

Elaboration of rules transforming national road accident values into European

GEORGE YANNIS
Lecturer

JOHN GOLIAS
Associate Professor

ABSTRACT

The objective of this research is the development of an original methodology for the improvement of the comparability of national road accident data held inside CARE, the European database with disaggregate road accident data. The originality of this methodology lies to the fact that comparability is achieved through restructuring of existing national road accident data files without any harmonisation of the original data collection methods of the EU member states. The methodology developed consists of three distinct phases concerning the selection of priority variables and values, the elaboration of common definitions and labels and the development of the related transformation rules at a very detailed level. The application of this methodology has produced a list of common-definition values of 48 priority variables, making thus CARE database a more useful tool for European-level comparisons.

1. INTRODUCTION

The Community database with disaggregate data on road accidents (CARE), developed by the European Commission, is an important tool for road accident analysis and the development of traffic and road safety policies [1]. The CARE database is composed of national road accident data in disaggregate form (information for each individual accident is contained in the database), supplied by the member states of the European Union (EU). A number of incompatibilities inherent to the national data contained inside CARE, are limiting CARE potential for international comparisons [2]. These incompatibilities refer to the different national accident data collection systems, the different national accident data collection systems, the different national data file structures, the different

information collected and the different corresponding definitions used [3] [4].

Within this framework, the objective of this research was the development of a methodology for the improvement of the comparability of national road accident data held inside CARE database. This was achieved by restructuring existing national road accident data files within the CARE system, rather than harmonising the collection methods of member states (e.g. use of a standard road accident form). Increased comparability among member states road accident data held in CARE is an important step towards enhanced use of the database, and therefore towards increased potential support to European and national road safety policies [5].

The developed methodology allows comparability between road accident data by the use of transformation rules. These rules transform original national road accident variables into common road accident variables, which can then be compared among all or several member states. It is however noted that achieving comparability does not take into account differences in the reporting level of accidents, or variations in data quality that exist between the road accident data systems of the various EU member states.

This research was carried out within CAREPLUS project, which was managed by a working group of national experts who brought together information on road accident data collection systems from all fifteen EU member states. In the first phase of this project [6] the methodology was developed and a first set of common-definition values of 29

priority variables was elaborated. In the second phase [7], the methodology was refined and another set of common-definition values of 19 priority variables was elaborated, together with the respective glossary of definitions for the values of all 48 priority variables.

2. TERMINOLOGY

Common Variables: common road accident variables are variables made comparable across all or several EU member states. They are referred to as European variables. They are derived from the original national variables using transformation rules.

Common Values: each road accident variable contains values. For example, the road accident variable "Vehicle Type" contains values such as Car, Pedal Cycle and Goods Vehicle. Common road accident values are values made comparable across all or several member states. They are sometimes referred to as European values. They are derived from the original national values using transformation rules.

Transformation Rules: transformation rules are logic statements which contain the Boolean operators "AND" (intersection), "OR" (union) and "=" (equal). Transformation rules are applied to national road accident values to obtain common or European road accident values.

Label: a label is allocated to each common value and describes the meaning of a European common value by using elements of each member states' national definition.

Grids: grids are Tables, which summarise the national definition components for each common value and indicate the availability of each component for each Member State. Grids are the building blocks for creating transformation rules.

Glossary: the glossary summarises the European meaning of each common value and its availability across member states.

Associated variable: an associated variable is not included in the common variable list, but its values are used in deriving transformation rules.

3. DEVELOPMENT OF THE METHODOLOGY

The methodology developed for the improvement of the comparability of national road accident data contained inside CARE database consisted of three distinct phases as these are described in the following sections:

- selection of priority variables and values
- elaboration of common definitions and labels
- development of transformation rules

3.1. Selection of priority variables and values

The first phase of the methodology aims to identify a core set of priority variables and associated values which most member states collect in the framework of their national data systems, and which at first sight appear to be similar. The remaining variables and values from which the possibilities for producing common data are either less obvious or less significant, from a road safety perspective, should be considered at later stages.

Selecting appropriate variables for inclusion in the comparability process is judged by four main criteria:

- Significant relevance in the context of road safety analysis
- The presence of the variable in the national data file of a sufficient number of countries
- Similarity in the definition of the variable across member states
- The usefulness of the variable and its values for comparative analysis

The choice of priority values is an iterative process, whereby the initial selection is modified in light of the collection of more detailed information about the definitions of values. The initial selection of values within each priority variable is made on the basis of wide availability among all member states, either directly or indirectly (after simple transformation rules). For example:

- either values such as 'rain' are collected directly in a national data set for the priority variable 'weather' and, in principle, are directly comparable between member states;
- or values can be derived by combining different values which are also collected for the same priority variable 'weather'. For example in Great Britain the value 'rain' is not directly available in their national data set but can be produced by combining the values 'rain with high wind' and 'rain without high wind'.

