occupants who were involved in crashes, along with the crash/vehicle/occupant descriptors that are needed for both risk models. The collection of occupants should approximate the likely distribution of crash-involved occupants in each region in the near future (i.e., the time when mutual recognition could be in effect). Thus, if older occupants make up 20% of crash-involved occupants in the EU, then the EU standard population should include about 20% older occupants. Similarly, if rollovers make up 5% of crashes in the EU, then about 5% of occupants in the standard population should be in rollovers.

For the US standard population, we will use the NASS-CDS population for 2012. CDS populations are used by NHTSA to estimate benefits and consequences of regulations and NCAP ratings. For the EU standard population, GIDAS employs a weighting method adopted by the European Commission (EC) for EuroNCAP Advanced Technologies Assessment. That method will be replicated in this analysis.

The example in Table 5 does not address the model uncertainty that was discussed after Equation 1 and shown in Equation 2. In exercising the risk models on the standard populations, we must also account for model uncertainty. Thus, the results of Approach 2 will be *distributions* that represent the probability of a variety of possible true overall consequences. The less uncertainty there is regarding the predicted probability of injury, the less uncertainty there will be in the bottom-line risk ratios.

Another issue that must be addressed in Approach 2 is the nature of the expected fleet penetration of other-region vehicles into the standard populations (items (e) and (f)). The simple model illustrated in Table 5 and discussed afterwards would compare the overall risk ratio across *all* crashes in each population. However, it is unlikely that US-regulated vehicles will replace EU-regulated vehicles in the EU fleet at random (and vice versa). Instead, it is likely that certain vehicle types will be more often purchased from among US-regulated vehicles and other vehicle types will be more often purchased from among EU-regulated vehicles. There is no expectation that mutual recognition would substantially change the fleet *composition* in either region with respect to vehicle type (e.g., percent of small cars vs. large cars), but if fleet penetration of other-region vehicles differs by vehicle type, bottom-line risk can be affected. Although (e) and (f) are not developed from data, they will need to be discussed and agreed upon.

Approach 2 provides a nuanced way of assessing risk models and their likely consequences. However, it is focused on the single best risk model for each vehicle group (EU-regulated and US-regulated). Approach 3 turns the question around, and instead of focusing on a single model, it assesses the evidence for groups of models that produce equal consequences.

In contrast with the first two approaches, Approach 3 evaluates a large number of possible models of risk in EU- and US-regulated vehicles. Many models are highly unlikely, but many other models are only slightly less well supported by the data than the best model. Moreover, there are many different models that lead to the same population-wide risk ratios.

Using Bayes Factors, it is possible to measure the evidence for groups of models. Models will be grouped according to their population-wide outcome for the EU standard population (using the appropriate fleet penetration model), and separately grouped according to their population-wide outcome for the US standard population. For a given group of possible models, evidence can be measured and also compared to the evidence for a different group of possible models. Bayes Factors are defined as the ratio of evidence for one group of models over another group, thus providing a measurement of the comparison of evidence. Approach 3 is computationally more intensive than the other two approaches, but it has the advantage of addressing the question in a different way, thus triangulating an answer to the original question of essential equivalence in field performance.

In summary, we plan to take three different analysis approaches to assess the evidence for various injury risk models, focusing on the consequences of these models and the evidence for them. Results from the three approaches form a more complete picture of how the regulatory environments affect injury or crash risk and how this would affect risk if mutual recognition were adopted. The three proposed strategies are:

1) Test the hypothesis that EU and US injury risk models are identical;
2) Choose the best EU and best US models independently and assess injury consequences, taking into account model uncertainty;
3) Assess relative strength of evidence for groups of models that result in equal consequences.

Throughout the next sections, we discuss the methods applied to crashworthiness analysis, which will use logistic regression. The same methods will be applied to the crash avoidance analysis, but we will use Poisson or Negative Binomial regression models for that case rather than logistic regression models. These models fall into the same class of general linear models as logistic regression, but are most appropriate for count data (as opposed to binary outcome data).

Statistical Modeling

Logistic Regression

The purpose of logistic regression in this methodology is to develop models of injury risk as a function of crash, vehicle and occupant predictors for EU- and US-regulated vehicles. Logistic regression is used in a wide variety of applications when the response variable has a small number of possible outcomes. The binary outcome case is most common, and will be described here.

Suppose the response variable, *y,* for an occupant is assigned a value of 1 if a crash resulted in MAIS2+F injury to the occupant, and is assigned the value of 0 if a crash resulted in little or no injury to the occupant. The dataset also has a set of *r* predictors, $x_1, x_2 \dots x_r$ each of which describes a characteristic of the crash, vehicle, or occupant in each case. Predictors might include direction of impact, weather conditions, and occupant age, among others.

The logistic regression model uses these data to predict the *risk* of MAIS2+F injury, given the predictors, according to the formula given in Equation 3.

$$\hat{p} = \frac{1}{1+e^{-(\beta_0+\sum_{i=1}^{r}\beta_i x_i)}} \tag{3}$$

where $\hat{p}$ is the predicted probability of injury, β_0 is the intercept, the β_i are coefficients of the predictors, and x_i's are the predictor values. The model-fitting process results in estimates of the β's and an additional error component that measures model uncertainty using any lack of fit of the model to the data.

Logistic regression is a type of general linear model in that a simple transformation of the predicted outcome is related to a linear function of predictors (though individual predictors can also be transformed). Here, the *odds* of serious injury are defined as p/(1-p). Logistic regression models the ln odds of MAIS2+F injury for each occupant on the left side of the model equation as a linear function of predictor variables on the right side of the model equation. For N occupants, the model equation is shown in Equation 4.

$$ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_r x_{ir} \qquad i = 1, \dots N \tag{4}$$

where β_0 is the intercept parameter, $\beta_j,\ j = 1, \dots, r$ are regression coefficients or slope parameters, and x_{ij} are known predictor variables such as Delta-V or occupant age. Note that this formulation is equivalent to Equation 3, but focuses on the linear portion of the equation. Like many models, logistic regression models are fit using the method of maximum likelihood, which is described in more detail in a subsequent section.

Since the left side of the model equation represents the ln odds of MAIS2+F injury, the slope parameters have interpretations as ln odds ratios. For a continuous predictor *x* such as Delta-V, the slope parameter attached to it represents the ln odds of MAIS2+F injury for a unit increase in Delta-V at fixed values of all other predictors. So if the coefficient for Delta-V in the model is 0.039, then the corresponding odds ratio for Delta-V is $e^{0.039} = 1.04$. This odds ratio indicates that the odds of MAIS2+F injury increase by 4% for each increment of Delta-V if all other predictors are kept constant. For a binary predictor such as seat belt use that is coded 0 if the occupant did not use a seat belt and 1 if the occupant did use a seat belt, the slope parameter represents the ln odds of MAIS2+F injury for occupants who wore a seat belt, compared to those who did not. If the coefficient in the model for belt use is -0.69, the odds of MAIS2+F injury for a belted occupant is 50% of the odds for an unbelted occupant ($e^{-0.69} = 0.5$). Because multiple predictors are included in the model, an odds ratio for a variable is interpreted under the assumption that it has been adjusted for the other predictors included in the model.

In this context, coefficients can be interpreted as describing the performance of occupant protection systems in a variety of contexts. For example, a positive age coefficient means that older occupants are at greater risk than younger occupants. This would be expected based on fragility of older people. However, a shallow age slope means that occupant protection works nearly as well for older occupants compared to younger ones, whereas a steep age slope means that younger occupants are protected much better than older occupants. Taking this a step further, the coefficients of the EU and US models can generally be compared, and any differences in those coefficients reflect differences in occupant protection systems in vehicles. Note, however, that coefficients are always in the context of the entire model—the presence of other predictors will affect the magnitude of a given coefficient. Thus, a simple comparison does not tell the whole story, and it is necessary to implement the model on the entire crash population to understand how two different models actually compare for different cases, as well as across the population as a whole.

An additional challenge surrounds estimation of the intercept, or β_0. With logistic regression, coefficients of predictors are unbiased, even when the underlying sample is biased (Breslow, 1996). For example, the underlying sample might use inclusion criteria that emphasize more severe crashes, but the coefficients will be the same as they would be with a random sample of crashes. In contrast, the intercept is biased when the sample is biased. If inclusion criteria for the EU and US datasets are equated, the intercepts should correctly reflect the risk of that type of crash. However, if inclusion criteria are implemented differently, the biased intercepts could introduce overall risk differences that do not reflect *true* risk differences.

To address this challenge, we can use cases where risk should be the same to calibrate the model. One example of such cases might be those where specific elements of specific vehicles are identified that are sold unchanged in both the US and the EU. In principle, identical safety systems should produce identical injury risk under identical crash/vehicle/occupant conditions. If the models produce risk differences under the same crash conditions (relevant to the unchanged safety systems), then the intercepts may be out of calibration and will need to be adjusted accordingly.

One advantage of logistic regression is that when the outcome of interest is rare, an odds ratio can be interpreted as relative risk. For crash outcomes, serious injury to occupants is rare compared to no injury. In other words, in crash datasets, the binary *y* variable often has many more 0 outcomes (no injury) than 1 outcomes (serious injury). In that case the odds ratio approximates a relative risk and it is appropriate use logistic regression as an exposure-based risk model (Greenland & Thomas, 1982).

Poisson Regression

Unlike logistic regression, in which the response is binary, the response variable for Poisson regression is a count. The Poisson model is the standard model for the analysis of rates. The numerator of the rate is a count, such as the number of crashes, and the denominator of the rate is a measure of exposure, such as vehicle-miles traveled (VMT). Therefore, data used in a Poisson model are aggregated in the sense that each observed rate is a ratio of a sum of crashes and a sum of exposures taken over combinations of predictor variables.

Poisson regression models the ln rate on the left side of the model equation as a linear function of predictor variables on the right side of the model equation. For N rates, the model equation is Equation 5.

$$ln\left(\frac{\mu_i}{t_i}\right) = \alpha + \beta_1 x_{i1} + \cdots + \beta_r x_{ir} \qquad i = 1, \ldots N \qquad (5)$$

where μ is the expected count and t is the measure of exposure in the denominator. The parameter α is the intercept, $\beta_j,\ j = 1, \ldots, r$ are regression coefficients or slope parameters, and x_{ij} are known predictor variables such as road type, time of day, or area of operation (rural/urban). As with logistic regression, Poisson models are fit by the method of maximum likelihood.

A relative risk (RR) is the ratio of two rates. If two rates are the same, then the RR=1 or the ln RR=0. Since the left side of the model equation represents the ln rate of crashes per unit of exposure, the slope parameters have interpretations as ln RRs. For a binary predictor such as rural/urban that is coded 1 for crashes occurring in rural areas, and 0 for crashes occurring in urban areas, the slope parameter represents the ln RR that compares crash rates in a rural area relative to an urban area.

As noted above, crash avoidance analysis using Poisson regression requires a measure of exposure. If sufficient exposure data is not included in current available datasets, we may use quasi-induced exposure instead (e.g., Cuthbert, 1994; Keall & Newstead, 2010).

Negative Binomial Regression

Often, the Poisson distribution is not sufficiently flexible to model the variance in a given dataset. The Poisson has only one parameter, so it represents a strong assumption about the relationship between the mean and variance of the rates (i.e., that they are the same).

In cases where variance is greater than the Poisson model suggests (called *overdispersion*), the negative binomial can be used instead. The modeling structure and process are exactly the same—ln rates are a linear function of predictors and coefficients. However, a separate parameter is estimated to better account for variance in the data.

In the present context, Poisson and Negative Binomial models serve the same function. We will use the best-fitting model to estimate crash risk as a function of parameters.

Likelihood

Since likelihood forms the basis of our statistical methodology for all three approaches, a brief overview is provided here. If we consider all possible risk models, we might include models such as those that take the forms listed below:

$\hat{p} = -6 + 0.14 * dV;$

$\hat{p} = -7 + 0.1 * (dV)^2 + 0.2 * age;$

$\hat{p} = -12 + 0.2 * age - 1.2(if\ belted) + 0.9 * \cos(direction\ of\ force);$

where *dV* is the change of velocity (also known as Delta-V), and $\hat{p}$ is the predicted risk of injury given a crash.

The evidence provided by the data for every model can be characterized by its *likelihood*, which is the probability of the data given the model (Equation 6).

$$L = \prod_D P(\boldsymbol{D}|\boldsymbol{\theta}) \tag{6}$$

where $\boldsymbol{\Theta}$ is a vector of coefficients that describes the model, $\boldsymbol{D}$ is the data, P is a probability measure, and $\boldsymbol{L}$ is the likelihood, or the product of all the individual probabilities of each data element under the model described by $\boldsymbol{\Theta}$**.** We can visualize this by creating a two-dimensional[12] space defining the parameters being considered. For this example, we use a simple logistic model with one predictor (*dV*) and two parameters (intercept and slope of *dV*, denoted by β_0 and β_1). The *predicted risk* of an MAIS2+F injury given a crash in this model, denoted by $\hat{p} = \hat{p}(dV)$, is shown in Equation 7.

[12] The number of dimensions will be two for the simple model that follows; in general, it depends on the number of predictors.

$$\hat{p} = 1/_{e^{\beta_0+\beta_{dV}*dV}} \quad (7)$$

Note that this value is the unique solution of the logistic regression equation (Equation 8).

$$ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = \beta_0 + \beta_{dV} * dV \quad (8)$$

The data are a series of observations of occupants in crashes, visualized in Table 6. Each observation has a value of Delta-V and an injury outcome. The injury outcome is treated as a binary variable, set to 1 if the occupant has an injury corresponding to AIS2 or higher (or fatality), and 0 if they are uninjured or have AIS 1 injuries only.

Table 6. Theoretical occupant dataset

Occupant Number	Delta-V	Injury outcome (injured=1)
1	15	0
2	30	1
3	10	0
4	12	1
...	...	...

Given a set of observed data, we can compute the probability of getting those particular outcomes (given the particular crashes), as the product of each individual outcome probability, under a particular model. For the example in Table 6, the likelihood is computed as in Equation 9.

$$= [1-\hat{p}(15)] * \hat{p}(30) * [1-\hat{p}(10)] * \hat{p}(12) * \ldots = \left[1 - 1/_{1+e^{-(\beta_0+\beta_{dV}*15)}}\right] * \ldots \quad (9)$$

Over the whole parameter space in this example (i.e., for each value of β_0 and β_1), the likelihood is a smooth function and can be graphed as shown in Figure 6. In logistic regression, the pair of parameters at the peak of the likelihood surface is chosen as the maximum likelihood estimator of the true parameters[13]. In other words, that parameter pair is the one most supported by the data; of all models of the form in Equation 7, it is the one with exactly this β_0 and β_{dV} pair (namely, β_0=-1, β_{dV} =0.9 in Figure 6) which maximizes the probability of seeing the set of outcomes (Table 6) that have been observed in the reality.

