iGLAD - Ein pragmatischer Ansatz für eine internationale Unfalldatenbank

iGLAD - A pragmatic approach for an international accident database

Hier werden wir Ihr Foto einfügen, bitte separat an den VDA senden

Jörg Bakker
Accident Research
Daimler AG

VDA Verband der Automobilindustrie

Abstract

The harmonization of accident data on a multinational level has always been a promising but ambitious target which has not been achieved to date. The development in recent years additionally emphasises the strong need for a balanced view on the global accident situation. This need arises from different areas: harmonization of legislation, improving road safety in emerging and low income countries, and generally reducing the number of road fatalities worldwide, which is also postulated in the UN Programme „Decade of Action for Road Safety 2011-2020".

Currently, different accident data already exists worldwide and in many cases it includes more detailed, so called in-depth information. There is a need for international accident data and as a consequence accident data projects appear in different regions of the world. This results in the demand to compare and merge data from different countries. Leveraging the resources of already available accident data worldwide can help speeding up the process of improving global road safety.

While previous approaches concentrated on building a somehow artificial data scheme and starting a complete accident database from scratch, the iGLAD project (Initiative for the Global Harmonization of Accident Data) tries to take a more pragmatic approach by building on top of the work already done in this area and complementing it. The target of iGLAD is to build an additional dataset as a compatibility layer between already existing data sets worldwide and integrating the structure of these datasets by defining a common data scheme. As a result, an individual data converter for each participating accident investigation team will be built, feeding each data set into a common data pool. This not only saves costs and time, and thus makes such a target more feasible, but also creates a pool of usable data at hand right from the start.

The iGLAD project kickoff took place in September 2011 at the ACEA offices in Brussels, bringing together representatives of accident data collection teams worldwide. This paper gives a short overview of the current status and first steps taken. Additionally, some methodological aspects are touched along with a glance at other projects working currently on related issues, providing additional input for iGLAD. Finally, an overview of next steps and possible future work is given.

1. Introduction

While certainly there is a strong demand for international accident data that provides some more level of detail than most national police recorded data, building up such a data pool is a big challenge and has not been achieved to date. Even limited to EU-27, a centralized multinational in-depth accident data project is very big in terms of organizational effort and financial demand and thus is hard to realize. The iGLAD project (Initiative for the Global Harmonization of Accident Data) takes a different approach which is bottom-up and of a more pragmatic and evolutionary character. Starting with different but already available pieces of data which, put together as they are, make up an inhomogeneous data set at first, iGLAD strives to build a usable and more homogeneous data set out of it. Also in the long run, iGLAD tries to initiate a convergence of in-depth accident data sources, as more and more data will become globally available.

2. History

Being a young project that kicked of at the end of 2011, iGLAD's history is rather brief. Basic discussion started within the GIDAS (German In-Depth Accident Study) [1] steering committee, triggered by requests from emerging in-depth data projects in other countries seeking for support and best practices in how to set up an in-depth data investigation. Central point of a detailed in-depth investigation is the code book which reflects essentially the complete data scheme. While there are differences between countries, for example in infrastructure and car-fleet, the core structure of such an accident data scheme and the needs of users of the data from different organizations (governmental, automobile industry OEMs and suppliers, educational and research institutes) are essentially the same.

Experiences with GIDAS and other accident investigation projects show that a full blown in-depth project can get very complex and a challenge to maintain. There has been agreement within the GIDAS administration on sharing knowledge about the well-proven data scheme of GIDAS to facilitate upcoming in-depth data to be better comparable. Tying in with this momentum and starting with the FIA Mobility Group in October 2010, iGLAD has been initiated as a working group to address this challenge. Supported by FIA and ACEA, this working group should have the goal to define a common standardized accident data set as effective foundation for developing and measuring road safety policy and interventions. It also should establish how this data set helps to achieve the goals of the "European Road Safety Action Programme" [2] and the „Decade of Action for Road Safety“ [3].

The taskforce iGLAD then was confirmed at the FIA Manufacturers Commission on 2nd March 2011. After presenting the basic concepts of iGLAD to NHTSA/NCSA, especially the NASS group in April 2011 and at the VDI congress [4], the project's kick-off meeting followed on 30th September 2011 at the ACEA offices in Brussels. This also marked the beginning of common and cooperative tasks of FIA and ACEA within the iGLAD project. One such task is a project started by FIA to analyse the situation in low-income and emerging countries, complementing the efforts of ACEA which initially address in-depth projects in higher and middle-income countries. The next section goes into more detail about participants and results of the kick-off workshop.