This is followed by a detailed collection of definitions to establish the true meaning of all the selected priority values in each Member State. The definitions provide the means for iteratively finalising the priority value labels to a state where, in principle, they should be comparable, in definition and data coverage, between member states. This iterative process means that some common English language value labels are repeatedly revised. As a result, some additional priority values are introduced to enhance comparability between member states.

3.2. Selection of priority variables and values

The second phase of the methodology concerns the collection of national definitions and the elaboration of the common definitions and labels for the priority values. The glossary with national definitions not only serves the refinement of the priority values, which will make up the initial core of comparable variables/values at European level, but also provides a glossary documentation for external users, containing the meaning of common European variables and values and highlighting the inherent problems of data comparison between EU member states.

The glossary is based upon, and developed from the initial identification of priority values (phase one), and in an iterative fashion assists the refinement of the formulation of the priority values. Figure 1 schematically describes the iterative process for developing the common definitions and labels of the priority variables and their values.

The first step is the agreement of the names in English (labels) of the priority variables and values, which is followed by the initial collection of general English text definitions (Textual). An example of the collection form for the collection of English text definitions (Textual) is shown for the value "slightly injured" of the variable "INJURY SEVERITY (PERSON)" in the following Figure 2.

These initial textual definitions for all member states are then broken down into definition components and represented in grid form (with components in rows and member states in columns), to facilitate detailed comparison. These grids, as shown in Figure 3, are in turn used to revise value labels. The finalisation of value labels provides the basis for establishing the transformation rules.

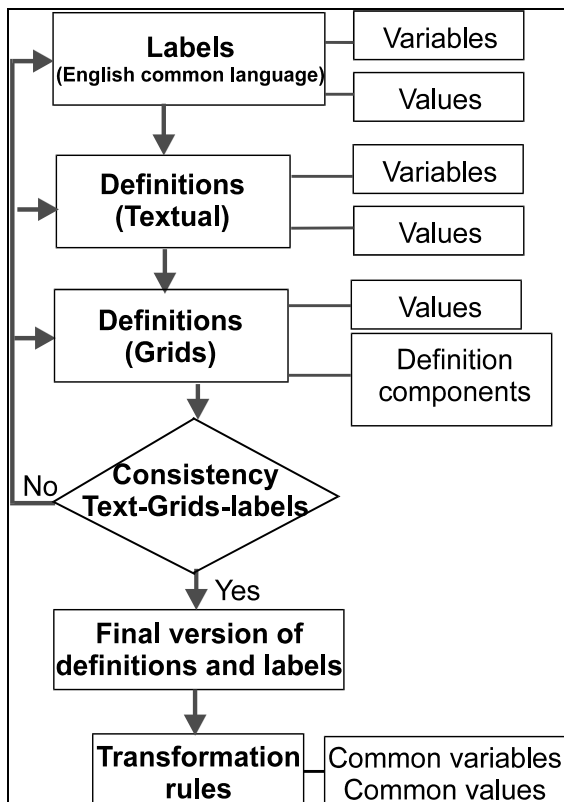


Figure 1. Flow chart for the elaboration of common definitions and labels

YEAR: 92		NATIVE VARIABLE NAME:	ENGLISH VARIABLE NAME:
COUNTRY: FRANCE (F)		GRAVITE DE L'USAGER	INJURY SEVERITY (PERSON)
VALUE NAMES		DEFINITION:	
NATIVE	ENGLISH		
blesse leger	slightly injured	Person involved in a road accident and who needs medical treatment without hospitalization or who needs to be hospitalized for a period of zero to six days etc.	