[13] The expression "true parameters" refers to the pair of parameters for which the predicted injury risk equals the actual real-world injury risk for the considered group of vehicles (e.g. US-certified vehicles).

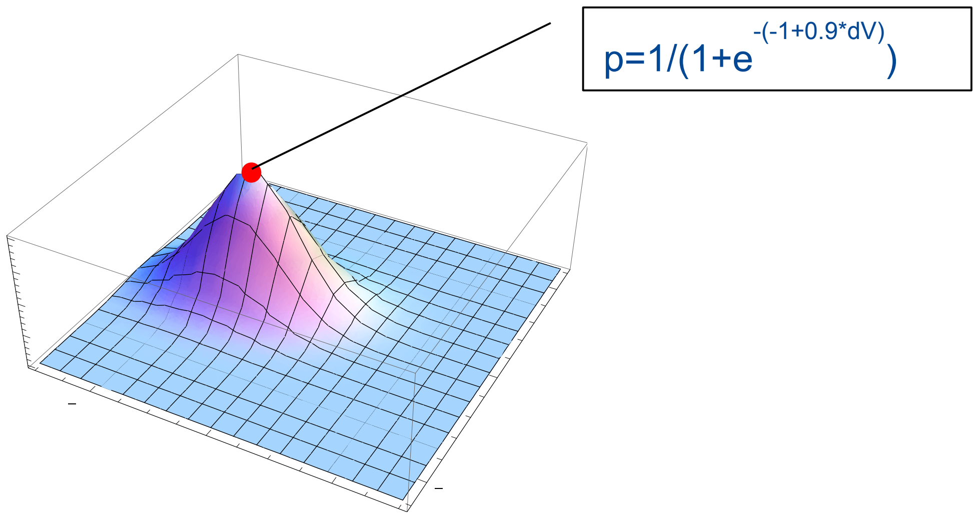


Figure 6. Illustration of likelihood of outcomes;
the maximum likelihood corresponds to the peak.

The next step would be to consider the likelihood space for EU vehicles separately from US vehicles, as visualized with hypothetical data in Figure 7. The EU model in this hypothetical example is of the same form as the US model; i.e., it also is described by Equation 7. The underlying data are different from crash-involved occupants of US vehicles. This explains why the likelihood surfaces are different.

For the EU model, it again is possible to estimate the true parameters by the pair of parameters at the peak of the EU likelihood surface; the maximum likelihood estimators may either be the same or different from those in the US model.

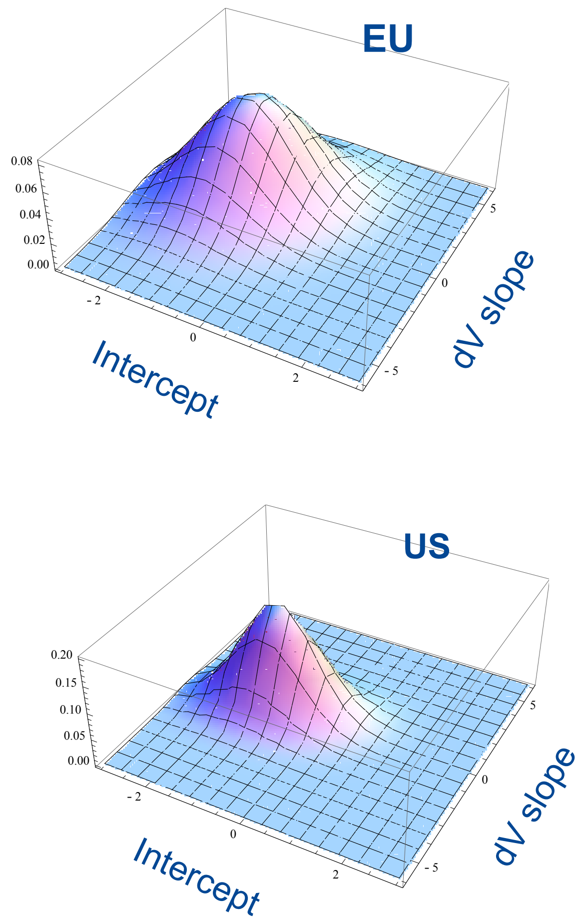


Figure 7. Example likelihood space for EU and US vehicles (not based on actual data)

Approach 1: Seemingly Unrelated Regression (Test hypothesis that EU and US injury risk models are identical)

In this context, using the crashworthiness problem to illustrate the example, a model can be any function of predictors that results in a predicted injury risk. Predictors are fixed qualities of each occupant's crash experience, such as age, gender, Delta-V, crash direction, model year, and restraint use. The real-life outcome is the presence or absence of injury. The model output is the predicted risk of injury. For this approach, the modeling process results in estimated coefficients of predictors that go into an equation that outputs injury risk.

After the model-selection process, we end up with two equations:

EU predicted risk = f(EU Coefficients * predictors)

US predicted risk = f(US Coefficients * predictors)

In addition, each equation has an *r X r* variance-covariance matrix that quantifies the uncertainty (variance) in each coefficient, as well as the covariance between coefficients. Note that the same predictors are used in both models. For example, if gender is a significant predictor in one dataset but not the other, we will still choose to include gender as a predictor in both models.

Seemingly Unrelated Regression (SUR) is a technique to test the hypothesis that the EU coefficients are equal to the US coefficients, including the intercept. The SUR method uses a standard hypothesis-testing framework in which "equal coefficients" is treated as the null hypothesis. With this approach, we must consider the potential for making two types of error: 1) claiming "different" when "same" is true, and 2) claiming "same" when "different" is true. As with all decision algorithms, increased confidence that one type of error is low results in lower confidence that the other type of error is low.

As described mathematically in Appendix E, the probability of a Type II error, or β, (1-β is power) is a function of the size of difference being detected and the probability of a Type I error, or α. The test statistic, *S,* and its variance are determined by the modeling process. As output of Approach 1, we can describe the relationship between α, β, and δ, using a Receiver Operating Characteristic (ROC) curve. A theoretical ROC curve is shown in Figure 8. When we refer to "large" and "small" differences, we might consider that a difference in risk ratio of less than 1% is a small difference, and a difference in risk ratio greater than 5% is a large difference. (The actual curves for these values may look different from those in Figure 8, although the curve for 1% will be below the curve for 5% for all values of α. The topic of quantifying "large" and "small" differences will be covered more fully in a later section.)

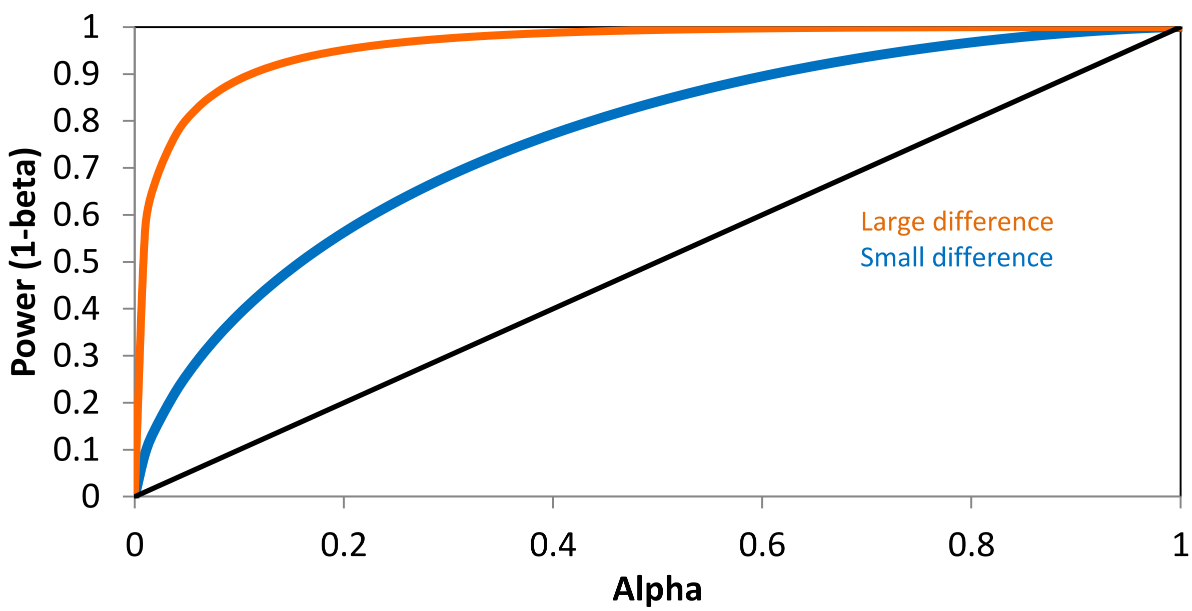


Figure 8. Theoretical ROC curve relating the relationship between power and alpha for two possible levels of difference between coefficients.

Figure 9 shows how Type I and Type II errors trade off. The figure also shows how a large difference, if it were to exist, can be detected with more certainty than a small difference.

To use this curve in the present context, we first need to select values for delta, alpha, and power; in choosing two items, the ROC curve defines the third. The choice of δ depends on the size of difference that is "close enough." Once δ is determined, we must select a cutoff for either α or the power, which depends on balancing between error types. In traditional hypothesis testing, α is selected to be 0.05 and power is often not measured. There, the goal is only to reject or fail to reject the hypothesis of "same." However, in this context, it would be reasonable to select a target value of power (e.g., 80%) and then determine α from the provided function. Figure 9 illustrates this approach.

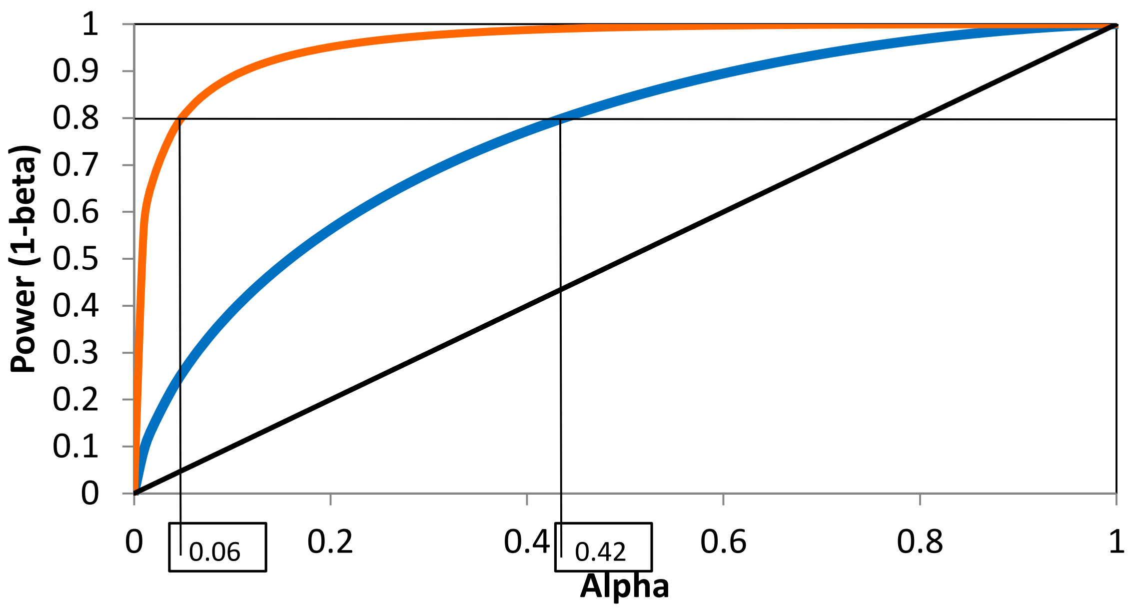


Figure 9. Theoretical ROC curve with examples of selection of criterion.

In Figure 9, setting power at 0.8 would result in an α of 0.06 in the case of the large difference and 0.42 in the case of the small difference. Finally, a p-value is compared to the chosen α level. The p-value represents the probability of getting *S* under the null hypothesis of "coefficients are the same". Note that *S* is independent of α, power, and δ. If the p-value is smaller than α, we conclude that the two risk models are different. If it is larger than α, we consider this to be evidence that they are the same (and the probability of being wrong is β).

As an example of how to interpret these curves, suppose that power is selected to be 80% and the p-value for the data is 0.29. If we have an 80% chance of detecting a small difference, then we will have a 42% chance of concluding that the coefficients are different when they are not. In this case, we would reject the null because the p-value is less than α, and we have a 42% chance of being wrong. However, if we have an 80% chance of detecting a large difference, then we have a 6% chance of incorrectly concluding "different." In this case, we conclude that the risk models are the same because the p-value is greater than α. This time, we have a 20% (1-80%) chance of being wrong.

To effectively use this approach, we will need to choose a size of difference that is important to detect, based on resulting likely injury consequences. Power is set as the requirement. In other words, we need to be able to conclude "different" if the difference really existed with some minimum probability (e.g., 80%). Alpha is then determined from the ROC curve based on the selected power level and the size of the difference. With larger variance and/or smaller sample size (i.e., weak evidence), we will need higher alpha values to maintain fixed power. The area under the ROC measures the strength of the evidence for detecting a particular difference in this context.

Appendix E contains more detailed equations relevant to Approach 1.

Approach 2: Consequences of Best Separate Models

For the second approach, we will develop separate models for EU and US injury risk and look at their consequences on a standard population. In this case, some predictors used in the EU model may not be used in the US model and vice versa. Since EU and US vehicles conform to different regulations, they would be expected to produce different models of injury risk. The underlying models take the following form:

Distribution of EU predicted risk = f(EU Coefficients * predictors) + EU uncertainty

Distribution of US predicted risk = f(US Coefficients * predictors) + US uncertainty

One might be tempted to conclude that separate models generated independently must result in different risk, but this is not the case. Different models can produce equivalent risk across the whole population. For example, one region may have higher risk in side impacts while the other region has greater risk in frontal impacts, but because of the differences in distribution by crash type or age, the overall injury risk to each driving population could still be the same. In fact, even in the simplest case with only one predictor it is possible to have different models resulting in the same point estimate of risk across the whole population. This case will be illustrated first in this section before turning to the more complex case including several predictors where the predicted injury risk is not one value but rather a probability distribution on several possible values.