3. Status

The kick-off meeting in September 2011 was primarily an organizational step to get representatives of the different projects and organizations together. Content related it was an initial step to get an overview of the principal intention to cooperate and to gather some feedback on possible pieces of data that could be available for use in a common database.

Participating organizations at the kick-off meeting of iGLAD were: IRTAD (OECD), MUARC (Australia), VSRC (UK), DaCoTA (EU), CEESAR and LAB (France), ARU and GIDAS (Germany), MTI (Poland), George Washington University (USA), IDIADA (Spain), FIA and ACEA (Belgium). Additionally, the following organizations have been kept in the loop: NHTSA, NCSA, IIHS (USA), EC, ERSO (EU), SRA (Sweden), JP Research (India, USA), Uni Pavia (Italy), ITARDA (Japan), KATRI (Korea), CDV (Czech Republic), and the "Kuratorium für Verkehrssicherheit" (Austria).

The participating organizations presented details about their relationship, support and possible contribution to iGLAD. A first impression of the potential of a common database could be derived from a survey prepared beforehand that consisted of a template reflecting a first and very rough draft of a common data scheme. The templates were completed by each participating organization according to their own in-depth data studies. A selected overview of the first draft common data scheme and availability in the participating projects is shown in table 1. Each row lists the information item that should be contained in the common data set and the participating organizations are listed in the columns. Ellipsis in table cells denote restricted availability for this particular item, details are omitted. The conclusion is that a common data set including the data of the currently participating organizations already should have quite some potential.

4. Methodological Aspects

This section addresses related projects and how iGLAD differs from them. Also, principal methodological issues in context of the iGLAD approach are shown with possible solutions to handle them.

Related Projects

While there already have been other projects addressing the need for a multinational in-depth accident database, these mostly have a different focus than what iGLAD is targeting for. Even more, iGLAD should be able to contribute to and improve the current situation by complementing the work previously done and not replacing it. This should lead to an overall more complete solution which should be beneficial for all involved parties.

Following, is an overview of what has been done or still is underway in this field, not striving for completeness. First there are projects that define an in-depth data scheme or standard themselves and some of them also generate data or set up own teams to collect data: STAIRS [5], EACS [6], Pendant [7], TRACE [8], SafetyNet [9] and the currently running project DaCoTa [10].

Then, there are projects with a special focus: motorcycle study MAIDS [11] and truck study ETAC [12].

Also very important in this context are national statistics on a macroscopic level. They contain basic accident numbers of a larger scale of different countries and take care of a harmonized understanding and definition of the parameters contained (for example the definition of fatality): IRTAD [13] and CARE [14].

CARE data is based on disaggregated data and thus has access to individual accidents, but is limited to EU countries. IRTAD also contains non European countries.

Finally, CADaS [15] has introduced a reduced data scheme and a proposal to gradually implement it on a national level in Europe. This scheme contains 73 variables and is mainly based on CARE. The detail of data is somewhere in between national level and in-depth, reconstruction information is not planned to be included. Out of the projects listed here, CADaS comes closest to what iGLAD is supposed to be.

However, the main differences of these projects compared to iGLAD are that most of them are limited to EU countries, they have a fixed time frame (except for the national data projects IRTAD and CARE) and some of them have a special focus. All of the projects either investigate data for themselves or design a data scheme to be filled by future projects, which can be considered as a top-down approach.

In contrary, iGLAD tries to follow a bottom-up like approach, by employing what is already there. The basic difference is that no accident investigation teams are installed and no new accident case data is created within iGLAD. It is supposed to provide some kind of glue between already existing projects. Also, iGLAD is intended to be very simple, leaving the sophisticated details to the different projects. This not only has practical reasons, as a very fine grained standard containing a long list of parameters is hardly a basis for a common data set of different data schemes. Also, with a simple standard, details and country specific issues are left to the particular in-depth study, complementing these studies and bringing them closer to a wider and more globally oriented audience of researchers.