Figure 2. Example of form for the collection of national data

COUNTRY :		FRANCE			YEAR :		91		
Priority VARIABLES & values				VARIABLES & values					
VARIABLES	values	=	VARIABLES	values	OP	VARIABLES	values	OP	
INJURY SEVERITY (DRIVER)	slightly injured	=	person class	driver	AND	injury severity (person)	slightly injured		

Figure 4. Example of transformation rules

CARE European road accidents database									
slightly injured as a result of the accident							F		
opinion of the police							F		
injured, not seriously							F		
MEDICAL TREATMENT :									
injured, no medical treatment required							no		
injured, medical treatment required							F		
HOSPITALISATION :									
injured and not hospitalised							F		
injured and not hospitalised more than 24 hours							F		
injured and hospitalised 6 days or less							F		
INJURY TYPES :									
sprain							F		
bruise							F		
cut, not judged severe							F		
slight shock requiring medical treatment							F		
information not collected							no		

Figure 3. Example of gridline of national definitions for France

3.3. Development of transformation rules

Once the definitions and labels of the priority values have been finalised, after verifying the consistency of definition texts, grids and labels, the way is open for deriving priority values, for which data should be available, and in principle comparable, for all EU member states. The priority value will either exist directly in a national data set or it will have to be derived by combining related values, either from the same variable or from different variables. The transformation rules for the required combinations use the 'BOOLEAN ALGEBRAIC' operators “AND”, “OR” and “NOT”. An example of a transformation rule is shown in the following Figure 4.

The integration within CARE system of the derived common-definition values of the priority variables by the use of the transformation rules developed, make possible the comparisons of road accident data between EU member states.

4. AN EXAMPLE OF THE APPLICATION OF THE METHODOLOGY

The methodology developed for obtaining comparable road accident variables and values at EU level, is illustrated in the following example, using the variable ROAD SURFACE CONDITION. This variable is targeted as useful for road accident analysis work. The example illustrates the procedure followed in obtaining one common value from this variable: 'Snow or ice'.

The list of values for ROAD SURFACE CONDITION (after the application of a series of translation revisions) found in the national road accident data systems is shown in Table 1. From this Table it is obvious that values 'Snow', 'Ice', 'Dry', 'Slippery' and 'Wet' present a good degree of comparability and were targeted as common values.

Table 1. Tables of values for ROAD SURFACE CONDITION

Value	countries	Value	Countries
Bare, dry	1	Light snow/ snow slush	1
Bare, wet	1	Mud	4
Black ice	4	Not stated	5
Black ice (not salted)	1	Oil	4
Black ice (salted)	1	Other	8
Black ice, snow	1	Puddles	1
Clean	1	Scattered gravel	1
Damaged surface	2	Slippery	3
Dirt	1	Slippery (after dry spell)	1
Dirt, grease	13	Slippery due to oil, manure, clay	1
Dry	2	Slippery due to rain, snow, black ice	1
Dry and clean	3	Slippery due to snow	1
Flood	3	Slush	1
Frost, ice	2	Snow, slush	1
Frost	2	Snow	8
Gravel, sand	1	Snow/ ice	1
Grease, oil	1	Thick ice/ packed	9
Grit	2	Unknown	1
Humid	1	Water in wheel tracks	12
Ice, dew, snow	1	Wet	3
Ice	2	Wet/ damp	1
Leaves		Wheel tracks bare	

Table 2. Tables of values for the common value 'Snow, ice' (Y means available)

No of countries	A	B	D	DK	E	F	FI N	G B	G R	I	IR	L	NI	NL	P	S
Black ice	4		Y			Y				Y						Y
Black ice (not salted)	1	Y														
Black ice (salted)	1	Y														
Black ice, snow	1		Y													
Frost, ice	3								Y		Y		Y			
Frost	2				Y				Y							
Ice, frost, snow	1														Y	
Ice	1						Y									
Light snow, snow slush	1															Y
Slippery due to rain	1											Y				
Slippery due to snow	1			Y												
Slush	1						Y									
Snow, slush	1	Y														
Snow	8				Y	Y	Y	Y	Y	Y	Y		Y			
Thick ice, packed snow	1															Y
Snow, ice	1													Y		
Snow or ice (Common value)	15	Y	Y	Y	-	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	

A subset of the values in Table 1, closely related to 'snow' and 'ice' is shown in Table 2. Initially, the possibility of having 'Black ice' and 'Snow' as two separate common values was discussed. However, given that values such as 'Black ice or snow', or 'Slippery due to rain, snow or black ice', or 'Ice or frost or snow' were used in some national systems, the use of value labelled 'Black Ice' was not retained. Using 'Black ice' alone as priority value, comparability for 5 countries could have been possible, while using 'Snow' alone could have made possible comparability for 8 countries. Consequently, the common value 'Snow or ice' was selected. This value allowed comparability for 15 countries.

Transformation rules to generate this common value 'Snow or ice' for each country could then be applied using the results of Table 1. The rule is derived by adding the value definitions listed in the respective member states' column in Table 2.