Suppose that we consider models defined in terms of an intercept plus Delta-V and fix two (β_0, β_{dV}) pairs as maximum likelihood estimators based on the observed data in each of the regions US and EU. This approach leads to two predicted risk functions of the form in Equation 7, namely $\hat{p}_{US}$ characterized by the coefficients $a_{US} = \beta_0$ and $b_{US} = \beta_{dV}$, and $\hat{p}_{EU}$ characterized by the coefficients $a_{EU} = \beta_0$ and $b_{EU} = \beta_{dV}$. Furthermore, we assume that these are different models; i.e., $a_{US} \neq a_{EU}$ and/or $b_{US} \neq b_{EU}$. Our goal here is to compute the effect of these predicted risk functions across the whole population. These models are components (a) and (b) of the seven model elements described above.

Next, we identify a standard crash-involved occupant population for each region that describes the characteristics of crash-involved occupants and the conditions they are exposed to. The standard population consists of a set of occupants in crashes, together with data describing the crash, vehicle, and occupant. The EU and US standard populations are separate and carried through all analyses in parallel. These correspond to components (c) and (d) in the list of model elements.

For the underlying simple model with only one predictor, let us consider standard populations in which N_i^{EU} and N_i^{US} occupants are involved in crashes with Delta-V level *i* in the EU and the US, respectively;

the total number of crash-involved occupants are $N^{EU} = \sum_i N_i^{EU}$ and $N^{US} = \sum_i N_i^{US}$. The *overall predicted risk of a vehicle with risk function $\hat{p}$ on the EU standard population* can be computed using Equation 10.

$$R^{EU}(\hat{p}) = \frac{1}{N^{EU}} \sum_i \hat{p}(i) * N_i^{EU} \quad (10)$$

and the *overall predicted risk of a vehicle with risk function $\hat{p}$ on the US standard population, $R^{US}(\hat{p})$*, is defined analogously, with N^{EU} and N_i^{EU} replaced by N^{US} and N_i^{US}, respectively. The *risk ratio* computed on one of the standard populations is the ratio of overall predicted risks in a fixed order; for this study, we will always have the risk for EU vehicles is in the numerator and the risk for US vehicles in the denominator. This way, the risk ratio computed on the EU standard population is

$$RR_{EU} = R^{EU}(\hat{p}_{EU})/R^{EU}(\hat{p}_{US});$$

Analogously, the risk ratio computed on the US standard population is

$$RR_{US} = R^{US}(\hat{p}_{EU})/R^{US}(\hat{p}_{US}).$$

With these definitions, RR_{EU} is a measure of the *crashworthiness of EU vehicles relative to US vehicles in Europe* while RR_{US} measures the *crashworthiness of EU vehicles relative to US vehicles in the United States*. For example, outcomes of $RR_{EU} < 1$ and $RR_{US} > 1$ would mean that the overall predicted risk of injury given a crash in Europe is lower in EU-certified vehicles than in US-certified vehicles, but in the United States, it is the other way around. Roughly speaking, this means that EU vehicles are more crashworthy in Europe than US vehicles would be, but US vehicles perform better in the United States than EU vehicles would. Note that the EU and US models being the same (i.e., $a_{US} = a_{EU}$ and $b_{US} = b_{EU}$) is a sufficient but not necessary condition for $RR_{EU} = RR_{US} = 1$.

For the analysis, we probably need to be able to identify risk ratios relating to a 5% difference or less to be able to detect changes resulting in fatality fluctuations on the order of 100 occupants (see last section under Statistical Modeling). The variation in outcomes for a given risk ratio is an inherent property of risk, not a reflection of modeling uncertainty. Thus for this phase of analysis, we will leave that uncertainty out of the comparison process and focus on risk.

The process described above provides point estimates of the risk ratio based on one predictor, Delta-V. A completely analogous process could be applied in case of several predictors; in that case, point estimates can be obtained based on the occupants' characteristics in the standard population (crash type, severity, restraint use, age, etc).

However, this method has not accounted for model uncertainty so far. Only the best fit predicted risk functions were taken into account; considering model uncertainty is especially relevant for cases when the likelihood space is less peaked (e.g. see the likelihood space for EU vehicles in Figure 7) and no choice of the parameters is substantially better than others. Therefore, a more appropriate approach to the problem is to consider a whole distribution of possible parameter values instead of a point estimate.

The next step is to estimate a *distribution* of risk ratios produced by the US and EU models for the standard populations via simulation[14]. This is done in several steps. First, we randomly sample an injury risk from the distribution of such risks for each occupant, using the EU injury risk model. Next, we repeat the process by selecting an injury risk from the distribution of injury risks using the US model. We then calculate the risk ratio across each of the two standard populations, producing two risk ratios. This process is repeated 1000 or more times to produce an estimate of the distribution of risk ratios for the EU standard population (RR_{EU}) and another distribution of risk ratios for the US standard population (RR_{US}). A detailed mathematical description of this process is found in Appendix E.

Figure 10 through Figure 12 show three possible theoretical outcomes of this process. In these figures, the distribution of RR_{US} is displayed, which is a measure of the crashworthiness of EU vehicles relative to US vehicles when driven in the United States. In Figure 10 the most likely EU/US risk ratio is 1.0. The shaded area indicates that there is a 71% likelihood that the risk ratio lies between 0.9 and 1.1, i.e., the distribution is narrow. This case indicates strong evidence for a claim that the crashworthiness of an EU-certified vehicle in the US would be similar to the crashworthiness of US-certified vehicles. Figure 11 shows a similar shape of distribution that is shifted to the right. This means that the best estimate of the EU/US risk ratio is 1.15, and there is a 32% likelihood that the ratio lies between 0.9 and 1.1 (shaded area). The narrow distribution indicates strong evidence for a risk ratio higher than 1; intuitively, this means that EU-certified vehicles are less crashworthy in the US traffic than US-certified vehicles. In Figure 12, the best single estimate of risk ratio is 1.05, and there is a 29% likelihood that the ratio lies between 0.9 and 1.1. However, the broad distribution of risk ratios indicates that the data are inconclusive regarding this value.

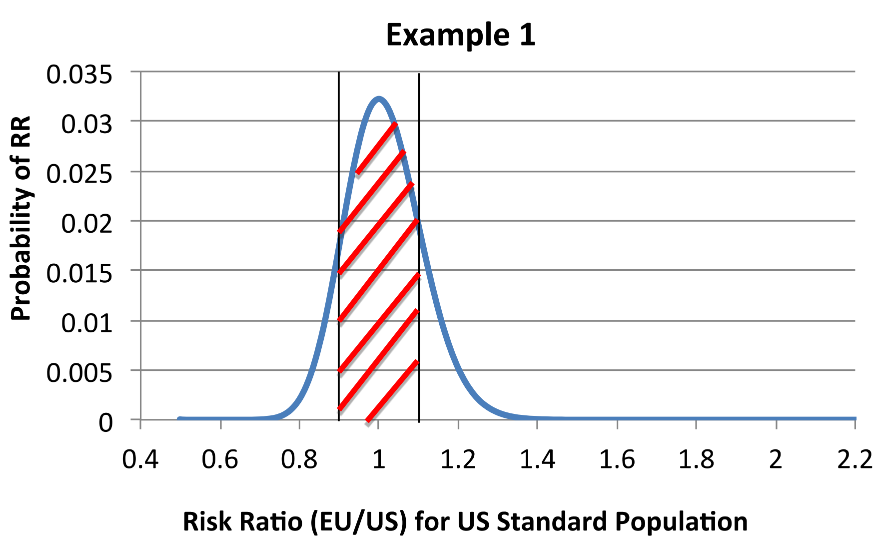


Figure 10. Distribution of EU/US risk ratio for US standard population, example 1.

[14] The simulation process described here is an instance of Monte Carlo simulations (Mooney, 1997).

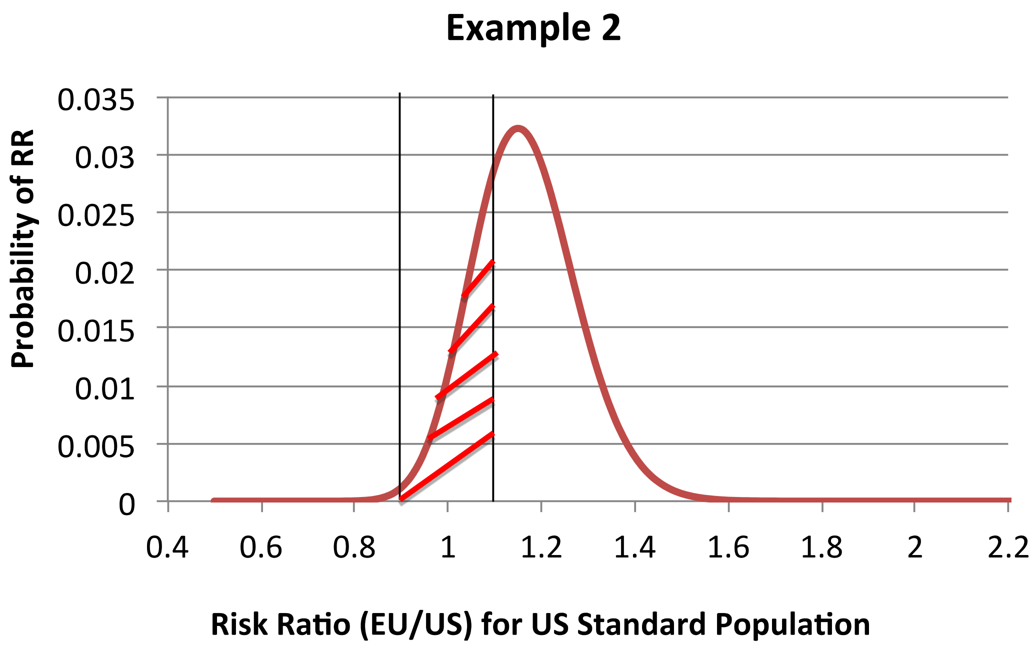


Figure 11. Distribution of EU/US risk ratio for US standard population, example 2.

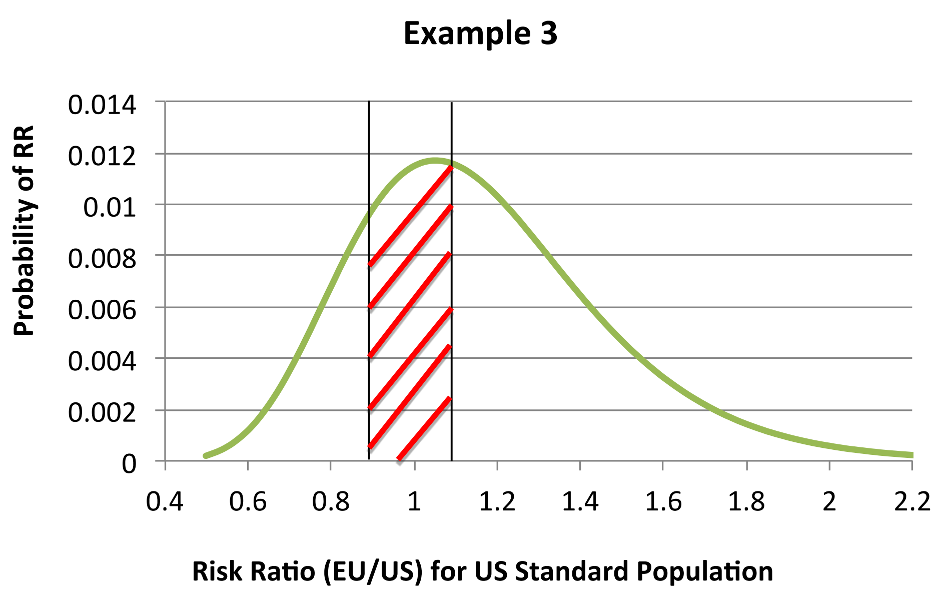


Figure 12. Distribution of EU/US risk ratio for US standard population, example 3.

This procedure would then be repeated for the EU population to produce parallel results. We will then have an estimated distribution of risk ratios for the standard US population (RR_{US}) and for the standard EU population (RR_{EU}). It is possible that the direction of the risk ratios is opposite. This case is similar to the outcome of $RR_{EU} < 1$ and $RR_{US} > 1$ for the point estimates that was discussed above and suggests that regulations have been optimized to the exposure in each of the regions.

The distribution of risk ratios is likely to include 1, corresponding to equal risk. The smaller the sample, the more likely it is to include 1. That is, weaker evidence leads to greater likelihood of failing to reject the null. As a result, we will not use a 0.05 cutoff, but will evaluate the strength of evidence, which is embodied in the spread of the risk-ratio distribution.

Approach 3: Evidence for Consequences

In Approach 2, we calculated the best model and evaluated likely consequences. In Approach 3, we do the opposite. We group models in terms of consequences, and then examine the relative evidence for each group.

In Approach 3, we will consider a large collection of possible hypotheses about the true risk models for EU- and US-regulated vehicles. For each hypothesis, there is an EU model and a US model. The models can be constructed using the same or different predictors. Figure 13 illustrates this idea using the likelihood surfaces first introduced in Figure 6 and Figure 7. Recall that in this simple two-parameter example, each point in the likelihood space is a pair of coefficients. Here, there is a separate pair for EU-regulated vehicles and one for US-regulated vehicles.