Define the common data scheme

The approach taken by iGLAD is very pragmatic. Basically, it boils down to: see what is already there and build on top of it. Also, the result is supposed to be kept small and simple. iGLAD strives to do the best job to unify a limited number of parameters while sticking to realistic targets and staying effective. To achieve this, the different interests of the supporting members have to be balanced carefully. The level of detail provided by the resulting common data subset is not only a technical question, but also depends on the interests of the consortium partners. The result should be a well-balanced data set, where each party gives and receives comparable values. An bonus for each participant is the detail that each in-depth project is able to provide additionally as a superset of the common data subset and so the common data subset can serve as an entry point for further data analysis (or contracted analysis) performed by a participating project.

Nevertheless, despite its targeted simplicity, it is important that the data forms a useful basis for typical questions in the domain of accident data analysis. To accomplish this, the working group has to prepare relevant use cases of the data for demonstration purpose.

Adapt different samples

One crucial step for bringing the iGLAD concept to life is adapting the different accident data samples to form a best possible homogeneous data set, resulting in a quality and expressiveness of the data that is sufficient for real life problems.

The issue of adapting the data samples has two largely independent dimensions. First, on a parameter level, data converters have to be built that are able to map the different schemes to the common data scheme. Second, the different sample characteristics must be compensated for differences or bias, which only affects the case level. The good thing is these two issues can be handled separately. Case and parameter level can be considered as orthogonal, that is they can vary independently from each other.

The first issue is mainly a work that has to be done only once by setting up a data converter for each sample. Close knowledge of the parameters in each sample are important for defining the most appropriate mapping between the values. Simple example: it is easy to convert between units like inch and cm, but mapping two different accident types involves accounting for regional and systematic differences.

The second issue can be addressed by the use of multinational statistics like IRTAD and CARE which can serve as a link between the in-depth data samples, provided that the in-depth samples include some parameters of the national statistics and that they are large enough to reflect the real world accident situation in the specific country.

IRTAD already does a lot of work to harmonize national statistics and is supporting iGLAD. Adapting a sample to national statistics can be accomplished by introducing weighting factors to compensate for differences in the sample characteristics of in-depth and national data.

Assumed that the raw data from national statistics can be accessed on a case level, arbitrary multidimensional tables of all available parameters can be generated easily for a specific country. In this ideal case, weighting factors can be calculated directly for the parameters that are present in both, the national statistics and the in-depth sample. These weighting factors should be updated regularly (for example once a year) to reflect changes over time. However, in most cases, accessibility of national data is more restricted and only distributions of single parameters or crosstabs of two or three parameters are available. Then, weighting factors can be calculated by filling up the contingency table of the weighting parameters with an appropriate statistical method. A simple example using the IPF (Iterative Proportional Fitting) algorithm [16] shall illustrate a possible weighting procedure, its successful application and possible failures.

As the name suggests, IPF adapts the frequencies (cell values) in a contingency table to a marginal distribution using an iterative process that must converge in order to get a result. Fortunately, in most real life scenarios, convergence of IPF is good and fast. Table 2 shows how IPF is applied for the two parameters “injury severity” (fatal, injured) and “location of accident” (rural, urban). Expanding this to n parameters leads to an n-dimensional crosstab that can be handled by IPF analogous.

Accident data used for this example is a GIDAS in-depth sample and national statistics from Germany and Austria, all from the accident year 2009. Input is a crosstab “location of accident” vs. “injury severity” in GIDAS and four marginal distributions of each of the two parameters for Germany and Austria. The example has quite some realistic aspects as the injury severity is often needed for assessment of the potential of a safety measure. Apart from the general availability of injury severity in the national statistics and thus its likely application in a weighting procedure, accurate estimates of the injury severity is desirable.

As the crosstabs for these two parameters are also provided for Germany and Austria, error checking can be conducted. Here, the overall error is calculated as the maximum relative error over all cells related to the whole sample size.

Start of the iteration is a table combined of the marginal values of the national statistics and the frequencies of the combined parameter values in the GIDAS sample, which seed the starting table (top of table 2). Seeding has an influence on the cell values of the resulting crosstab, but not on its margins, they remain the same as in the starting table. Zero values in the starting table are a possible reason for non-convergence, as zeros are invariant throughout the iteration. This can be avoided by injecting small numbers replacing the zeros. In the Iteration step, alternating rows and columns are calculated by weighting the corresponding value in the previous table with a quotient of the original marginal value and the one from the previous iteration step.

