Quantifying the Underestimated Burden of Road Traffic Mortality in Mexico: A Comparison of Three Approaches

Martha Híjar, Aruna Chandran, Ricardo Pérez-Núñez, Jeffrey C. Lunnen, Jorge Martín Rodríguez-Hernández & Adnan A. Hyder

To cite this article: Martha Híjar, Aruna Chandran, Ricardo Pérez-Núñez, Jeffrey C. Lunnen, Jorge Martín Rodríguez-Hernández & Adnan A. Hyder (2012) Quantifying the Underestimated Burden of Road Traffic Mortality in Mexico: A Comparison of Three Approaches, Traffic Injury Prevention, 13:sup1, 5-10, DOI: 10.1080/15389588.2011.631065

To link to this article: http://dx.doi.org/10.1080/15389588.2011.631065
Quantifying the Underestimated Burden of Road Traffic Mortality in Mexico: A Comparison of Three Approaches

MARTHA HÍJAR,1 ARUNA CHANDRAN,2 RICARDO PÉREZ-NÚÑEZ,3 JEFFREY C. LUNNEN,2 JORGE MARTÍN RODRÍGUEZ-HERNÁNDEZ,1 and ADNAN A. HYDER2

1Centro de Investigación en Salud Poblacional del Instituto Nacional de Salud Pública, Cuernavaca, Mexico
2Johns Hopkins International Injury Research Unit, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland
3Centro de Investigación en Sistemas de Salud del Instituto Nacional de Salud Pública, Cuernavaca, Morelos, Mexico

Introduction: We present a novel multistep technique to estimate the actual burden of road traffic mortality in Mexico during the time period 1999 to 2009 by comparing 3 approaches for redistribution of nonspecific (“garbage”) International Classification of Diseases (ICD)-coded deaths.

Methods: Road traffic (RT) mortality data were extracted using a secondary analysis of the Mexican mortality databases for the period 1999 to 2009. In an attempt to correct for underestimation due to inappropriately coded deaths, those deaths assigned to nonspecific codes were redistributed utilizing 3 different adjustment methods. A comparison of the 3 adjustment approaches (proportional, multiple imputation, and regression) is presented. A Poisson regression analysis was utilized to model mortality trends in the raw data and the 3 estimates.

Results: After adjustments, the total number of RT deaths increased by 18 to 45 percent, showing significant underestimation when only the raw data are used. All 3 approaches showed statistically significantly higher RT mortality rates than the crude figures. The proportional approach resulted in the highest RT mortality rate estimate of 23 per 100,000 in 2009 and showed a statistically significant positive increase of 1.5 percent per year across the decade. The 60+ age group and pedestrians had the highest mortality rates of 40 and 10.3 per 100,000, respectively. Over the decade, there was an alarming 332 percent increase in the mortality rate for male motorcyclists.

Conclusion: Though efforts to improve coding should continue to be implemented, we present an additional and often overlooked contribution to the underestimation of road traffic mortality: the ICD nonspecific codes. Improved estimates of road traffic mortality are important in Mexico for policy change and decision making, highlighting the importance of targeting road traffic deaths as a public health problem. The approach presented here may also be useful for estimating the burden of other deaths with similar coding problems.

Keywords Road traffic mortality; Unintentional injury; Mexico; Adjusted disease burden

INTRODUCTION

It is widely recognized that a reliable estimate of the burden of death by cause is essential for informing national and global health priorities (Bhalla et al. 2010). However, a significant proportion of mortality information, particularly from low- and middle-income countries, tends to be misclassified; thus, the burden of deaths from many diseases and injuries is significantly underestimated and therefore considered to be of poor quality (Naghavi et al. 2010). This is a problem that has particularly affected the analysis of mortality due to unintentional injuries, including estimates within the global burden of disease studies to establish global mortality patterns (Bhalla et al. 2010). Bhalla et al. (2010) documented recently that in many countries, over 20 percent of unintentional injury deaths are labeled as an unspecified external cause, using the International Classification of Disease, 10th Revision (ICD-10) code X59, which represents “exposure to an unspecified factor” (Bhalla et al. 2010). Other commonly utilized nonspecific codes include V99 and Y85.9 (unintentional transport injury not otherwise specified; 20 percent of deaths in Georgia and Serbia were recently shown to be allocated to these codes) and V89 (unspecified motor vehicle accident, traffic), which is also utilized frequently in many countries. Categories V87 and V88, other specified personal
Table I  Categories of corrections to reclassify nonspecifically coded road traffic deaths in Mexico, 1999 to 2009

<table>
<thead>
<tr>
<th>Category</th>
<th>Existing code examples</th>
<th>Nonspecific codes</th>
<th>Adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of deaths</td>
<td>Communicable diseases</td>
<td>All R codes</td>
<td>Correction 1</td>
</tr>
<tr>
<td></td>
<td>Noncommunicable diseases</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Injuries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of injury deaths</td>
<td>Intentional injuries</td>
<td>Y32–34</td>
<td>Correction 2</td>
</tr>
<tr>
<td></td>
<td>Unintentional injuries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of unintentional injury deaths</td>
<td>Poisoning</td>
<td>X59</td>
<td>Correction 3</td>
</tr>
<tr>
<td></td>
<td>Smoke</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transport</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of transport injury deaths</td>
<td>Pedestrian</td>
<td>V87, V892, V899</td>
<td>Correction 4</td>
</tr>
<tr>
<td></td>
<td>Motorcycle</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Etc.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

exposures and history, are less frequently but also often used as nonspecific/garbage codes (Bhalla et al. 2010).