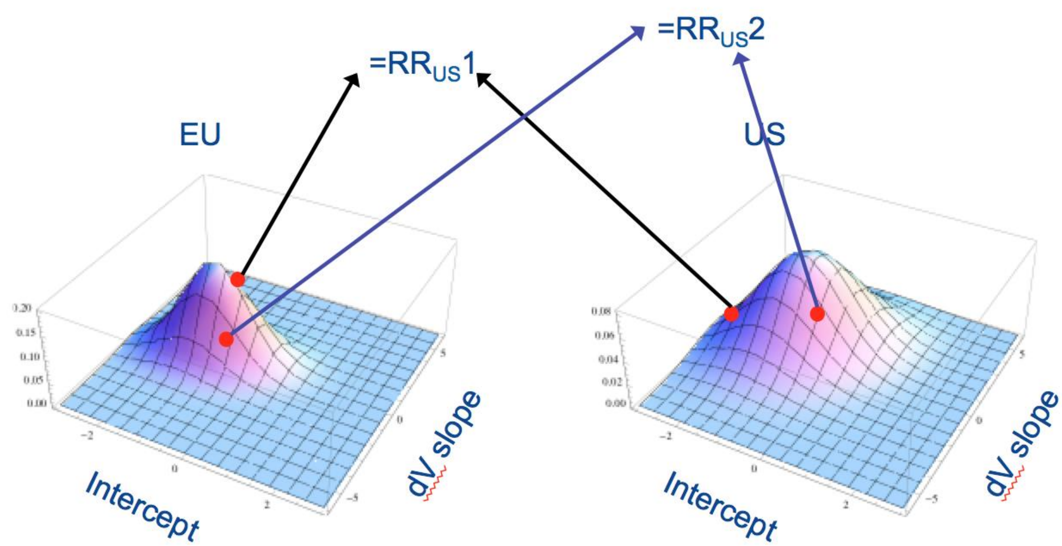


Figure 13. Hypothetical likelihood surfaces for EU and US vehicles. Risk ratios for the US standard population (RR_{US}) can be computed for each pair of models.

Each pair of models can be scored in terms of the resulting risk ratio in a single standard population. Figure 13 illustrates scores for two particular hypotheses, using the US standard population. These scores are labeled as $RR_{US}1$ and $RR_{US}2$. These pairs (as all pairs) will also have a score using the EU standard population, but here we will illustrate using only one standard population.

It is important to note that in this approach, each hypothesis is treated *as if it were true*. Thus, the consequences on the standard population are calculated without considering any uncertainty. We will score each potential hypothesis (i.e., pair of models) and then later measure the evidence for them.

Once all model combinations are scored, we group them into RR_{US} bins, such as .69-.71, .71-.73,…, .99-1.01, 1.01-1.03, 1.03-1.05….. For each risk-ratio bin representing a group of hypotheses, we use the data to measure the relative evidence for that group. The measure of evidence is the probability of getting the data we observed under the hypothesis (or model) being assessed; i.e., likelihood. The evidence for a bin is calculated as the weighted average of the evidence for each of the models in the bin.

Theoretical examples of the evidence are shown in Figure 14 and Figure 15. In Figure 14, the evidence peaks narrowly with the 0.98-1.02 bin, supporting a risk ratio in this range. The evidence shown In Figure 15 has a wide distribution that does not strongly favor any one risk ratio.

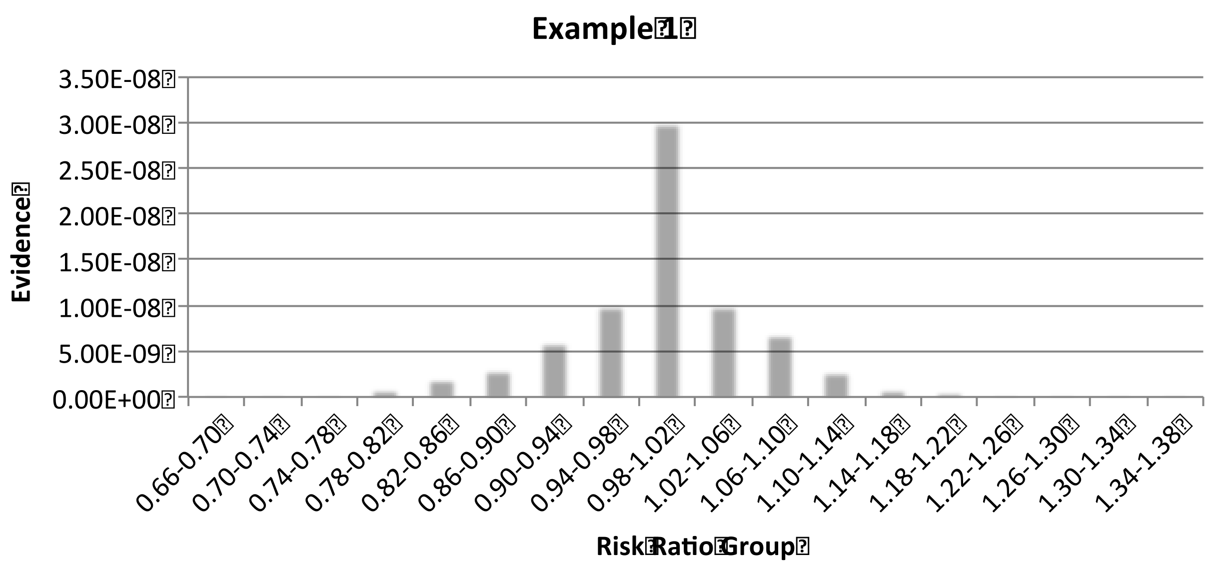


Figure 14. Evidence supports risk ratio in 0.98-1.02 range

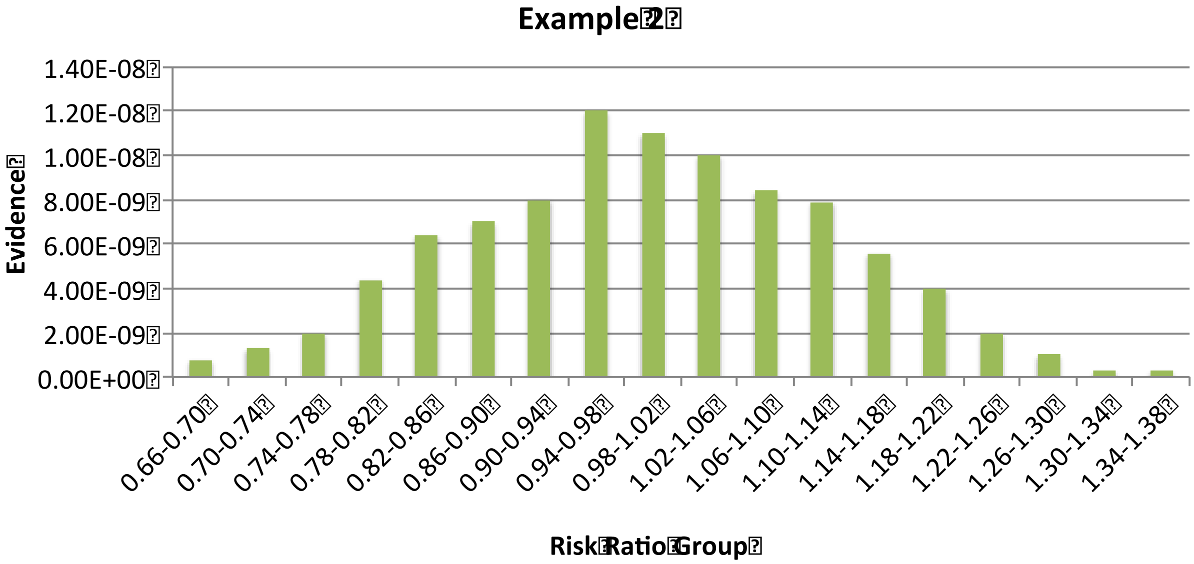


Figure 15. Evidence does not support any particular risk ratio.

Bayes Factors are ratios of the evidence, or average likelihood, for groups of hypotheses and can be interpreted as a direct comparison of evidence for one risk ratio vs. another. They allow us to compare the evidence provided by the data for one group/bin of hypotheses to the evidence provided for another group/bin of hypothesis.

A Bayes Factor calculation for Figure 14 is shown below. Bayes Factors ≥ 3 are considered positive evidence for the numerator hypothesis, relative to the denominator hypothesis.

$$BayesFactor1 = \frac{evidence\ for\ RR_{US} \in [0.98, 1.02]}{evidence\ for\ RR_{US} \in [0.94, 0.98)} = 3.0$$

As described above, groups of hypotheses (that is, risk ratios grouped in the same bin) can be considered together in the numerator and denominator. Generating the likelihood of the data within any group of hypothesis is done using a weighted average, where the likelihood of each hypothesis in the group is weighted by the prior probability of that hypothesis.

$$B = \frac{\sum_{\theta_1} p(D|\theta_1)\pi(\theta_1)}{\sum_{\theta_2} p(D|\theta_2)\pi(\theta_2)} \quad (12)$$

We will build the multi-dimensional likelihood space for the EU and US based on all predictors that showed some value in Approaches 1 and 2. The joint likelihood is the product of the separate likelihood values. The range of these parameter values will also be based on relevant results from the first two approaches. We will compute likelihoods for each group of hypothesis that falls within each RR_{US} bin. We then use the Bayes factors to compute ratios for each pairing of hypothesis bins to assess the level of differentiation of hypothesis, or to measure the ability of the data to support any particular bin of hypotheses. The process will be repeated for RR_{EU}.

One of the criticisms frequently leveled against Bayesian approaches is that the selection of the prior distribution, in this case weights for the weighted average of evidence, influences the results and is arbitrary. For this analysis, we need to choose a prior distribution to determine the weights in the weighted average of models in each RR bin. If we use equal weight across all hypotheses, the overall evidence for the group of hypotheses will be diluted and evidence for all bins will be similar. Some models are highly improbable, and their very low likelihood (and poor match to data) will reduce the average of all hypotheses in the group. Instead, "empirical Bayes" priors can be constructed that use the data to estimate the prior, hence minimizing influence of the prior on the outcome (Kass & Raftery, 1995). We will also perform sensitivity analysis to ensure that our results are not overly influenced by the selection of prior distribution.

Understanding Consequences

One of the issues that must be addressed in the risk comparison analysis of EU and US vehicles is what size of injury risk difference matters, or "how close to same is close enough?" To address this we need to have an idea of what the consequences of any given risk difference are. To illustrate this in Phase 1, we developed a simple model of fatality risk across the US fleet that was designed to give a general idea of consequences of various risk differences. However, in Phase 2, this model will need to be revisited, as described at the end of this section.

The estimation approach uses the US fleet and fatalities as the base comparison condition. Using GES 2011 data and limiting analysis to passenger cars, SUV, minivans and light trucks, we estimate 10,179,303 occupants involved in crashes. Using FARS (from Traffic Safety Facts on the NHTSA website), we observe that 21,253 occupants of those vehicle types were killed in 2011. This results in a base fatality risk of 0.20879% per crash-involved occupant.

We assume that in a (near) future year, the number of crash-involved occupants will be identical to 2011 and we want to compare the total fatalities under two scenarios:

1) Vehicles sold in the US must meet US standards and fatality risk for all occupants remains the same as in 2011
2) Vehicles sold in the US may meet either US or EU standards.

Scenario 1 is straightforward because the risk is determined. However, under scenario 2, we need to evaluate some different possibilities. We allow two numbers to change: fleet penetration of EU vehicles, and fatality risk ratio for the EU vehicles compared to US vehicles (which can be higher or lower). In Table 7, we set EU vehicle fleet penetration to be 10%. This value is a parameter of the model that could easily be adjusted according to different requirements. To put the currently chosen value of 10% into perspective, it can be compared to the annual vehicle turnover rate in the US, which in 2011 was about 5%.

Table 7.
Difference in US Annual Fatalities for Different Relative Risks between EU and US vehicles.

Relative Risk (EU Compared to US)	Difference in US Annual Fatalities are 50% Likely to Be Less Than…	Likelihood that US Annual Fatalities Will Be Reduced
1% Greater	21	44.4%
1% Less	-21	55.5%
5% Greater	106	24.4%
5% Less	-106	75.7%
Same Risk	0	50.0%

For the simulation, we assumed that risk for EU vehicles was either 1% or 5% higher or lower than the fatality risk for US vehicles, as illustrated in Figure 16.

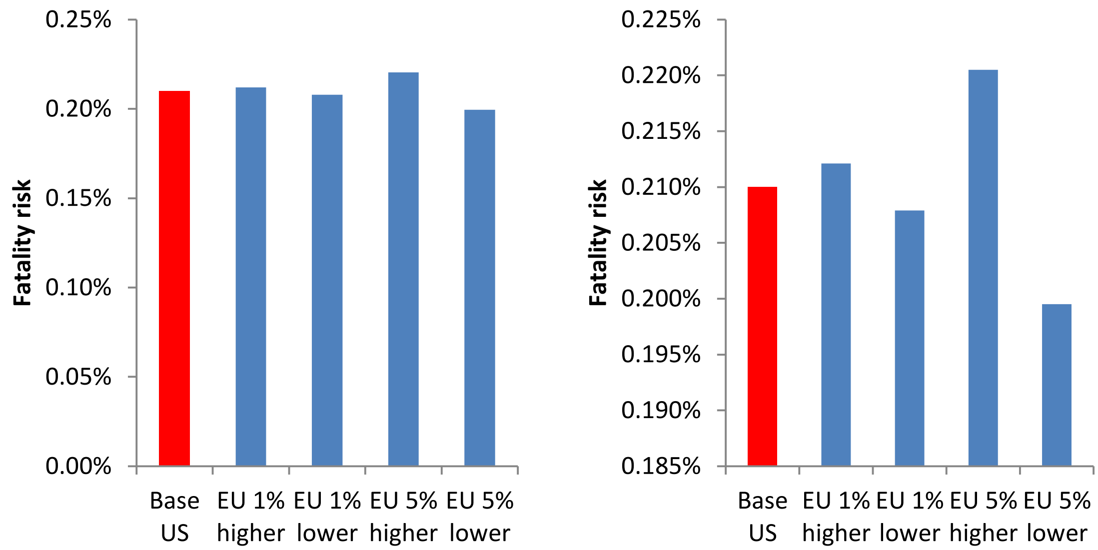


Figure 16. Estimated fatality risk if EU risk is 1% or 5% higher or lower than US (right plot includes close-up of variation).

For a fleet penetration of 10% EU vehicles, the second column of Table 7 is computed by multiplying 90% times the original risk plus 10% times the EU risk (which is some percentage increase or decrease relative to the US risk). Graphically, this is shown in Figure 17 (note the nonzero origin of y-axis). The red portion of the fatality total is from the vehicles meeting US standards, while the blue portion is from vehicles meeting EU standards. Overall, the estimated variation in total fatalities is small relative to the total number of fatalities, but the absolute differences are on the order of those used by NHTSA to justify regulation (NHTSA 2005).