In this example, convergence is sufficiently reached after five steps. The resulting table is shown in the lower right corner of table 2 for the German national data and both results for Germany and Austria are shown in chart 1. Although there is bias in the sample, IPF provides a good estimate for the cells of the crosstab of the two weighting parameters. This is especially true for larger groups (injured persons). The maximum error rate amounts for 0.2% in the estimates for Germany and Austria, which is quite good in the context of the bigger group of injured persons. However, for smaller groups like the fatalities the error rate can have quite some impact on the stability of the results, which then have to be interpreted carefully.

Finally, weighting factors can be derived as the quotient of percentages of the national estimates (after applying IPF) and the original in-depth sample percentages. Each accident in the sample then can be weighted with a factor that is given by the estimated crosstab cell value entry with the particular combination of weighting parameter values for this accident. This step is not very revealing to show for this small example, as more than two weighting parameters should be involved for acceptable results.

Of course, the more interesting situation where parameters not contained in the national statistics (a typical example may be airbag deployment) are estimated is likely to introduce additional errors with the constraint that these errors cannot be determined, even in a test case.

A reason for the results being quite good in this example is the inner relation between “location of accident” and “injury severity”, which is usually also not affected by regional differences. Reasons for this relation are higher speeds and more single car accidents in rural areas.

Adapting to a sample is theoretically not a matter of country borders. Hence, work is also underway to provide cluster methods to put similar countries in terms of accident data characteristics together. This also has potential to increase the size of the in-depth subsamples.

5. Next Steps, Long Term Goals

After primarily dealing with organizational issues at the kickoff-meeting, the first working group meeting is planned for early 2012. There, the first steps towards a common data subset will be addressed. This includes identifying the common parameters and finding an agreement which of the parameters may be included in the common data subset. Having accomplished this, the next step is to build data converters, to transform the data into a first version of the common data subset. In some cases this may involve some contract work to be accomplished, leaving funding issues to be addressed by FIA and ACEA. However, most of the work should be done self contained by the iGLAD working group and its associated organizations, reflecting the light-weight nature of the project. Long term goals of iGLAD could be some form of official standardization of the data scheme and also a certification procedure that could address quality maintenance issues of the data set.

6. References

[1] German In-Depth Accident Study, http://www.gidas.org

[2] EU Commission, European Road Safety Action Programme, ISBN 92-894-5893-3, 2003.

[3] WHO, http://www.who.int/roadsafety/decade_of_action/en/

[4] D. Ockel, J. Bakker, R. Schöneburg, “Internationale Harmonisierung von Unfalldaten; Fortschrittsbericht des FIA / ACEA Projekts iGLAD (Initiative for the Global Harmonization of Accident Data)“, VDI Kongress 2011, Berlin.

[5] STAIRS, EC 4th Framework Programme, “Standardisation of Accident and Injury Registration Systems”, Final Report, Contract N° : RO-96-SC.204,1999

[6] EACS, ACEA / EC project 1996 – 2001 coordinated by CEESAR, “European Accident Causation Survey - The Databank Questionnaire”, 2001

[7] Pendant, http://www.vsi.tugraz.at/pendant/

[8] TRACE Project, http://www.trace-project.org

[9] SafetyNet, http://erso.swov.nl/safetynet/content/safetynet.htm

[10] DaCoTa, http://www.dacota-project.eu/

[11] MAIDS, http://www.maids-study.eu/

[12] ETAC, EU Commission, “European Truck Accident Causation – Final Report”, 2006

[13] IRTAD, http://internationaltransportforum.org/irtadpublic/index.html

[14] CARE, http://ec.europa.eu/transport/road_safety/specialist/statistics/index_en.htm

[15] CADaS, SafetyNet Integrated Project D.1.14, “The Common Accident Data Set”, 2008

[16] W.E. Deming, F.F. Stephan, On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known, Annals of Mathematical Statistics 11 (4): 427–444. doi:10.1214/aoms/1177731829. MR3527, 1940.