For example, in order to transform the values from the Austrian national file to the common-definition value 'Snow or ice' the following rule is applied:

If the Austrian national accident record contains the value 'Black ice (nor salted)' or 'Black ice (salted)' or 'Snow,

slush' then the common-definition value 'Snow or ice' is assigned.

This example of the procedure required to derive just one common value from one of the simpler common variables targeted, illustrates the complex and time consuming task involved in obtaining comparability between road accident data used in the national data files of the EU member states.

5. ESTABLISHING THE LIST OF COMMON VARIABLES AND VALUES

The application of the methodology developed, resulted to the establishment of two respective lists of common variables and values as these are presented in Tables 3 and 4. The common-definition values elaborated, concerned 29 priority variables in the first phase of CAREPLUS project, and 19 additional priority variables in the second phase. The proposed common values for each common variable sum up to 100% as all non common values are contained inside value labeled "other".

Table 3. Common road accident variables and values of CARE PLUS 1

AGE	JUNCTION	VEHICLE TYPE
1 - 99	Yes	Agricultural tractor
SEX	No	Bus or coach
Male	Not stated	Car
Female	JUNCTION TYPE	Car or taxi
Unknown	Crossroad	Moped
INJURY SEVERITY	Level crossing	Motor cycle
Killed	Not at a junction	Heavy goods vehicle
Seriously injured	Roundabout	Lorry >= 3.5 tons
Slightly injured	t or y junction	Lorry < 3.5 tones
Injured	Other junction	Other motor vehicle
not injured	Unknown	Other non-motor vehicle
Unknown	AREA TYPE	Pedal cycle
PERSON CLASS	Inside urban area	Road tractor
Driver	Outside urban area	Road tractor with semi-trailer
Passenger	Not stated	Taxi
Pedestrian	MOTORWAY	Other
Unknown	Yes	Unknown
CAR PASSENGER TYPE	No	VEHICLE AGE
Front passenger	Not stated	0
Rear passenger	ROAD SURFACE CONDITION	1 - 2
Unknown	Snow or ice	3 - 5
ALCOHOL TEST	Dry	6 - 10
Not stated	Slippery	More than 10
Tested	Wet, damp, flood	Unknown
Unknown	Other, unknown	DRIVING LICENCE AGE
ALCOHOL LEVEL	SPEED LIMIT	0 years old
Less than 0.5	Less than 30 Km/h	1 years old
More than 0.5	30 - 50	2 years old
Unknown	51 - 80	3 years old
NATIONALITY DRIVER	81 - 100	4 years old
National	About 110 km/h	5 years old
Foreigner	120 Km/h	More than 5 years old
Unknown	Unknown	Unknown
PSYCHOPHYSICAL CIRCUMSTANCES	CARRIAGEWAY TYPE	HIT AND RUN
Drugs or medicine or fatigue	Dual carriageway	No
Others, none	Single carriageway	Yes
MOVEMENT PEDESTRIAN	Unknown	ACCIDENT SEVERITY
Crossing	JUNCTION CONTROL	Fatal accident

6. DEVELOPMENT OF A GLOSSARY OF DEFINITIONS FOR THE COMMON VALUES

Following the selection of the priority variables and values, the elaboration of common definitions and labels and the development of the related transformation rules, a

glossary providing short definitions for each common-definition value was elaborated. This glossary was derived from the definition grids, which contained detailed information on each common value for each Member State. The definition glossary has three components:

- a) a definition of the common value
- b) the availability of the common value in the data file for each Member State
- c) the availability of the common value in associated values

The value definition defines the data contained in the common value. When necessary, clarification of definitions is presented to avoid ambiguity. Exceptions to the definitions are also highlighted in the text. For example, the definition of the common value “no street lights” is as follows: “roadside not equipped with lights; opinion of the police; information only collected if dark for E, F, GR, IRL and S”.

The availability of the common value lists the countries, which contain the common value directly in their own data files or indirectly through the use of transformation rules. Exceptions to the definition found in various member states are also mentioned. For example, in the field listing data availability for the common value “no street lights” the following text is presented: “DK, F, GB, GR, IRL, NI, NL, S” This is the list of countries which have data for the common value “no street lights”.

The availability of the data in associated values provides information about other values (associated), where the required data can be found. For the countries listed, the common value is available only through the use of another common value, called an associated value. For example, the field concerning “availability in associated value field” for the common value “no street lights” the following text is listed: a) “darkness, lights unlit or no lights (A, B, E, FIN, L)” b) “darkness, no lighting or no moonlight (P)”. This means that the common value “no street lights” can be found only as part of the common value “darkness, lights unlit or no lights” for the member states A, B, E, FIN, L and as part of the common value “darkness, no lighting or no moonlight” for the Member State P. These two common values are defined as associated values.