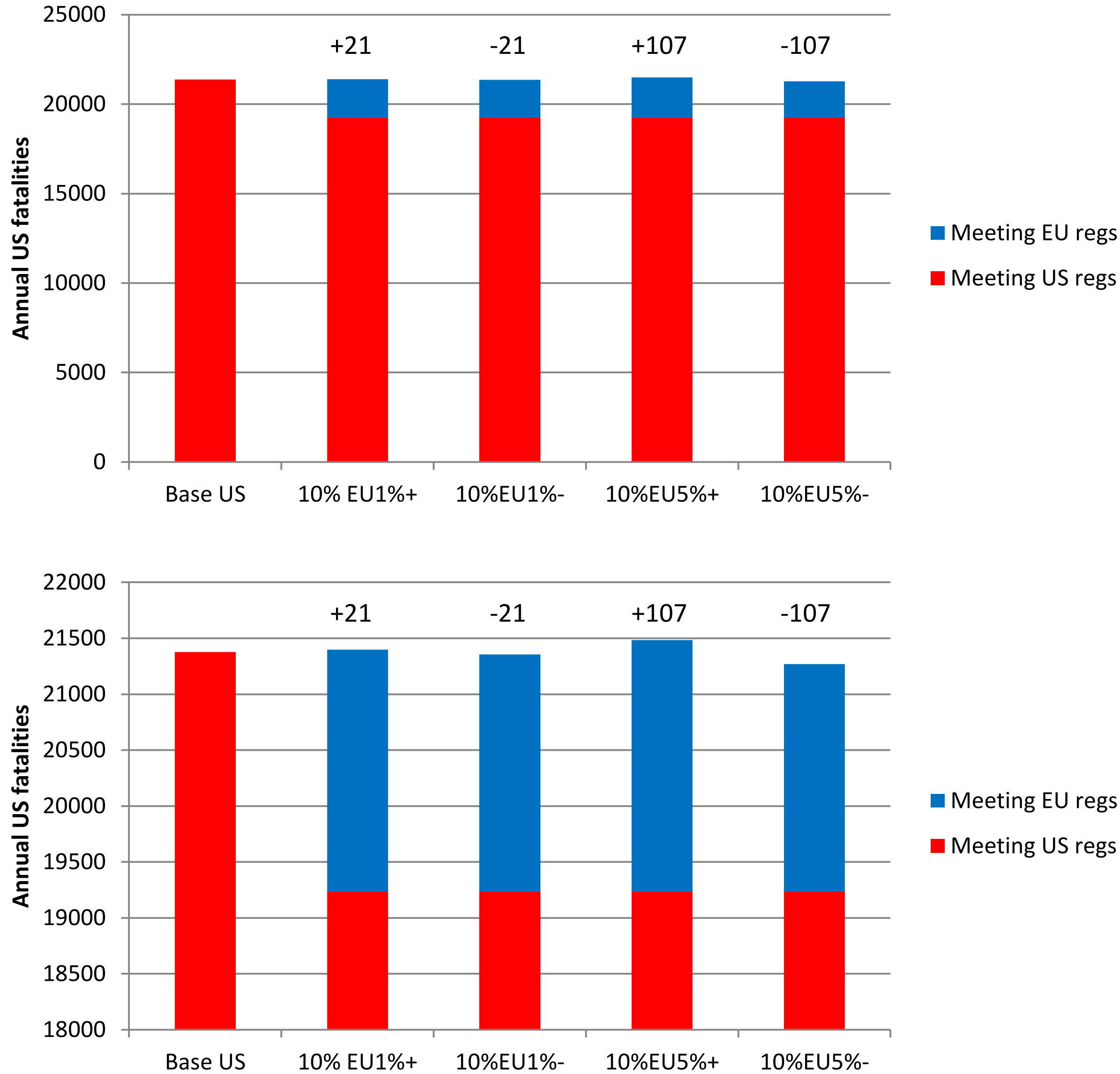


Figure 17.
Possible consequences to US fatality count for 10% EU vehicle fleet penetration with different levels of relative injury risk
(note non-zero y-axis origin on the lower graph).

Although these simulations convey the magnitude of change on average, they do not convey the variability in total fatalities that could occur even if risk was demonstrated to be the same. Within Table 7, variability in potential outcome is conveyed by the third column.

To compute the range of possibilities and their associated probabilities, we use a binomial distribution. The binomial has one parameter, *p*, which is the risk of fatality, and it returns the probability of getting a particular number of fatalities, given the risk, *p,* and total crashes, *n.*

The numbers in Table 7 and Figure 17 come from the *difference* between the two scenarios (i.e., the one with only US vehicles and the one with 10% EU vehicles and 90% US vehicles). For every difference, we can compute the probability of that happening, given the original US risk (from data) and the risk ratio between EU and US vehicles (up or down). The second column in Table 7 gives the expected fatality difference, which is the 50th percentile of the options.

The third column turns the question around and asks how likely it is that we would see lower fatalities in the EU-allowed scenario. This can happen even if EU risk is higher because the outcome is a random variable. If risk is higher, the *average* outcome is guaranteed to be an increase in fatalities, but the *actual* outcome can be many things. With a small risk increase and low fleet penetration, it is still possible there would be fewer fatalities when mixing in some EU vehicles (though never more than 50% likely). If the risk difference is much larger, there would be a lower chance of seeing fewer fatalities.

Another way of showing this concept is to present the 90% confidence interval on the fatality difference, as shown in Table 8. In all of these cases, that interval would include zero because of the low fleet penetration, the relatively small risk difference, and the selected confidence level.

Table 8.
90% confidence interval for expected fatality count change
given 10% fleet penetration of EU vehicles with different fatality risks

Relative Risk (EU Compared to US)	Difference in US Annual Fatalities Are 90% Likely to Be Within…
1% Greater	-230 – 273
1% Less	-272 – 230
5% Greater	-146 – 358
5% Less	-356 – 144
Same Risk	-251 – 251

One of the complexities of this calculation that is not reflected above is the fact that if EU-regulated vehicles are sold in the US, it is unlikely that such vehicles will enter the fleet at random. Instead, current European vehicles tend to be smaller than American vehicles (on average), and it is plausible that EU-regulated vehicles will initially replace smaller American vehicles. Thus, even if EU-regulated fleet penetration reaches 10%, it may not be a random 10%. This, too, will affect estimated consequences.

As part of the overall methodology, we listed components (e) and (f), which are economic models of how US vehicles might enter the EU fleet and EU vehicles might enter the US fleet. These models are not estimated from the data, but are decided upon *a priori*. A variety of plausible models can also be implemented in a sensitivity analysis.

The role of the fleet penetration models is to refine interpretation of risk consequences. The fleet penetration models will only be implemented in Approach 2. It is possible to use them in Approach 3, but because of the computational intensity of that approach, implementation is likely to take an inordinate amount of time. Instead, the primary value of looking at the implications of different fleet-penetration models can be gained by focusing on Approach 2 models.

Fleet-penetration models are added to the simulation of consequences of EU and US risk models in each of the standard populations. In Approach 2, we focused on RR_{EU} and RR_{US}. However, to understand consequences, we would take the next step and simulate outcomes. For each case in a standard population, we would incorporate not only the predicted risk and its model uncertainty, but also the possible outcomes. These outcomes would be weighted based on the fleet penetration model.

PHASE 2: RECOMMENDATIONS AND LIMITATIONS

Phase 1 focused on defining a methodology and potential datasets for analysis in Phase 2. Given that we have not yet applied any of the methods to the suggested datasets, it is difficult to predict how definitive the results will be. In other words, we do not know how much unexplained error will remain. There is a distinct possibility that the results, particularly from Approaches 2 and 3, will be inconclusive due to large variance.

However, the risk models developed in Approach 2 represent point estimates, which are our best guesses as to the true underlying risk curves. This will allow us to look at collections of crash types for which different injury risks may balance out. In addition, performing the basic comparisons has value, as this type and level of analysis has not been previously done. While the results may not directly be helpful in the TTIP negotiations, performing the analysis may provide guidance as to what might be possible with the available data for future analyses. It could also point to regulatory areas where harmonization or item-specific recognition (as in BASA) is least likely to result in major risk consequences.

To complete Phase 2 within the targeted window, analysis needs to begin as soon as possible if the Alliance decides to fund it. In the meantime, UMTRI and SAFER will begin preliminary negotiations with other database owners and experts about collaborating on Phase 2, if funded.

REFERENCES

Bayly, M., Fildes, B., Regan, M., & Young, K. (2006). *Review of crash effectiveness of intelligent transport systems. Deliverable d4.1.1 - d6.2, EU Project No.027763 - TRACE.* TRACE.

Breslow, N. E. (1996). Statistics in epidemiology: the case-control study. *Journal of the American Statistical Association*, *91*(433), 14-28.

Cliff, W. and Montgomery, D., (1996). Validation of PC-Crash - A Momentum-Based Accident Reconstruction Program. SAE Technical Paper 960885, doi:10.4271/960885.

Cuthbert JR. (1994) An Extension of the Induced Exposure Method of Estimating Driver Risk. *J. Roy. Stat. Soc. Stat. Soc.,* Vol. 157, pp. 177–190.

Evgenikos, P., Papantoniou, P., Yannis, G., Stanzel, M., & Kohsiek, A. (2011). *Review of the existing evaluation procedures related to safety systems, Deliverable 5.3 of the EC FP7 project DaCoTA.* DaCoTA.

Greenland, S., & Thomas, D. C. (1982). On the need for the rare disease assumption in case-control studies. *American Journal of Epidemiology*, *116*(3), 547-553.

Hampton, C., Gabler, H. (2009). NASS/CDS Delta-V Estimates: The Influence of Enhancements to the WinSmash Crash Reconstruction Code. Ann Adv Automot Med. 2009 Oct;53:91-102.

Hampton, C., Gabler, H. (2010). Evaluation of the Accuracy of NASS/CDS Delta-V Estimates from the Enhanced WinSmash Algorithm. Ann Adv Automot Med. Jan 2010; 54: 241–252.

Hill, J., Aldah, M., Talbot, R., Giustiniani, G., Fagerlind, H., and Jänsch, M., (2012). *Final Report*, Deliverable 2.5 of the EC FP7 project DaCoTA

Ireus, J., Lindquist, M. (2014). Analysis of Delta-Velocity and PDOF by Means of Collision Partner and Structural Involvement in Real-Life Crash Pulses With Modern Passenger Cars. Traffic Injury Prevention, 15:1, 56-65, DOI:10.1080/15389588.2013.793796

Johnson, N., Gabler, H. (2014). Evaluation of NASS-CDS Side Crash Delta-V Estimates using Event Data Recorders. Evaluation of NASS-CDS Side Crash Delta-V Estimates using Event Data Recorders, Traffic Injury Prevention, DOI: 10.1080/15389588.2014.881995

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773-795.

Keall, M. & Newstead, S. (2010). Selection of comparison crash types for quasi-induced exposure risk estimation. *Traffic Injury Prevention, 10(1),* pp. 23-29.

Kessler, C., Etemad, A., Alessandretti, G., Heinig, K., Selpi, Brouwer, R., et al. (2012). *Deliverable D11.3 Final report.* euroFOT.

Malta, M., Ljung-Aust, M., Faber, F., Metz, B., Saint-Pierre, G., Benmimun, M., et al. (2012). *Deliverable 6.4 Final results: Impacts on traffic safety.* euroFOT.

Mooney, C. Z. (Ed.). (1997). *Monte carlo simulation* (No. 116). Sage.

National Highway Traffic Safety Administration, Office of Regulatory Analysis and Evaluation, National Center for Statistical Analysis (2005). Preliminary Regulatory Impact Analysis: FMVSS No. 216, Roof Crush Resistance. http://www.nhtsa.gov/cars/rules/rulings/RoofCrushNotice/NPRM-216Final/#Iv

Niehoff, P., Gabler, H. (2006); The Accuracy of WinSmash Delta-V Estimates: The Influence of Vehicle Type, Stiffness, and Impact Zone. Annu Proc Assoc Adv Automot Med. 2006;50:73-89.

Nodine, E., Lam, A., Stevens, S., Razo, M., & and Najm, W. (2011). *Integrated Vehicle-Based Safety Systems (IVBSS) Light Vehicle Field Operational Test Independent Evaluation.* NHTSA.

Nolan, J., Preuss, C., Jones, S., O'Neill, B. (1998). An Update on The Relationships between Computed Delta-Vs and Impact Speeds For Offset Crash Tests. 16th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Windsor, Ontario, Canada, May 31- June 4, 1998.

Richards, D. and Cuerden, R. (2009). The relationship between speed and car driver injury severity. Department for Transport, Road Safety Web Report No. 9, 2009

Sayer, J., Bogard, S., Buonarosa, M., LeBlanc, D., Funkhouser, D., Bao, S., et al. (2011). *Integrated Vehicle-Based Safety Systems (IVBSS) Light Vehicle Field Operational Test Key Findings Report.* NHTSA.

UCLA: Statistical Consulting Group. Ordered Logistic Regression. From http://www.ats.ucla.edu/stat/stata/dae/ologit.htm (accessed April 25, 2014).

Wege, C., Will, S., & and Victor, T. (2013). Eye movement and brake reactions to real world brake-capacity forward collision warnings a naturalistic driving study. *Accident Analysis & Prevention , 58*, 259–270.

