Statistical annex

Note. This annex is extracted from the *Global status report on road safety*.

Explanatory notes

Background

The data presented in the following pages were obtained from the 20 countries that participated in the first global survey on road safety in 2008. The survey focused on the recommendations of the *World report on road traffic injury prevention* (*1*) as the basis for its structure and content. Most countries used the same methodology for data collection, as outlined in a survey protocol developed for the study. In over 95% of the countries, the implementation of the survey was coordinated by a National Data Coordinator (NDC) identified by the country, and was completed by teams of 6–8 key respondents including the NDC. The NDCs were trained in the methodology and coordinated the collection, validation and clearance of data, as well as the data entry, in the countries concerned. The survey instrument, protocol and accompanying guidelines and training materials were all available in the six WHO languages (Arabic, Chinese, English, French, Russian and Spanish). Where needed, NDCs coordinated the translation of these documents into the local language and then back-translated them for the data entry which was done in English. More details on the methodology used for data collection can be found at www.who.int/violence_injury_prevention/road_traffic/road_safety_status/2009.

The following sections contain country-by-country data obtained from the survey.
- Table A.1 includes the list of NDCs who, in collaboration with national authorities, played a key part in conducting the survey. Respondents from different sectors are also listed.
- Table A.2 provides detailed data on the 30-day adjusted number and rates per population of road traffic deaths, and on the proportion of road users by country/area. The table also includes modelled road traffic death numbers which have been generated. A short description of the process is presented below.
- Tables A.3–A.6 provide information on the status of laws, the enforcement of laws, and the coverage/wearing rates relating to the five road traffic risk factors (alcohol, seat-belts, child restraints, speed, and helmets)
- Table A.7 contains data on policy-related responses.
- Table A.8 includes information related to the availability of pre-hospital care.

The following section gives a brief description of the data processing, the comments on the obtained results (specifically on the reported, adjusted and modelled country-level death data) and the method used to develop the model.

Data processing

The data processing involved completion of the survey instrument and data entry at a country level, and validation at a regional level. Data cleaning and analysis were done at WHO headquarters in Geneva.

Reporting of country-level data

The final country responses were entered by NDCs into an online database specially prepared by WHO for this project. NDCs also uploaded supporting documents where applicable and available. Data was then validated at a regional level. Once finalized and approved by Regional Data Coordinators (RDCs), the data were then exported into Microsoft Excel for cleaning. At this stage, each country's data were examined for accuracy, consistency and validity on a question-by-question basis. Where necessary, NDCs were contacted and additional supporting documents were requested to clarify inconsistencies. A copy of the survey instrument and study protocol can be found at www.who.int/violence_injury_prevention/road_traffic/road_safety_status/2009.

As part of the data cleaning and validation process, exploratory analysis was done using STATA (*2*). The same software was used for all analysis and results presented in the earlier sections of this report.

Types of data utilized

Three types of data are used in this report:
– reported data from countries and secondary sources;
– data adjusted for the 30-day definition of a road traffic death in order to facilitate comparability;
– modelled numbers.

Reported data

In addition to the data obtained directly from countries, secondary data sources were used to:
– classify countries into income categories;
– generate road safety indicators such as the adjusted road traffic deaths and modelled road traffic death rates (with a 90% confidence interval) as reported in Table A.2.

Population and income data from the United Nations Population Division (*3*) and the World Bank (*4*) were used for this analysis.

Population estimates for 2007 are reported in Table A.2. Where there was no estimate available for a country for that year, published data for the latest year were used. For the modelling process, population estimates corresponding to the year of reporting were used (*4*).

In Table A.2, World Bank (Atlas method) gross national income per capita (GNI) for 2007 (*5*) (or latest available year) was used to categorize countries into:
– low-income = $935 or less;
– middle-income = $936 to $11 455;
– high-income = $11 456 or more.

More detailed subgroupings were used in the modelling process.

Adjusted data

Underreporting has been acknowledged for many years as an important reason for the difficulty in comparing road traffic crash data between countries. Additionally, the lack of harmonized definitions for road traffic deaths, the use of different data sources, and the quality of the reporting system have also been documented. Consequently a number of mechanisms were employed to address some of these issues in order to make data more comparable. This global survey employed the following two methods:
– the European Conference of Ministers of Transport (ECMT) standardized 30-day road crash fatality factors (*6*) to adjust all reported country/area data;
– a model using negative binomial regression.