Picture Annex

	BASt	LAB	IDIADA SP	IDIADA CZ	ITS	JP Research	Uni Pavia	OTS	CCIS	DaCoTA
General Data										
1. Date, place	yes	yes	yes	yes	yes	yes	yes	yes (...)	yes (...)	yes (...)
2. Original police recorded data (for weighting)	yes	yes	police agreement	police agreement	yes	yes (...)	yes	yes	yes	yes
3. Accident description	yes	yes	yes	yes	no	yes	yes	yes	yes	yes
4. Pictures	yes	yes	yes	yes	no	yes	yes	yes	yes	yes
5. Accident sketch (scaled) with final positions and objects	yes	yes	police agreement	yes	no	yes	yes	yes	no	yes
6. Accident and collision type	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
7. Environment: type of street, light & weather conditions, urban/rural	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
8. Emergency arrival (time)	yes	no	police agreement	yes	no	yes (...)	no	yes	no	yes
Vehicle / Pedestrian										
1. Type of vehicle (make, model)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
2. Registration year	yes	yes	yes	yes	possible	yes	yes	yes	yes	yes
3. Vehicle parameters (mass, engine type, number of seats, shape)	yes	yes	yes	yes	possible	yes	yes	yes	yes	yes
4. Deformation (VDI/CDC)	yes	yes	yes	yes	possible	yes	yes	yes	yes	yes
5. Systems (passive/active safety)	yes	yes	integral safety	no	possible	yes	yes	yes	yes	yes
6. Pedestrian information (if any involved in accident)	yes	yes	yes	yes	yes	yes	yes	yes	No peds	yes
Occupant										
1. Age, gender, weight, height	yes	yes	certain	yes	age, gender	yes (...)	yes	yes	yes	yes
2. Injury severity (MAIS), AIS of body region, fatal injury	yes	yes	yes	yes	possible	yes (...)	yes	yes	yes	yes
3. Restraint systems (presence, use, and deployment)	yes	yes	yes	yes	no	yes	yes	yes	yes	yes
Reconstruction										
1. Collision opponent	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
2. Collision speed	yes	yes (...)	yes	certain	no	yes (...)	yes	yes	no	yes
3. Driving speed	yes	yes (...)	yes	yes	no	yes (...)	yes	yes	no	yes
4. Delta-v	yes	yes	yes	certain	no	yes (...)	yes	yes	yes	yes
5. EES	yes	yes	yes	certain	no	yes (...)	yes	yes	yes	yes
6. Variables above for each collision (primary, most severe)	yes	yes (...)	yes	yes	no	yes	yes	yes	yes	yes

Table 1 – Overview of data availability survey for first draft of common data scheme (excerpt).

Germany	Sum	Urban	Rural	
Sum	401,823	258,919	142,904	← input
Injured	397,671	257,694	139,977	
Fatalities	4,152	1,225	2,927	

Germany	Sum	Urban	Rural
Sum	100.0%	64.4%	35.6%
Injured	99.0%	64.1%	34.8%
Fatalities	1.0%	0.3%	0.7%

GIDAS	Sum	Urban	Rural	
Sum	2,203	1,680	523	
Injured	2,178	1,676	502	← input
Fatalities	25	4	21	

GIDAS	Sum	Urban	Rural
Sum	100.0%	76.3%	23.7%
Injured	98.9%	76.1%	22.8%
Fatalities	1.1%	0.2%	1.0%

Iteration 1	Row			Col	
401,823.0	306,677.4	95,145.6		258,919.0	142,904.0
397,671.0	306,013.1	91,657.9	396,023.8	258,358.1	137,665.7
4,152.0	664.3	3,487.7	5,799.2	560.9	5,238.3

Iteration 2	Row			Col	
401,823.0	259,834.3	141,988.7		258,919.0	142,904.0
397,671.0	259,432.7	138,238.3	397,648.2	258,518.9	139,129.4
4,152.0	401.6	3,750.4	4,174.8	400.1	3,774.6

Iteration 3	Row			Col	
401,823.0	258,931.6	142,891.4		258,919.0	142,904.0
397,671.0	258,533.7	139,137.3	397,670.7	258,521.1	139,149.6
4,152.0	398.0	3,754.0	4,152.3	397.9	3,754.4

Iteration 4	Row			Col	
401,823.0	258,919.2	142,903.8		258,919.0	142,904.0
397,671.0	258,521.3	139,149.7	397,671.0	258,521.1	139,149.9
4,152.0	397.9	3,754.1	4,152.0	397.9	3,754.1

Iteration 5	Row			Col	
401,823.0	258,919.0	142,904.0	401,823.0	258,919.0	142,904.0
397,671.0	258,521.1	139,149.9	397,671.0	258,521.1	139,149.9
4,152.0	397.9	3,754.1	4,152.0	397.9	3,754.1

Result of IPF: GIDAS sample adapted to German national statistics

Germany	Sum	Urban	Rural
Sum	100.0%	64.4%	35.6%
Injured	99.0%	64.3%	34.6%
Fatalities	1.0%	0.1%	0.9%

max error 0.2%

Table 2 – IPF algorithm applied to GIDAS as a sample of German National Statistics.