APPENDIX A. SUMMARY OF DATABASES

Table A1. Summary of Databases, Countries, Owners, Homepages

Database Name	Full Name	Country	Owner	Homepage
Country-specific in-depth databases in the EU				
ADAC	ADAC Accident Investigation Study	DE	ADAC	http://www.adac.de/infotestrat/unfall-schaeden-und-panne/Unfallforschung/default.aspx
Applus+ IDIADA	Applus+ IDIADA	ES	Applus+ IDIADA	http://www.applusidiada.com/en/service/Road_safety-1328274977266
CCIS	Co-operative Crash Injury Study	UK	Consortium (Department for Transport)	http://www.trl.co.uk/research_development/improving_safety/collision_investigation/incident_investigation.htm
CHICC	Child Safety in Car Crashes	SE	Consortium (Saab Automobile)	http://publications.lib.chalmers.se/records/fulltext/155673.pdf
CzIDAS	Czech In-Depth Accident Study	CZ	IDIADA CZ and CDV Transport Research Centre	http://www.kfv.at/fileadmin/webcontent/Publikationen_englisch/5th_CEE_2nd_Day/4_Fric_In-Depth_Accident_Investigation_CZ.pdf
DIANA	System Analysis on Real Accidents	ES	CIDAUT	http://www.cidaut.es/en/
EDA	Études Détaillées des Accidents	FR	LAB, IFSTTAR	http://www.innovations-transports.fr/Etudes-detaillees-d-accidents?lang=en
GIDAS	German In-Depth Accident Study	DE	Consortium (BASt and several manufacturers/suppliers)	http://www.gidas.org/?L=1
In-SAFE	In-depth Study of road Accidents in Florence	IT	Department of Mechanics and Industrial Technologies (DMTI) at the University of Florence and the Anaesthesia and Intensive Care Unit at the Emergency Department (ICU) of the Careggi University Hospital (Florence)	http://www.ptw.unifi.it/projects.php?id=14
INTACT	Investigation Network and Traffic Accident Collection Techniques	SE	Consortium (leader: Swedish Transport Administration)	http://www.intact-project.se
LAB	Laboratory of Accidentology and	FR	LAB - PSA Peugeot Citroën / Renault	http://www.psa-peugeot-citroen.com/en/psa_group/security_b3.php

Database Name	Full Name	Country	Owner	Homepage
	Biomechanics - Secondary Safety Database			
OTS	On The Spot	UK	Department for Transport (TSRC/TRL)	http://hdl.handle.net/2134/9170
RAIDS	Road Accident In-Depth Studies	UK	Department for Transport (TSRC/TRL)	http://www.trl.co.uk/research_development/improving_safety/collision_investigation/incident_investigation.htm
SIRSS	Sistema Integrato Regionale per la Sicurezza Stradale	IT	Osservatorio Regionale sulla Sicurezza Stradale	http://www.regione.toscana.it/-/sirss-sistema-integrato-regionale-per-la-sicurezza-stradale
Swedish national in-depth fatal crash database	Swedish national in-depth fatal crash database	SE	Swedish Transport Administration	http://www.trafikverket.se/Privat/Trafiksakerhet/Vart-trafiksakerhetsarbete/Sa-utreder-vi-olyckor/Djupstudier-av-vagtrafikolyckor/
VOIESUR	Vehicule Occupant Infrastructure Etudes de la Sécurité des Usagers de la Route	FR	Consortium (CEESAR, CETE NC, IFSTTAR, LAB)	http://www.agence-nationale-recherche.fr/en/anr-funded-project/?tx_lwmsuivibilan_pi2[CODE]=ANR-11-VPTT-0007
ZEDATU	ZEntrale DAtenbank Tödlicher Unfälle in Österreich	AT	Graz University of Technology	http://www.vsi.tugraz.at/index.php?id=47
Multinational in-depth databases in the EU				
CASPER	Child advanced safety project for European roads	DE, ES, FR, IT, UK	Consortium (PSA - Renault)	http://cordis.europa.eu/projects/rcn/91144_en.html
CHILD	CHild Injury Led Design	DE, ES, FR, IT, SE, UK	Consortium (Renault)	http://www.casper-project.eu/child%20web%20site/
CREST	Child Restraint System for Cars	DE, FR, IT, UK	Consortium (Renault)	http://cordis.europa.eu/projects/rcn/31115_en.html

Database Name	Full Name	Country	Owner	Homepage
DaCoTA	Road Safety Data, Collection, Transfer and Analysis	AT, CZ, DE, DK, EE, ES, FI, FR, GR, IS, IT, MT, NL, NO, PL, SE, SI, UK	Consortium (VSRC Loughborough)	http://www.dacota-project.eu
EACS	European Accident Causation Survey	DE, ES, FR, IT, NL, FI	European Automobile Manufacturers' Association (ACEA) + European Commission (EC)	http://ec.europa.eu/transport/wcm/road_safety/erso/data/Content/european_databases.htm
ECBOS	Enhanced Coach and Bus Occupant Safety	AT, DE, ES, FR, GB, IT, NL, SE	Consortium (Graz University)	http://ec.europa.eu/transport/road_safety/pdf/projects/ecbos.pdf
ETAC	European Truck Accident Causation	DE, ES, FR, HU, IT, NL, SI	Consortium (CEESAR) (initiated by the International Road Transport Union (IRU) & European Commission (EC))	http://ec.europa.eu/transport/roadsafety_library/publications/etac_final_report.pdf
MAIDS	Motorcycle Accidents In-Depth Study	DE, ES, FR, IT, NL	Consortium (European Association of Motorcycle Manufacturers, ACEM)	http://ec.europa.eu/transport/road_safety/pdf/projects/maids.pdf
PENDANT	Pan-European Co-ordinated Accident and Injury Databases	AT, DE, ES, FI, FR, NL, SE, UK	Consortium (VSRC Loughborough)	http://www.vsi.tugraz.at/pendant
RISER	Roadside Infrastructure for Safer European Roads	AT, ES, FI, FR, NL, SE, UK	Consortium (Chalmers University)	http://ec.europa.eu/transport/roadsafety_library/publications/riser_final_report.pdf
ROLLOVER	Improvement of Rollover Safety for Passenger Vehicles	AT, DE, ES, UK	Consortium (Graz University)	http://www.vsi.tugraz.at/rollover
SafetyNet Causation	SafetyNet Causation Database	DE, FI, IT, NL, SE, UK	Consortium (VSRC Loughborough)	http://erso.swov.nl/safetynet/content/safetynet.htm
SafetyNet Fatal	SafetyNet Fatal Database	DE, FI, FR, IT, NL, SE, UK	Consortium (VSRC Loughborough)	http://erso.swov.nl/safetynet/content/safetynet.htm

Database Name	Full Name	Country	Owner	Homepage
In-depth databases in the US				
NASS-CDS	National Automotive Sampling system - Crashworthiness Data System	US	NHTSA	http://www.nhtsa.gov/NASS
CIREN	Crash Injury Research Network	US	NHTSA	http://www.nhtsa.gov/Research/Crash+Injury+Research+(CIREN)/Data:
SCI in NiTS	Special Crash Investigations of Not in Traffic Surveillance	US	NHTSA	http://www.nhtsa.gov/SCI
EU level crash data				
CARE	Community Road Accident Database	All EU countries & CH, IS, NO, i.e.: AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MT, NI, NL, NO, PL, PT, RO, SE, SI, SK, UK	European Commission (EC)	http://ec.europa.eu/transport/road_safety/specialist/statistics/index_en.htm
US level crash data				
NASS-GES	National Automotive Sampling System - General Estimated System	US	NHTSA	http://www.nhtsa.gov/NASS
FARS	Fatal Accidents	US	NHTSA	http://www.nhtsa.gov/FARS

Database Name	Full Name	Country	Owner	Homepage
	Recording System			
Global databases				
iGlad	Initiative for the Global Harmonization of Accident Data	AT, AU, CZ, DE, ES, FR, IN, IT, SE, US	Consortium (Administrator: Chalmers University)	http://www.iglad.net
IRTAD	International Road Traffic and Accident Database	OECD countries: AT, AU, BE, CA, CH, CZ, DE, DK, ES, FI, FR, GR, HU, IE, IL, IS, IT, JP, KR, LU, NL, NO, NZ, PL, PT, SE, SI, UK, US	OECD	http://internationaltransportforum.org/irtadpublic/index.html

Table A2. Summary of Databases, Accessibility, Sizes, Years, Inclusion Criteria

Database name	Accessibility			Number of crashes	Data years	Inclusion criteria, other details
	Public	Consortium	Private			
Country-specific in-depth databases in the EU						
ADAC			x	11456	2005-present	Air rescue was called to crash scene; therefore, 90% of the collected crashes involve a severely/critically injured occupant.
Applus+ IDIADA			x	400	1999-present	Varying (as data collection has been conducted in projects with different focus). Generally, crashes with the involvement of cars, motorcycles, coaches or vulnerable road users are collected. PC-Crash reconstruction.
CCIS		x		15000	1983-2010	Passenger car <7 years old and towed was involved in the crash, at least one occupant of the car was injured.
CHICC		x		87	2004-2007	Crashes including a child of age <=12 admitted to a hospital in the western region of Sweden are collected; no Delta-V values are available.
CzIDAS		x		300	2011-present	At least one person was injured in the crash; all traffic modes are included; PC-Crash reconstruction.
DIANA			x	400	2003-present	Injury crashes; random notifications from police control rooms regardless of the type of crash or the type/age of vehicle; crashes are collected in Valladolid province, both in urban and in rural areas. PC-Crash reconstruction.
EDA		x		1100	?-2003	Car, pedestrian and truck crashes; in-depth on-the-spot investigation; PC-Crash and Madymo reconstruction.
GIDAS		x		22000	1999-present	At least one person injured; all traffic modes; PC-Crash reconstruction.
In-SAFE			x	150	2011-present	Severe and fatal accidents (accidents with major trauma, ISS>15) in metropolitan areas of Florence; the deformation energy and Delta-V are estimated with Crash3; all the previous data are verified and validated by PC-Crash 8.3 and Virtual CRASH 2.2.
INTACT		x		275	2007-2009 and 2012-present	Ambulance was called to the crash scene; at least one passenger car, bus or truck was involved in the crash; PC-Crash reconstruction.
LAB		x		1000+ (~400 / year)	1993-present	At least one person in a passenger car was injured.

Database name	Accessibility			Number of crashes	Data years	Inclusion criteria, other details
	Public	Consortium	Private			
OTS		x		4500	2000-2010	All road users, all injury severities (including property damage only). PC-Crash reconstruction.
RAIDS		x		300	2013-present	Police reported collisions in the Thames Valley and Hampshire regions; crashes occurring anywhere in Nottinghamshire or Leicestershire.
SIRSS			x	3887	2008-2012	Randomly selected crashes from the Region of Tuscany.
Swedish national in-depth fatal crash database	x (access can be granted to selected research projects)			3481	1997-present; digitalized from 2004	All fatal crashes in Sweden; no Delta-V values are available.
VOIESUR		x		9000	2011	All information coded in the database, including crash reconstruction, relies on police reports investigated by accident investigation experts. The sample includes: all fatal crashes in France (~3500), all crashes occurring in Rhone department (~2500), 1/20 of injury crashes in France (~3000).
ZEDATU		x		3000	2004-present	Fatal car accidents in Austria; PC-Crash reconstruction.
Multinational in-depth databases in the EU						
CASPER		x		137	2009-2012	Same as in CREST with two differences: children <=13 years are considered and rear impacts are also reviewed on a case-by-case basis.
CHILD		x		264	2002-2005	Same as in CREST.
CREST		x		405	1996-2000	Car-to-car or car-to-fixed obstacle crashes, with at least one child <=12 yrs restrained in CRS or with adult seat belt involved in the crash. Vehicles with <=9 occupants are considered. At least one occupant suffered an MAIS2+F injury in the crash. Only frontal impacts with DeltaV>=40 km/h and lateral impacts with >=200 mm intrusion are considered.
DaCoTA		x		99	2012	At least 3 out of 5 cases from each country must include a road user who was taken to hospital immediately after the crash.
EACS		x		1904	1996-2002	A light vehicle with weight <3500 kg was involved in the crash, at least one person was injured. The cases from each region must be closely related to the distribution of

Database name	Accessibility			Number of crashes	Data years	Inclusion criteria, other details
	Public	Consortium	Private			
						crashes in the country in terms of road type, vehicle types and drivers. Cases where the drivers sustained serious brain injuries were excluded.
ECBOS		x		36	1994-2001	Crashes with the involvement of M2, M3 and city buses.
ETAC		x		624	2004-2006	All injury accidents involving at least one heavy goods vehicle.
MAIDS		x		921	1999-2001	Crashes with powered two-wheelers in which at least one rider was injured.
PENDANT		x		1086	2003-2005	At least one car manufactured after 1998 was involved in the crash. At least one vehicle occupant was injured. Among other crashes, the database includes 119 reconstructed rollovers (with Delta-V values).
RISER		x		211	1998-2002	Single vehicle crashes only; all passenger vehicles, trucks and motorcycles, and all severity levels are included.
ROLLOVER		x		145	2002-2005	Rollover crashes, up to one turn, direction longitudinal, belted occupant(s); the sample includes 40 crashes that started with roll followed by impact; 40 crashes that started with side impact followed by roll; 40 cases that started with front impact followed by roll. In each category there are 1-2 cases with SUVs, but convertibles are excluded.
SafetyNet Causation		x		1006	2005-2008	Random selection of cases to the best of the teams abilities, all kinds of crashes were investigated. No Delta-V values are available.
SafetyNet Fatal		x		1296	2003-2004	Fatal crashes investigated retrospectively; the sample should be representative to country level with respect to the following measures (separately): road user killed, road class, urban-rural, month of the year. No Delta-V values are available.
In-depth databases in the US						
NASS-CDS	x			~3300-5000 per year (fewer in recent years)	1988-present (annually updated in fall of following year)	Probability sample of crashes in the US involving at least one light vehicle that is towed away due to disabling damage.

Database name	Accessibility			Number of crashes	Data years	Inclusion criteria, other details
	Public	Consortium	Private			
CIREN	x (partial)		x (UMTRI)	~400 per year	1997-present (2004-2013 avail in public files)	Convenience sample from participating CIREN hospitals; cases are crash-involved occupants admitted to participating trauma center who meet inclusion criteria: vehicle<=6 yrs, AIS 3+ injury, frontal and restrained (front row airbag only allowed) or any side impact or any rollover without 100% ejection or any fire or any pregnant occupant regardless of all other criteria).
SCI in NiTS	x			113	2006-2013	Vehicle related injury incidents that either occur on private property or do not involve a moving vehicle. The special crash investigation subset of the NiTS has done in-depth investigation of 113 cases from the complete database that includes approximately 600,000 documented injuries per year.
EU level crash data						
CARE	x (access by selected organizations only)			1000000+	1991-present	Injury crashes that are included in the national statistics of the countries involved.
US level crash data						
NASS-GES	x			~50,000 per year	1988-present (annually updated in fall of following year)	Probability sample of police-reported crashes in the US; all data elements are coded from the police reports.
FARS	x			~33,000 per year (has been declining)	1975-present (annually updated in summer of following year)	Census of fatalities on US public roads; death within 30 days of crash and due to injuries sustained in the crash; some additional information from death certificates and coroner's reports are included, but in-depth crash investigation is not done.
Global databases						
iGlad	x (access can be purchased)			1580	2007-present	Varies across the data providers. A general requirement is that more than 80% of the variables have a known value for each crash included in the database.
IRTAD	x (access can be purchased)			1000000+	1970-present	Aggregated crash data and exposure data from the OECD countries.