The "reported" data in Table A.2 have been adjusted to this 30-day definition (see Table 1 for adjustment factors) and the new adjusted number is therefore used in the corresponding model and its result is presented in Table A.2.

Table 1. ECMT standardized 30-day road crash fatality adjustment factors

	30-day total	adjustment factor
On the scene/1 day	77%	1.30
3 days	87%	1.15
6 days	92%	1.09
7 days	93%	1.08
30 days	100%	1.00
365 days	103%	0.97

Modelled data

Developing a model

Before the modelling exercise, simple exploratory analyses were done to evaluate the distribution of the reported data, to identify potential outliers and to determine the extent of missing data. Decisions as to whether to include these outliers in the analyses, or whether to exclude them, were taken at several stages of the analysis. Where appropriate, imputation was done to compensate for missing information.

Completeness of data

Having adjusted the reported data to a 30-day definition to facilitate comparability, the next step in the process was to explore the completeness of the reported death data on the basis of reported vital registration (VR) data. Information on the completeness of VR data was obtained from previous WHO published reports (7,8) and was updated with the latest information from the WHO mortality database. This information was then used to classify countries into two groups, namely:

Group 1: countries with VR completeness greater or equal to 85% and external causes of death coded to undetermined intent less than 30% (Table 2);

Group 2: countries with VR completeness less than 85% or external causes of death coded to undetermined intent greater than 30%.

Group 1 countries include 3 high-income countries (Table 2). Data from these countries were used as a reference in constructing the negative binomial model. As such, no estimation was done for these countries.

Group 2 countries include 2 high-income, 12 middle-income and 3 low-income countries. Estimated data based on the prediction model described above are provided for these countries.

In the global process, countries/areas with populations of less than 100 000 and thus very low numbers of deaths were also excluded from the modelling process.

Table A.2 gives the 30-day adjusted number of deaths for all countries, and for group 2 countries the modelled number of deaths with a 90% confidence interval. Those without a range are Group 1 countries.

Table 2. Countries in Group 1

Country/area	Income Level
Bahrain	HIC
Kuwait	HIC
Qatar	HIC

HIC = high-income countries

Variables used in the model

Table 3 summarizes the independent variables used in the model and the data source.

Table 3. Independent variables used in modelling process

| Variable	Source of information	Comments
Income (GNI)	World Bank (*4,5*)	
Income level	World Bank (*5*)	Grouping used: low, middle, lower middle, upper middle, high Designation based on 2000–2004 World Bank data, corresponding to year of income level used
Population		Estimated figures from UN Population Division (*3*)
Vehicle density: Number of cars per population	Number of vehicles: The GSRRS survey Population: World Bank (*3*)	
Road density: Total road per land area	2000–2006 World Road Statistics, International Road Federation, 2008 (*9*)	Total road/land km^2 Corresponding or latest year data
Existence of national helmet law	The GSRRS survey	WHO questionnaire
National policies that encourage walking and/or cycling	The GSRRS survey	WHO questionnaire
National policies that support investment in public transport	The GSRRS survey	WHO questionnaire
National speed limits on urban roads	The GSRRS survey	WHO questionnaire
National speed limits on rural roads	The GSRRS survey	WHO questionnaire
Alcohol consumption	World Health Statistics, 2008 (*10*)	Alcohol consumption among adults aged ≥15 years for 2003
Strength of health system	World Health Statistics, 2008 (*10*)	Hospital beds (per 10 000 population)

GSRRS = Global Status Report on Road Safety
GNI = Gross National Income per capita

Estimation method

The third stage used data from Group 1 countries to develop a statistical model to predict road traffic deaths (point estimates) for Group 2 countries including 90% confidence limits. The framework used to predict road traffic mortality was constructed using selected variables (identified through a literature review) which have direct relationship to the outcome variable (road traffic death). These variables were later grouped into three categories: exposure factors (E_j), risk or preventive factors (R_j), and mitigating factors (M_j). Gross national income (I_j) has an influence on the first two categories (Figure 1).

