Accident Analysis and Prevention 108 (2017) 220–226

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Availability and consistency of health and non-health data for road traffic fatality: Analysis of data from 195 countries, 1985–2013

Helai Huang[a], Qingyi Yin[b], David C. Schwebel[c], Peishan Ning[d], Guoqing Hu[e,*]

[a] Urban Transport Research Center, School of Traffic and Transportation Engineering, Central South University, 22 South Shaoshan Road, Changsha, 410075, China
[b] Design Department of Permanent Way, Station Yard and Terminal, China Railway Design Corporation, Tianjin, 300142, China; Urban Transport Research Center, School of Traffic and Transportation Engineering, Central South University, Changsha, 410075, China
[c] Department of Psychology, College of Arts and Sciences, University of Alabama at Birmingham, 1720 2nd Avenue South, HHB 560, Birmingham, AL, 35294 USA
[d] Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, 110 Xiangya Road, Changsha, 410078, China
[e] Department of Epidemiology and Health Statistics, Xiangya School of Public Health, 110 Xiangya Road, Changsha, China

ARTICLE INFO

Keywords:
Road traffic mortality
Data availability
Consistency
Data quality

ABSTRACT

High-quality data are critical for validly monitoring progress toward global initiatives related to road traffic crash prevention. We assessed the availability and consistency of road traffic mortality data from health and non-health departments in national governments, plus changes in data consistency over time from 1985 to 2013. Using freely accessible data, we systematically assessed availability and consistency of health and non-health data from 1985 to 2013 in 195 countries. Data availability was reflected by the presence of data on motor vehicle mortality rates in that country at any points between 1985 and 2013. The ratio of 'health data divided by non-health data' was calculated to measure the consistency of the two types of data sources. We found that 77 of the 195 countries in the world (39%) had both health and non-health data sources available from 1985 to 2013. None of the 34 low-income countries had both kinds of data sources simultaneously available, while 41 of 55 high-income countries had both kinds of data sources. Of the 71 countries having both kinds of data for five years or more, 33 countries demonstrated high consistency between data sources, 25 countries showed moderate data consistency, and 13 countries displayed low consistency. Jamaica, Mexico, and China had the largest data inconsistencies. 26 of the 71 countries witnessed improved data consistency between 1985 and 2013, but nine experienced decreasing data consistency, in a few cases to a large degree. Efforts are needed to identify reasons leading to data quality changes, and to develop approaches to improve data quality in those nations where it is inadequate.

1. Introduction

Over the past three decades, road traffic crashes have become the world's eighth leading cause of deaths globally, and the leading cause of global death for youth ages 15–29 years (World Health Organization, 2013). It is estimated that about 1.24 million individuals die from road traffic crashes around the world each year (World Health Organization, 2015a). In response to this rising public health challenge, the United Nations launched the Global Plan for the Decade of Action for Road Safety 2011–2020 and included road traffic injury prevention in the Post-2015 Millennium Development Goals, with the goal of halving global road traffic crash casualties by 2020 (Global Road Safety Collaboration, 2011; United Nations General Assembly, 2015).

High-quality data are required to validly monitor progress toward global initiatives related to road traffic injury prevention (Ning et al., 2016). Since 2009, the World Health Organization (WHO) and the Global Burden of Disease (GBD) study group have regularly released road traffic death estimates for more than 180 countries (World Health Organization, 2013; World Health Organization, 2009; World Health Organization, 2015b; Institute for Health Metrics and Evaluation, 2017). Both datasets are based on advanced statistical models and provide valuable updates on the progress of global road traffic injury control. Advanced statistical models, however, cannot remedy the limitations presented by the absence of raw data and poor raw data quality, and thus they may yield biased estimates. For example, alcohol-attributed road traffic death estimates from the GBD 2010 study were significantly lower than reported data in countries reputed to have accurate public health data, including the USA, Canada, UK and New

* Corresponding author.
E-mail addresses: huanghelai@csu.edu.cn (H. Huang), YinQingyi@csu.edu.cn (Q. Yin), schwebel@uab.edu (D.C. Schwebel), ningpeishan@csu.edu.cn (P. Ning), huguoqing009@gmail.com (G. Hu).

http://dx.doi.org/10.1016/j.aap.2017.08.033
Received 3 August 2016; Received in revised form 18 November 2016; Accepted 29 August 2017

Zealand (Hu and Keita, 2013).

Limited research addresses availability, data quality, and changes in data quality related to road traffic fatality, and published research is available only for select countries (Butchart et al., 2001; Loo and Tsui, 2007; Hu et al., 2011; Richard, 2002). No published systematic reviews address road traffic mortality data quality worldwide.

One way to evaluate the quality of data is to examine data consistency between two different data sources at the same time period. Globally, health data (typically from national health departments) and non-health data (typically from national transport departments or traffic police departments) are commonly cited sources of road traffic mortality in both decision-making and scientific research. The present study examines the availability of these two data sources in multiple countries, the consistency between the two types of data, and trends in consistency between the two types of data from 1985 to 2013 for 195 countries worldwide.