APPENDIX B: KEY VARIABLES

Table B1. Summary of key variables in datasets

	GIDAS		iGLAD	INTACT	NASS	
Type of variable	**Description**	**Name**	**Name**	**Name**	**Description**	**Name**
accident description	type of accident according to the catalogue of the HUK from 1977	UTYP	accident type	DaCoTA accident type (data years 2012-present); CADAS accident type (data years 2012-present); INTACT accident type (data years 2007-2009)	crash type	ACCTYPE
	moving direction of the involved vehicles	UART	collision type	-	crash configuration	RSHDSC2
surrounding environment	whether the scene of the accident is inside or outside a built-up area	ORTSL	location	area classification; accident environment	-	-
road characteristics	road level	STRART	road type	road type	-	RELATION TO INTERCHANGE, TRAFFICWAY FLOW , NUMBER OF TRAVEL LANES
	material of road surface	STRDECK	-	roadway surface type	roadway surface type	SURTYPE
	road surface condition due to weather	STROB	road condition	road condition	roadway surface condition	SURCOND
	maximum permitted speed of the road	VZUL	-	speed limit at accident time	posted speed limit	SPLIMIT
	information of road section	STFUHO	-	accident spot type	Traffic way-relation to junction	RELINTER

	GIDAS		iGLAD	INTACT	NASS	
Type of variable	**Description**	**Name**	**Name**	**Name**	**Description**	**Name**
	type of traffic control	VKREG	-	traffic regulation	traffic control device	TRAFCONT
vehicle characteristics	date of first registration (year)	TDEZJ	registration year	-	vehicle model year	MODELYR
	type of vehicle (own definition)	FART	participant type	element type	body category	-
	vehicle (make and model)	FZG2	vehicle make	make - car; model - car	vehicle make , vehicle model	MAKE, MODEL
	body shape	ABF	-	body style - car	class of vehicle	BODYTYPE
	curb weight	LGEW	-	kerb weight - car	vehicle curb weight	CURBWGT
	gross vehicle weight	ZULGEW	-	*not for cars, but possible to guess*	total GVWR (kgs)	GVWR
	energy source of the engine	EQUELLE	vehicle engine type	fuel type - car; alternative fuel type - car	fuel type	FUELTYP
vehicle crash characteristics	the principal direction of force according to CDC 1, 2	VDI1	primary collision - CDC	CDC1,2	clock	CLOCK
	main deformed vehicle area according to CDC 3	VDI2	primary collision - CDC	CDC3	deformation location	GAD
	specific horizontal location of the damage according to CDC 4	VDI3	primary collision - CDC	CDC4; CDC4e	long/lateral	SHL
	VDI vertical	VDI4	primary collision - CDC	CDC5	vertical/lateral	SVL

	GIDAS		iGLAD	INTACT	NASS	
Type of variable	**Description**	**Name**	**Name**	**Name**	**Description**	**Name**
	VDI type of impact	VDI5	primary collision - CDC	CDC6	distribution	TDD
	VDI degree of deformation	VDI6	primary collision - CDC	CDC7	extent	EXTENT
	EES	EES	primary collision - EES	EES	-	-
	delta-v	DV	primary collision - delta-v	Delta V (DV) total [km/h]; DV longitudinal [km/h]; DV lateral [km/h]	highest severity impact computer generated Delta V--total	HDVTOTAL
	vehicle weight in the crash	GEWGES	vehicle mass	-	vehicle cargo weight + vehicle curb weight	-
	backward dislocation of engine	MOTORDEF	-	powertrain hit - car	damage to fuel cell	FUELDAM
	impact number which caused worst damage	KOLLS	-	*can be computed from the variable "relative impact severity"*	highest severity impact--event number	ACCSEQDV
vehicle pre-crash characteristics	initial velocity	V0	primary collision - Driving speed	initial speed (IS)	-	-
	collision velocity / sequence end velocity	VK	primary collision - collision speed	collision speed [km/h]	highest severity impact computer generated Delta V—impact speed	HIMPCTSP
	skidding occurrence prior to impact	SCHLEU	-	event type	pre-impact stability	PREISTAB

	GIDAS		iGLAD	INTACT	NASS	
Type of variable	**Description**	**Name**	**Name**	**Name**	**Description**	**Name**
	equipment of traction control system	ASR	type of safety system	traction control system - car	-	-
	equipment of electronic stability program	ESP	type of safety system	electronic stability control - car	-	-
	equipment of cruise control	TMAT	type of safety system	cruise control - car	-	-
	equipment of collision warning	COLLWARN	type of safety system	forward collision warning - car; Rearward collision warning - car	equipment type-crash avoid	-
VRU crash characteristics	wrap around distance (VRU)	ANABWL	-	wrap around distance	-	-
	impact zone on the vehicle (VRU)	ANZONE	-	vehicle area description - car; x-measurement; y-measurement [mm]; z-measurement	-	-
human characteristics	gender	GESCHL	gender	gender	sex	SEX
	age in years	ALTER1	age	age	age	AGEMONTH, AGEYEAR
	height	GROESP	height	stature	height	HEIGHT
	body weight	GEWP	weight	weight	weight	WEIGHT
human behavior	result of digital alcohol test	TESTDIG	-	alcohol level measured by; Alcohol per millage	alcohol test result	ALCTEST
	blood alcohol content	BLUTALK1	-	alcohol level measured by; Alcohol per millage	alcohol test result	ALCTEST
	equipment of alcohol lock system	ALKOLOCK	type of safety system	alcolock - car	-	-

	GIDAS		iGLAD	INTACT	NASS	
Type of variable	**Description**	**Name**	**Name**	**Name**	**Description**	**Name**
	equipment of driver distraction control	ABLENKON	type of safety system	impairment warning - car	-	-
occupant movement	ejection occurrence	AUSSCHL	-	ejection route	ejection -type	EJECTION
	be ejected from where	AUSWO	-	seat position; seat	ejection—area	N/A
injury description	Maximum AIS	MAIS	MAIS	-	maximum known occupant AIS (AIS98 FORMAT)	MAIS
	Maximum AIS (by AIS2005)	MAIS05	MAIS	MAIS	maximum known occupant AIS (AIS08 FORMAT)	MAIS 08
	injury severity by the investigation team	PVERL	injury severity	police injury severity	PAR Severity	INJSEV
	type of hospital treatment	BEHAND	-	hospitalized; number of days in hospital	treatment	TREATMNT
	AIS region 1 head	AISREG1	AIS region 1: head w/o face	AIS-grade, AIS-localizer 1, AIS-localizer 2 for each injury	A.I.S. Severity (AIS08 Format) & Body Region (AIS08 Format)	AIS & REGION08
	AIS region 2 face	AISREG2	AIS region 2 face	AIS-grade, AIS-localizer 1, AIS-localizer 2 for each injury	A.I.S. Severity (AIS08 Format) & Body Region (AIS08 Format)	AIS & REGION09

	GIDAS		iGLAD	INTACT	NASS	
Type of variable	**Description**	**Name**	**Name**	**Name**	**Description**	**Name**
	AIS region 3 neck w/o spine	AISREG3	AIS region 3 neck w/o spine	AIS-grade, AIS-localizer 1, AIS-localizer 2 for each injury	A.I.S. Severity (AIS08 Format) & Body Region (AIS08 Format)	AIS & REGION10
	AIS region 4 thorax w/o shoulder	AISREG4	AIS region 4 thorax w/o shoulder	AIS-grade, AIS-localizer 1, AIS-localizer 2 for each injury	A.I.S. Severity (AIS08 Format) & Body Region (AIS08 Format)	AIS & REGION11
	AIS region 5 abdomen	AISREG5	AIS region 5 abdomen	AIS-grade, AIS-localizer 1, AIS-localizer 2 for each injury	A.I.S. Severity (AIS08 Format) & Body Region (AIS08 Format)	AIS & REGION12
	AIS region 6 spine	AISREG6	AIS region 6 spine	AIS-grade, AIS-localizer 1, AIS-localizer 2 for each injury	A.I.S. Severity (AIS08 Format) & Body Region (AIS08 Format)	AIS & REGION13
	AIS region 7 upper extremities	AISREG7	AIS region 7 upper extremities	AIS-grade, AIS-localizer 1, AIS-localizer 2 for each injury	A.I.S. Severity (AIS08 Format) & Body Region (AIS08 Format)	AIS & REGION14
	AIS region 8 lower extremities	AISREG8	AIS region 8 lower extremities	AIS-grade, AIS-localizer 1, AIS-localizer 2 for each injury	A.I.S. Severity (AIS08 Format) & Body Region (AIS08 Format)	AIS & REGION15
	AIS region 9 not specified injuries	AISREG9	AIS region 9 not specified injuries	AIS-grade, AIS-localizer 1, AIS-localizer 2 for each injury	A.I.S. Severity (AIS08 Format) & Body Region (AIS08 Format)	AIS & REGION16
	number of days until death	UELTG	-	number of days until death	time to death	DEATH
vision	Daytime or nighttime of the accident	TZEIT	light condition	light condition	conditions--light	LGTCOND

	GIDAS		iGLAD	INTACT	NASS	
Type of variable	**Description**	**Name**	**Name**	**Name**	**Description**	**Name**
	clouds and fog conditions at the time of the accident	WOLK	-	cloud cover; fog/mist	conditions--atmospheric	CLIMATE
	equipment of lane departure warning	SPURHAW	type of safety system	lane departure warning - car	equipment type -crash avoid	-
	equipment of lane keeping support	SPURHAE	type of safety system	-	equipment type -crash avoid	-
	equipment of backup warning aid	RHILF	type of safety system	back-up alarm - truck; back-up alarm - bus	-	-
	equipment of night vision	NIGHTV	type of safety system	night vision - car	-	-
lamps	type of lighting - daylight running lights	SWARTT	type of safety system	*no variable, but interview question*	equipment type -crash avoid	-
	bending light / adaptive front lighting system	KURVENL	type of safety system	active headlamps - car	-	-
	adaptive light distribution	ADAPTLV	type of safety system	active headlamps - car	-	-
headrest	type of headrest	KSTART	-	neck restraint	head restraint -type	HEADTYPE
	type of headrest protection system	KSTSCH	type of safety system	whiplash protection	head restraint - active	HEADACT

	GIDAS		iGLAD	INTACT	NASS	
Type of variable	**Description**	**Name**	**Name**	**Name**	**Description**	**Name**
seat belt	seat belt usage	RHSBEN	use of safety system	signs of seat belt usage; friction marks on belt webbing; claimed seatbelt usage; verified seatbelt usage	used in this crash?	ACTUSE
	type of seat belts	GURTE	type of safety system	Seat belt type	availability	ACTAVAIL
	whether the seat belt functioned during the accident	RHSFUNKT	-	belt malfunction; pretensioner activated	Malfunction (Seatbelt)	MANFAIL
airbag	activation of front airbag or not	AIRBF	type, Deployment / activation	airbag type	function – location& function— deployment (airbag)	BAGLOC & BAGDEPLY
	activation of roof-integrated airbag or not	AIRBDI	type, Deployment / activation	airbag type	function – location& function— deployment (airbag)	BAGLOC & BAGDEPLY
	activation of seat-integrated airbag or not	AIRBSI	type, Deployment / activation	airbag type	function – location& function— deployment (airbag)	BAGLOC & BAGDEPLY
	activation of door-integrated airbag or not	AIRBTI	type, Deployment / activation	airbag type	function – location& function— deployment (airbag)	BAGLOC & BAGDEPLY

	GIDAS		iGLAD	INTACT	NASS	
Type of variable	**Description**	**Name**	**Name**	**Name**	**Description**	**Name**
	activation of knee airbag or not	AIRBKN	type, Deployment / activation	airbag type	function – location& function— deployment (airbag)	BAGLOC & BAGDEPLY
	activation of seat ramp airbag or not	AIRBSR	type, Deployment / activation	airbag type	function – location& function— deployment (airbag)	BAGLOC & BAGDEPLY
	activation of rear airbag or not	AIRBH	type, Deployment / activation	airbag type	function – location& function— deployment (airbag)	BAGLOC & BAGDEPLY
child seat	type of child seat	KISIART1	type of safety system	CRS type as used in car; CRS ECE approval classification	type_child restraint	CHTYPE
	type of child seat	KISIART2	type of safety system	CRS type as used in car; CRS ECE approval classification	type of child safety seat & child seat used	CHTYPE
tire	tire pressure front right of the passenger car	REIFDRVR	-	tire pressure	tire measured pressure	PRES
	tire pressure rear right of the passenger car	REIFDRHR	-	tire pressure	tire measured pressure	PRES
	tire pressure rear left of the passenger car	REIFDRHL	-	tire pressure	tire measured pressure	PRES

	GIDAS		iGLAD	INTACT	NASS	
Type of variable	**Description**	**Name**	**Name**	**Name**	**Description**	**Name**
	tire pressure front left of the passenger car	REIFDRVL	-	tire pressure	tire measured pressure	PRES
brakes	equipment of brake assistant system	BREMSASS	type of safety system	ABS - car; brake assist - car	equipment type -crash avoid	-
steering	damage of steering wheel or not	ICLRAD	-	steering wheel deformation - car	location steering rim/spoke deformation	RDEFLOC
fire	the place fire started	BRANDURS	-	fire - car; fire start location - car	origin of fire	FIREORIG
door	condition door front left of passenger car	TUERZVL	-	front left door function - car; left frontal door opening longitudinal deformation - car	location& opening& damage/ separation	DLOCAT& DOPEN& DFAILURE
	condition door lock front left of passenger car	TUERSVL	-	front left door function - car	location& opening& damage/ separation	DLOCAT& DOPEN& DFAILURE
	condition door front right of passenger car	TUERZVR	-	front right door function - car; right frontal door opening longitudinal deformation - car	location& opening& damage/ separation	DLOCAT& DOPEN& DFAILURE
	condition door lock front right of passenger car	TUERSVR	-	front right door function - car	location& opening& damage/ separation	DLOCAT& DOPEN& DFAILURE