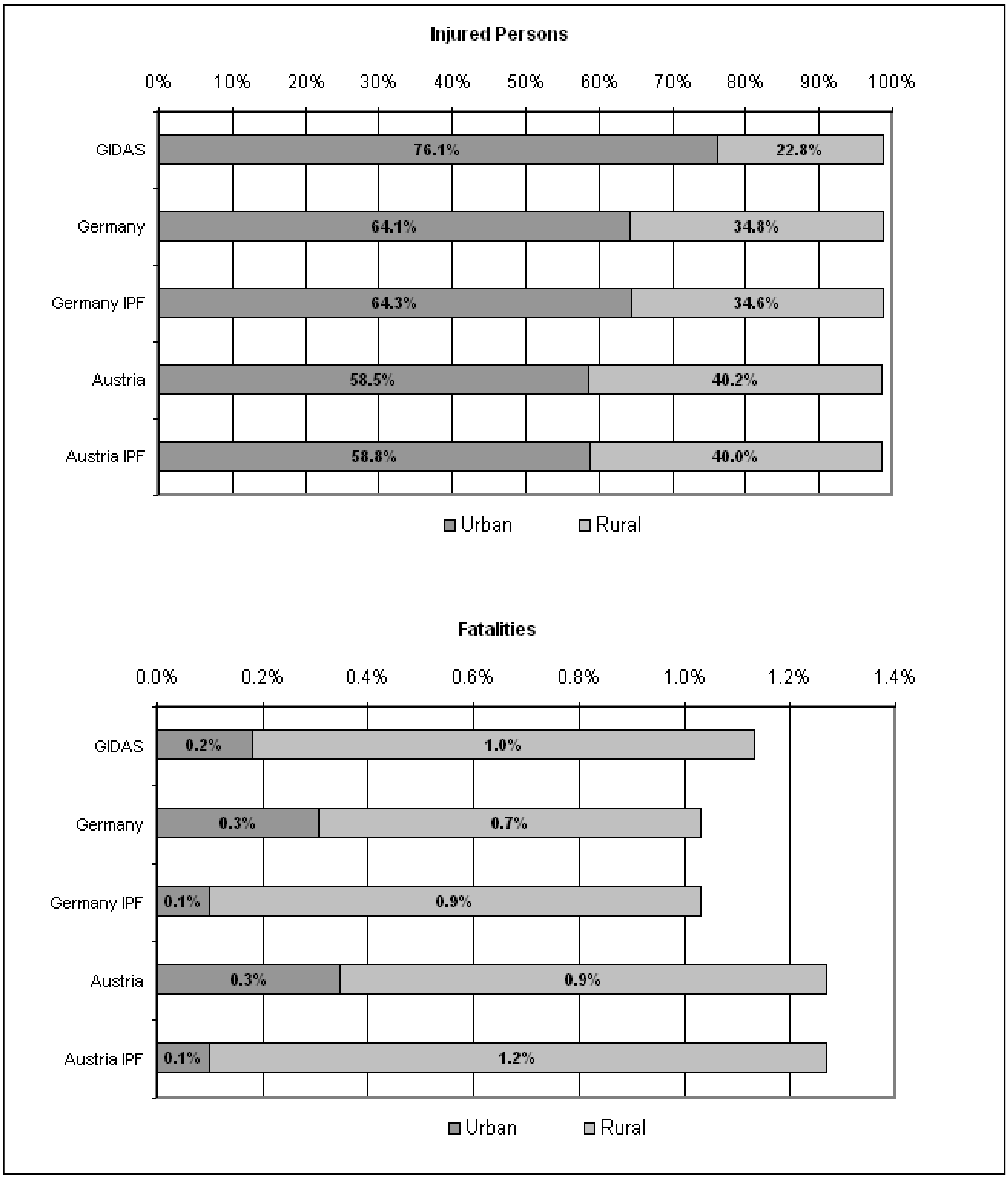

Chart 1 – Injury severity and location of accident in GIDAS sample adapted to Germany and Austria.