7. CONCLUSIONS AND RECOMMENDATIONS

This research consisted of the development of an original methodology for the improvement of the comparability of national road accident data held inside CARE, the European database with disaggregate road accident data. The originality of this methodology lies to the fact that comparability is achieved through restructuring of existing national road accident data files without any harmonisation of the original data collection methods of the EU member states. Perhaps, this approach could also provide interesting results in other fields of EU statistics, where harmonisation of original data at national level presents similar difficulties.

The methodology developed consisted of three distinct phases concerning the selection of priority variables and values, the elaboration of common definitions and labels and the development of the related transformation rules at a very detailed level. This methodology has been applied and a list of common-definition values of 48 priority variables has been produced, making thus CARE database a more useful tool for European-level comparisons.

It is necessary that the results of the proposed methodology should be further validated by the use of the CARE system for the production of a series of EU-comparable road accident statistics. Within this process, the produced set of common-definition values should be used for the identification not only of the methodology limits but also of the CARE system potential [8]. This work could provide valuable feedback, from a user perspective, to the further development of the CARE system design and functionality [9].

The work for the further enhancement of road accident data comparability at European Union level should be continued [10]. Initially, the proposed methodology should be used for the elaboration of more values of the priority variables as well as values of new priority variables. Additionally, a great improvement of the EU-wide road accident data comparability could be the elaboration of a

common-definition of serious injury accidents (hospitalised-persons), allowing for use of a larger than the fatal accident sample files. This could be of particular benefit to member states with relatively few fatal accidents interested in generating comparable reports using more detailed accident categories requiring larger sample sizes.

Additionally, an attempt should be made for EU harmonisation of data at national level [11]. A basic action, and not necessarily very difficult to implement, could be the progressive adoption by all European countries of a common road accident data collection form containing uniform basic information on the accident, allowing for direct international comparisons [12]. These harmonisation actions could be implemented progressively; first the common approach should be defined in detail, then each country could optionally implement it whenever an update of the national data collection form is revised, so that in few years, common data collection methods are used in all European countries (including accession and other European countries).

8. REFERENCES

1. COMMISSION OF THE EUROPEAN COMMUNITIES, Council decision 93/704/EC of 30 November 1993 establishing a Community database on road accidents, Commission of the European Communities, OJ No L 329, Luxembourg, December 1993.
2. EUROPEAN COMMISSION, Report from the Commission on progress with the project and its future prospects: CARE Community database on road traffic accidents, European Commission, COM(97) 238, Brussels, May 1997.
3. IRTAD, (1992), Definitions and data availability. OECD/IRTAD special report, BAST, Bergisch Gladbach, Germany.
4. TRB, (1993), Accident data quality: a synthesis of highway practice. Transportation Research Board, NCHRP Synthesis 192, Washington DC.
5. ETSC, Transport accident and incident investigation in the European Union. European Transport Safety Council, Brussels, 2001.
6. CETE-Sud Ouest, A proposal for improving the comparability of road accidents within EU Member States using the CARE data base, CETE-Sud Ouest, Bordeaux, September 1997.
7. CETE-Sud Ouest, Development of a methodology for deriving comparable data variables from national accident data, CETE-Sud Ouest, Bordeaux, May 2000.
8. NATIONAL TECHNICAL UNIVERSITY OF ATHENS / DTPE, (1996), Current and future potential of a European road accident data base with disaggregate data. NTUA/DTPE, Athens.
9. ETSC, EU transport accident, incident and casualty database: -current status and future needs. European Transport Safety Council, Brussels, 2001.
10. YANNIS, G., GOLIAS, J., KANELLAIDIS, G., (1998), A comparative analysis of the potential of international road accident data files. IAATS Research, Vol. 22 No. 2.
11. GOLIAS J., YANNIS G., Dealing with lack of exposure data in road accident analysis, Proceedings of the 12th international conference "Traffic Safety on three Continents", Federation of European Road Safety Institutes, Transportation Research Board, Moscow, September 2001.
12. OECD, (1995), International road traffic and accident databases. OECD Seminar, Helsinki.

George Yannis, Lecturer

John Golias, Associate Professor

National Technical University of Athens, Faculty of Civil Engineering, Department of Transportation Planning and Engineering, 5, Iroon Polytechniou str., GR-15773, Athens, Greece.