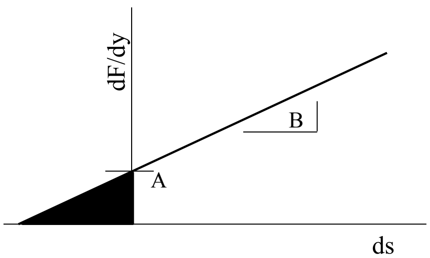
	GIDAS		iGLAD	INTACT	NASS	
Type of variable	**Description**	**Name**	**Name**	**Name**	**Description**	**Name**
	condition door rear left of passenger car	TUERZHL	-	rear left door function - car; left rear door opening longitudinal deformation - car	location& opening& damage/ separation	DLOCAT& DOPEN& DFAILURE
	condition door lock rear left of passenger car	TUERSHL	-	rear left door function - car	location& opening& damage/ separation	DLOCAT& DOPEN& DFAILURE
	condition door rear right of passenger car	TUERZHR	-	rear right door function - car; right rear door opening longitudinal deformation - car	location& opening& damage/ separation	DLOCAT& DOPEN& DFAILURE
	condition door lock rear right of passenger car	TUERSHR	-	rear right door function - car	location& opening& damage/ separation	DLOCAT& DOPEN& DFAILURE
under ride	whether vehicle was under-ridden or under-rides	ARTUFAHR	-	-	override/ under-ride	RIDEUP
rollover	rollover occurrence	ROLLWANN	-	*included in CDC 6 code; for each event, the variable "occurred during rollover" describes whether the event occurred during rollover*	Rollover data-type	ROLLTYPE
reconstruction method	theory principal of reconstruction	REKOART	-	*variable-specific, for Delta-V: "DV source"*	basis for total Delta V (highest)	DVBASIS

	GIDAS		iGLAD	INTACT	NASS	
Type of variable	**Description**	**Name**	**Name**	**Name**	**Description**	**Name**
	program used for reconstruction	REKOPROG	-	*for momentum -based reconstruction, PC-Crash is used; for damage-based reconstruction, AI Crash is used*	basis for total Delta V (highest)	DVBASIS

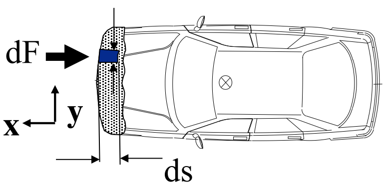
APPENDIX C: CALCULATION OF DELTA-V FROM CRUSH MEASUREMENT

Estimation of the deformation energy from crush pattern

Given a linear force-deflection characteristic



for each infinitesimal voxel (dy,ds) in the deformed area on the vehicle



the absorbed energy can be estimated as following, with B as the gradient and A as the intercept of the force-deflection function. A can be explained as a certain amount of force has to be applied before a residual deformation occurs.

$$E = F \cdot s$$

$$E = \int As(y)dy + \iint Bs(y)dyds$$

$$E = A\int s(y)dy + \frac{B}{2}\int s(y)^2 dy + \frac{A^2}{2B}\int dy$$

In case the resulting deformation force is applied from an angle α to the normal vector of the vehicle side, the resulting force and collinear deformation depth becomes:

$$F_R = \frac{F_N}{\cos(\alpha)},\ s_R = \frac{s_N}{\cos(\alpha)}$$

Thus the deformation energy has to be corrected:

$$E = \left(A\int s(y)dy + \frac{B}{2}\int s(y)^2 dy + \frac{A^2}{2B}\int dy \right) \cdot \left(1 + \tan^2(\alpha)\right)$$

Calculating of the Delta-V from deformation energy

Knowing the collision opponent masses and absorbed energy the Delta-V can be calculated from the equations for conservation of momentum and conservation of energy (AI Damage):

$$\Delta v_1 = \sqrt{\frac{2m_2(E_1 + E_2)(1-\varepsilon)}{m_1(m_1 + m_2)(1-\varepsilon)}}$$

When a restitution coefficient is not considered (WinSMASH) the equation is simplified to:

$$\Delta v_1 = \sqrt{\frac{2m_2(E_1 + E_2)}{m_1(m_1 + m_2)}}$$

The coefficient of restitution is dependent on the impact velocity and the impacted vehicle area. The corresponding values are most often derived from crash test, where all pre- and post-impact velocities are known and the deformation energy can be calculated.

An adjustment of vehicle stiffness and coefficient of restitution might be an additional alternative to compensate the bias between the trajectory-based and crush-based Delta-V estimation.

APPENDIX D. ASSESSMENT OF CRASH AVOIDANCE SYSTEMS

Table D1. Summary of crash avoidance technologies and related assessments and standards

Crash avoidance technology (other specific names)	Effectiveness Estimates and Sources	Potential risks	EuroNCAP	NHTSA	IIHS	Related ISO Standards
Adaptive cruise control	Adaptive cruise control is estimated to reduce 6-29% of rear-end crashes (Elvik 2006, Najm & Mironer 1998). Many other findings suggesting effectiveness of ACC can be found from (TraceD4.1.1-D6.2 2007 p15-16).	Driver's behavioral adaptation may negate the safety benefit of the system (TraceD4.1.1.-D6.2 2007 p.16, Kulmala 1997), for example due to increase average speed, reduced headway, larger braking force, greater lane position variability and longer hazard detection reaction (as reviewed in Humanist 2006, Regan et al. 2001)				ISO 15622:2010; ISO 22179:2009 for Full speed range ACC
Alcohol interlock	Alcohol interlocks are estimated to reduce 18% of crashes where alcohol was a factor (eSafety Forum 2005). Many other findings suggesting effectiveness of alcohol interlock can be found from (TraceD4.1.1-D6.2 2007 p18).	Studies by (Jullgren et al. 2005, Schonfeld & Sheehan 2004, as reviewed in TraceD4.1.1-D6.2 2007 p19) suggested that this system's effectiveness lowers over time.				
Attention assist (Driver Alert)	There has been an indication that the system "could prevent a third of all accidents involving a passenger car every year in Europe caused by drowsiness" (from http://www.euroncap.com/rewards/ford_driver_alert.aspx", "could prevent 1875 injury accidents involving a passenger car every year in Europe" (from http://www.euroncap.com/rewards/mercedes_benz_attention_assist.aspx).		Yes, in terms of rewards			
Automatic emergency call (eCall, (SYNC Emergency Assistance, Assist Advanced eCall,			Yes, in terms of rewards			

Crash avoidance technology (other specific names)	Effectiveness Estimates and Sources	Potential risks	EuroNCAP	NHTSA	IIHS	Related ISO Standards
Localized Emergency Call, Connect SOS)						
Autonomous emergency braking (City Brake Control, Forward Collision Mitigation, Front Assistant, Pre Sense Front, Pre Sense Front Plus, Front Assist, Active City Stop, Forward Alert, Collision Prevention Assist, City Emergency Brake, Collision Mitigation Brake System, PRE-SAFE Brake, City Safety)			Yes, in terms of rewards		Yes	
Blind spot monitoring (Rear Vehicle Monitoring System, Side Assist)			Yes, in terms of rewards			
Electronic Stability Control (ESP)	Electronic stability control is estimated to reduce 40% of all single-vehicle crashes and 75% of rollovers (IIHS). Many other findings suggesting effectiveness of ESC can be found from (TraceD4.1.1-D6.2 2007 p29-31).		Yes, in ratings	Yes, in ratings		

Crash avoidance technology (other specific names)	Effectiveness Estimates and Sources	Potential risks	EuroNCAP	NHTSA	IIHS	Related ISO Standards
Frontal collision warning (Forward Collision Warning)	FCW is estimated to reduce 7-80% of rear end crashes (FWHA 1998, Kanianthra and Murtig 1997, Kullen 2005, NHTSA 2001, Regan et al. 2002, and Sugimoto 2005) and 50-80% of head-on and object crashes (Lee et al. 2002). Many other findings suggesting effectiveness of FCW can be found from (TraceD4.1.1-D6.2 2007 p34-35).	Driver may experience behavioral adaptation that could result in a safety critical situation (Wege et al. 2003).		Yes, in ratings	Yes	ISO 15623:2013
Intelligent lighting systems	Intelligent lighting systems are estimated to reduce 18% of pedestrian/cyclist low-visibility crashes (eSafetyForum 2005).	Potential risks may occur as drivers increase their speeds as a result of better visibility (eSafetyForum 2005).				
Lane support systems (Lane change warning, Lane departure warning; Ford Lane Keeping Alert,Opel Eye, Lane keeping assistance; Lane Assistant, Active Lane Assist, Lane Assist, Lane Keeping Aid, Lane Departure Prevention)	Lane change warning systems are estimated to reduce 37-40% of drifting and lane change crashes (FHWA 1998, Kaniantrha and Murtig 1997, McKeever 1998). See also TraceD4.1.1-D6.2 2007 p40-41.					ISO 17387:2008
	Lane departure warning is estimated to reduce 25% of head-on collisions and 25% of off-path crashes (Abele et al. 2005, Regan et al. 2001). See also TraceD4.1.1-D6.2 2007 p42.		Yes, in terms of rewards	Yes, in ratings		ISO 17361:2007;
	Lane keeping assistance systems are estimated to reduce 17-25% of off-path crashes (eSafety Forum 2005, FHWA 1998, eImpact Project 2005), 24% of head-on collisions (eImpact Project 2005), and 60% of sideswipe collisions (eImpact Project 2005). See also TraceD4.1.1-D6.2 2007 p43.		Yes, in terms of rewards			
Programmable key system (MyKey)	"It is estimated that, if every car in Europe were fitted with such a system, some 4000 accidents and over 150 fatalities could be prevented" (from http://www.euroncap.com/rewards/ford_mykey.aspx)		Yes, in terms of rewards			
Multi Collision Brake						

Crash avoidance technology (other specific names)	Effectiveness Estimates and Sources	Potential risks	EuroNCAP	NHTSA	IIHS	Related ISO Standards
Pedestrian detection system	No published estimates of effectiveness of pedestrian detection systems					
(Pre-) Crash systems (Crew Protect Assist, Pre-Sense Basic, Proactive Occupant Protection, Collision Mitigation Brake System, PRE-SAFE, PRE-SAFE Brake)			Yes, in terms of rewards			
Reverse collision warning system	Reverse collision warning systems are estimated to reduce 50-81% of backing crashes (Lee 2002)	Some concern related to the systems were reported in Regan et al 2005 (See TraceD4.1.1-D6.2 2007 p47). The concern was mainly related to the limitation of the sensors used and therefore the concern may change as sensor technologies advances.				
Road departure warning	Road departure warning systems are estimated to reduce 24% of off-path crashes (Kaniantrha and Murtig 1997)					
Speed alert systems, or intelligent speed assistance (ISA)	The systems are estimated to reduce 36% crashes with injury and 59% of fatal crashes given 100% ISA penetration rate (Carsten & Tate 2005).		Yes, in ratings			
Vision Enhancement Systems (Adaptive Forward Lighting)			Yes, in terms of rewards			

APPENDIX E: ADVANCED STATISTICAL DESCRIPTIONS

Seemingly Unrelated Regression (Approach 1)

Seemingly Unrelated Regression (SUR) can be written as a system of equations as follows:

$$\begin{bmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \end{bmatrix} = \begin{bmatrix} \mathbf{X_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{X_2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

where the 1 subscripts refer to US and 2 refer to EU. X_i is an n_i x r_i matrix of predictor values for each case; r_i is the number of predictors in the i^{th} dataset. β_i is a r_i x 1 vector of coefficients estimated in the modeling process. Finally, ε_i is an n_i x 1 vector of residual error terms determined in the modeling process. These equations produce the same maximum-likelihood models as logistic regression on separate datasets, but with joint variance estimates. Once the models are selected, we can test the hypothesis that coefficients are the same:

H_0: $\beta_{j1} = \beta_{j2}$ (Hypothesis 1)

Where j indexes predictors from 1… $r_{i.}$

The test statistic, *S*, is equal to the difference between the coefficients, or β_{j1}-β_{j2}. Using a Wald test, *S* has a chi-square distribution with 1 degree of freedom when H_0 is true. Thus, the *p*-value for *S* is its probability under that distribution.

We can also choose an alternative hypothesis H_a, which can be described as:

H_a: $\beta_{j1} = \beta_{j2} - \delta$ (Note that δ is not a risk in this case, but part of a coefficent)

Under H_a, power is calculated using Equation 13.

$$\beta = p(S < \alpha | \delta\ is\ true) = 1 - \left(\Phi^{-1}(\alpha) - \frac{\delta}{\sqrt{var\ of\ S}}\right) + \left(\Phi^{-1}(1-\alpha) - \frac{\delta}{\sqrt{var\ of\ S}}\right) \qquad (13)$$

where *S* is the test statistic, α is the selected Type I error rate, δ is the size of difference being evaluated, and Φ is the standard normal distribution.

Calculating distribution of risks for Approach 2

Let us denote the distribution of predicted risk functions for EU vehicles by $\mathcal{D}_{EU}$ and the corresponding distribution for US vehicles by $\mathcal{D}_{US}$. For every Delta-V value *i* and $j = 1, \ldots, \max(N_i^{EU}, N_i^{US})$, *let* $\hat{p}_{EU}^{i,j,1}$ be a realization of $\mathcal{D}_{EU}$ (i.e., an injury risk function which was randomly chosen according to the distribution $\mathcal{D}_{EU}$) and let $\hat{p}_{US}^{i,j,1}$ be a realization of $\mathcal{D}_{US}$. The overall predicted risk of a vehicle with the realization of injury risk functions corresponding to EU vehicles computed on the EU standard population is

$$R_1^{EU,EU} = \frac{1}{N^{EU}} \sum_i \sum_{j=1}^{N_i^{EU}} \hat{p}_{EU}^{i,j,1}(i)$$

while the overall predicted risk of a vehicle with the realization of injury risk functions corresponding to US vehicles on the EU standard population is

$$R_1^{US,EU} = \frac{1}{N^{EU}} \sum_i \sum_{j=1}^{N_i^{EU}} \hat{p}_{US}^{i,j,1}\,(i).$$

Analogous quantities $R_1^{EU,US}$ and $R_1^{US,US}$ can be defined for the US standard population by replacing $\frac{1}{N^{EU}}$ by $\frac{1}{N^{US}}$ and letting the index j in the second sum run from 1 to N_i^{US} instead of N_i^{EU}. This way, for these realizations of the injury risk functions, the risk ratios computed on the standard populations can be computed as follows:

$$RR_{EU}^1 = R_1^{EU,EU} / R_1^{US,EU}$$

and

$$RR_{US}^1 = R_1^{EU,US} / R_1^{US,US}.$$

This process is repeated 1000 or more times (let $K \geq 1000$ be the total number of times), each time with newly drawn realizations with $\hat{p}_{EU}^{i,j,k}$ and $\hat{p}_{US}^{i,j,k}$ $(k = 1, \dots, \mathrm{K})$ of the injury risk functions. This procedure yields sometimes the same and sometimes different values of RR_{EU}^k; more precisely, the end-product is an *empirical distribution of* RR_{EU} and an *empirical distribution of* RR_{US}.