In this framework, the road traffic mortality outcome (Y_j) is a function of a set of independent variables described as exposure factors (E_j), risk or preventive factors (R_j), mitigating factors (M_j) and gross national income (I_j). This can be expressed as follows: $Y_j = f (R_j, M_j, I_j, E_j)$.

The relationship between the outcome and the independent variables is a nonlinear function. The number of deaths (Y_j) is a non-negative integer count data; thus the standard approach (*11*) to be used is the Poisson regression or another form of regression based on the Poisson. The most commonly used regression model for count data treats the response Y as a Poisson variable. In the Poisson regression model, the mean equals the variance, conditional on explanatory variables. In practice, however, this assumption was not satisfied. For this reason, a negative binomial regression model was chosen where the assumption for the dependent variance and Poisson's particular case of negative binomial model are adequately satisfied.

DETERMINANTS OF ROAD TRAFFIC MORTALITY
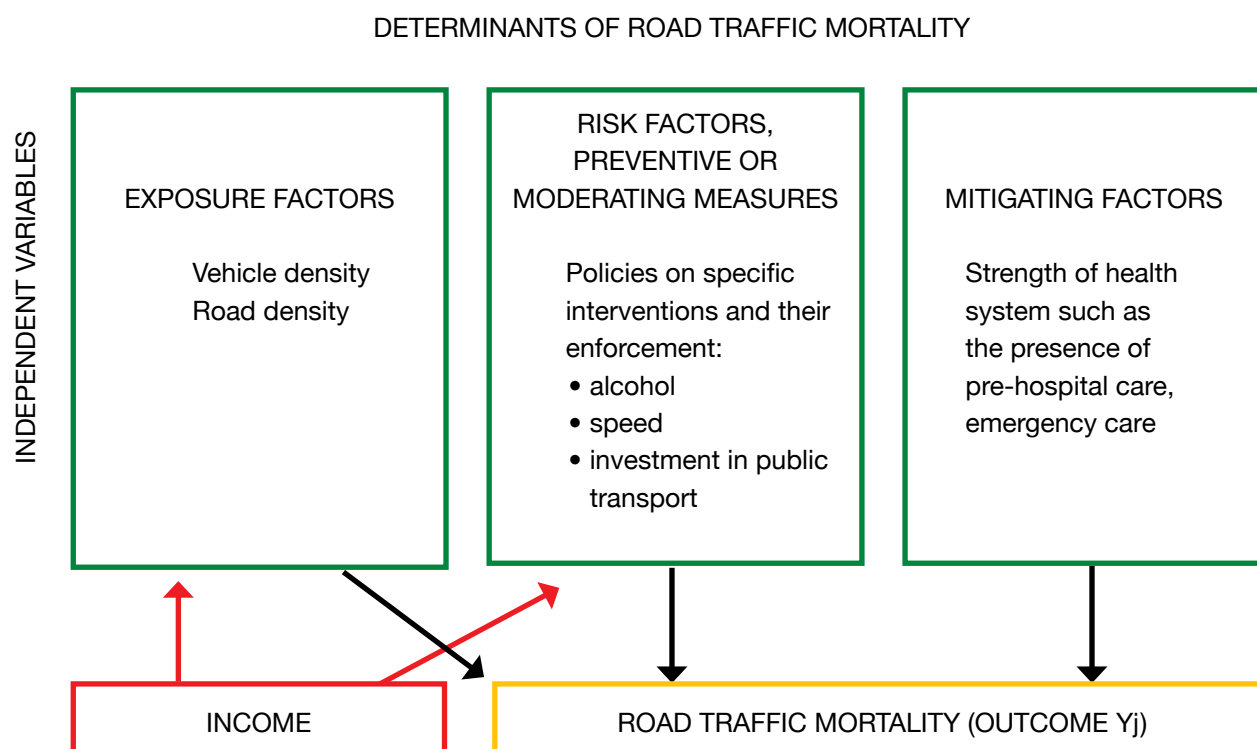
Figure 1. Framework for determinants of road traffic mortality

At the end, a negative binomial regression modelling technique using STATA software (*2*) was used to predict the number of road traffic fatalities with population size as an exposure facture. The model was constructed on the basis of reported data from Group 1 countries.

The full in-depth description of the methodology and formulas for the modelling process are available at the following website www.who.int/violence_injury_prevention/road_traffic/road_safety_status/2009.