2. Methods

2.1. Data sources

Based on *the Global Health Observatory country views* and *the United Nations (UN) report* (World Health Organization, 2017a; UN News Centre, 2013), we included 195 countries in our analyses, including the 193 member states of the UN and two disputed areas (West Bank and Western Sahara). These 195 countries cover 98% of the world's population (complete country list in **Online Supplement 1**) (World Health Organization, 2017a; UN News Centre, 2013).

Health data were accessed through the WHO Mortality Database, which is updated online annually and provides mortality in all member states by country, age, sex and cause (World Health Organization, 2017c). The WHO Mortality Database is a compilation of mortality data reported by Member States from their civil registration systems. According to recommendations by the WHO, mortality rates are not computed for data with very low statistical coverage (World Health Organization, 2017b).

Non-health data for the 51 member countries of the Organization for Economic Cooperation and Development (OECD) were obtained from the online OECD library (Intelligent Transportation Forum (ITF), 2017). Mortality Data in the OECD library are released by the International Transport Forum, which collects transport statistics from police, transport, or other governmental departments of ITF member countries (International Transport Forum, 2014). Because the governmental websites of six OECD countries (Australia, Austria, Azerbaijan, India, Lithuania and Malta) provide road traffic data for longer amounts of time than the ITF online library, we adopted data for these six countries directly from their governmental websites. (Note that mortality data from the governmental sites and from the ITF online library are identical for data available concerning the same years). In addition, we conducted a systematic search to extract road traffic mortality data from countries excluded by the WHO Mortality Database and the OECD library data, or whose data were missing within the study time period. The search was conducted using public search engines (GOOGLE and BAIDU). These supplemental data sources typically were available at the official websites of governmental departments (such as police, public security, or statistical authorities) or mentioned in authoritative reports such as the *Global Status Report on Road Safety* (World Health Organization, 2013; Ning et al., 2016; World Health Organization, 2009). We limited the search to the years 1985 through 2013 and to data presented in English or Chinese.

Both the WHO Mortality Database and the OECD online library data use a 30-day time period after the occurrence of a road traffic crash to calculate road traffic mortality statistics. Based on the transformation coefficients across various definitions used by the WHO (from 1 to 30 days, 1.30; from 3 to 30 days, 1.15; from 6 to 30 days, 1.09; from 7 to 30 days, 1.08) (Ning et al., 2016), we used the standardized time period of 30 days to adjust the impact of operational definition on road traffic mortality.

2.2. Data analysis

We assessed the availability of data using two criteria, the presence of data on motor vehicle mortality rates and the years data were available. Descriptive analysis of data availability was performed by country income. The 195 countries were divided into three groups using World Bank methods: low-income, gross national income (GNI) $<$ \$1046; middle-income (\$1046 $\le$ GNI $<$ \$12,746), and high-income (GNI $\ge$ \$12,747) (World Health Organization, 2013).

The ratio of 'health data divided by non-health data' was calculated to measure the consistency of the two types of data sources. For both kinds of data, we used mean and coefficient of variation (CV) to reflect the average degree of approximation and variation of consistency over time for each country in the covered years. The consistency of the two types of data sources was classified into three categories based on the ratio of 'health data divided by non-health data': high consistency ($0.9 \le$ mean ratio ≤ 1.1), moderate consistency ($0.7 \le$ mean ratio $<$ 0.9 or $1.1 <$ mean ratio ≤ 1.3), and poor consistency (mean ratio < 0.7 or > 1.3). A box plot was graphed to demonstrate variations in mean road traffic mortality ratio of health data divided by non-health data for the 77 countries with two kinds of data sources between 1985 and 2013.

For countries with both kinds of data sources available for at least five years, we further performed simple linear regression to examine time trends in data consistency. The linear regression model used the absolute value of 'ratio of health data divided by non-health data -1' as the dependent variable and year as the independent variable. A positive regression coefficient indicated worsening data consistency, while a negative regression coefficient indicated improving data consistency. Non-significant regression coefficients indicated the lack of substantial linear change. To quantify the variation of the regression coefficient (b), we calculated the 95% confidence interval of b in addition to presenting significance test results for b. Data analysis was performed using MATLAB 2013. "$p < 0.05$" was regarded as statistically significant.

2.3. Ethical issue

This study uses open access data and does not involve the information of individuals. This study was approved by the Medical Ethics Committee of Central South University.

3. Results

3.1. Summary of included countries and data availability

Of the 195 countries, 47 were located in Africa (except for the Middle East), 36 in the Americas, 31 in Asia, 45 in Europe, 22 in the Middle East, and 14 in Oceania. 77 of the 195 countries (39%) had both kinds of data sources available (health and non-health) for at least one year from 1985 to 2013. 105 countries (54%) had only one kind of data source and 13 countries (7%) had neither kind of data source (Fig. 1**A**).

Data availability corresponded somewhat with income status. None of the 34 low-income countries had both kinds of data sources available for the same year, for example, while 41 of the 55 high-income countries (75%) had both kinds of data sources simultaneously available. Similarly, the countries having neither data source were either low-income countries (six) or middle-income countries (seven) (Fig. 1**B, C**). Details on country-specific data availability and the years they cover are presented for all 195 countries in **Online Supplement 1**.

Among the 77 countries with both kinds of data sources, 30 had data for 24 or more years in the study time period of 1985–2013. These included four middle-income countries and 26 high-income countries.