Interestingly, Bhalla et al. (2010) reported that Mexico has good information systems to classify road traffic (RT) deaths, although no further details were documented. However, in Mexico, it has been recognized that death certificates are not filled out adequately. Analyzing only one city in 2003, only 12 percent of death certificates were useful for coding deaths using the ICD-10 classification; therefore, reported data actually showed an artificial decrease in RT mortality figures during the years immediately following the introduction of the ICD-10 (Lozano-Ascencio et al. 2003). This phenomenon also occurred in other countries such as Japan, Croatia, Germany, and Romania when the ICD-10 was introduced as the official death coding method (Lozano-Ascencio et al. 2003).

Previous efforts have been undertaken to adjust for the misclassification of deaths (Shahraz et al. 2008). In this study, we have extended the work undertaken by other authors by introducing 3 potential methodologies for the adjustment of road traffic mortality estimates from ICD-10-coded deaths by reallocating those deaths that were labeled with nonspecific codes. This results in adjusted (higher and potentially more accurate) road traffic mortality rate estimates. We also show mortality rates disaggregated by age, sex, and road user type utilizing the proportional adjustment method, highlighting subpopulations at particularly high risk for RT deaths.

For this study, we extracted all RT mortality information for the years 1999 to 2009.

In order to correct for the misclassification of RT deaths the through the utilization of nonspecific codes, adjustments were done to 4 categories of nonspecific codes using three different statistical approaches (Table I). The first category was the “R” codes or deaths of ill-defined or unknown cause; these codes have been referred to as “garbage” codes because they are not useful for public health analysis of cause-of-death data (Naghavi et al. 2010). The second category was codes Y32-34, utilized for deaths of undetermined intent. These codes have been referred to as “garbage” codes because they are not useful for public health analysis of cause-of-death data (Naghavi et al. 2010). The second category was codes Y32-34, utilized for deaths of undetermined intent. The third code was X59, a widely used code for “accidental exposure to other and unspecified factors”; previous analysis has shown that this code was overused in Mexico (Lozano 2003). Finally, the fourth category is codes V87, V892, and V899 for those road traffic deaths in which the victim’s mode of transportation was not specified.

Statistical Adjustments

Three different statistical techniques were utilized for the reallocation of deaths that were categorized using nonspecific codes.

1. Proportional categorization: Nonspecific codes were assigned to categories based on the observed proportion of appropriately coded observations within age–sex groups. This assumes that the deaths in the nonspecific categories occur in the same demographic proportions as the appropriately coded deaths. The age groups used in this analysis were as follows: <10 years, 10 to 19 years, 20 to 34 years, 35 to 59 years, and 60+ years.

2. Multiple imputation of missing values (Little 1992; Little and Rubin 2002; Schafer 1997; Schafer and Graham 2002): Using the “uvis” module (Royston 2004) in Stata 10.1 (2009), missing values (including nonspecific codes) were imputed in the dependent variable based on multiple regression of a dependent variable on a set of covariates, resulting in 10 regression coefficients sets that were then combined with random draws from the conditional distribution of missing observations, given the observed data and covariates. This assumes that those deaths coded with nonspecific values were random and therefore similar to those deaths with appropriate codes. Briefly, the multiple imputation procedure follow

METHODS

Categorization of Deaths

Mexican death certificates are completed mainly by the health sector (by health care providers in hospitals or forensic medical examiners’ offices) and include basic information such as sociodemographic variables and ICD-10 diagnosis codes. Death certificates from autopsies are completed by the forensic medical service; the information is entered into a database at the state level and then transferred to a national database (Secretaría de Salud 2007). The mortality database is then compiled and validated by the Instituto Nacional de Estadística y Geografía (National Institute of Statistics and Geography; INEGI).
these steps: (1) Using a multiple regression model, the vector of coefficients and residual variance are estimated from observations with complete information; (2) a random draw from the posterior distribution of $\beta$ is provided, based on a previously drawn variance; (3) point estimates and variances are calculated for each “missing” value based on the coefficients, their uncertainty, and the covariates; and (4) the results are combined (pooled) to create one overall estimate and standard error (van Buuren et al. 1999).

3. Regression approach: nonspecifically coded data were assigned the code of the highest probability estimated after multiple logistic regression analyses were performed on the appropriately coded data (binary when the dependent variable was bivariate and multinomial when the dependent variable was multivariate). Demographic covariates were inserted into a regression model to identify predictor parameters. The substituted value is that which is best predicted based on the additional demographic variables. This assumes that the group of nonspecifically coded deaths is similar to the group of appropriately coded deaths.

The proportional approach has been traditionally used by global burden of disease studies in order to reassign nonspecific codes for public health analysis purposes (Naghavi et al. 2010). The other 2 approaches rely on the use of additional demographic variables that were available in the Mexican mortality data sets. The available variables utilized for all 3 approaches were age categories, gender, medical insurance status, employment status, level of schooling, marital status, whether medical attention was received, and place of death. In the multiple imputation and regression approaches we also included place of injury as a variable. Other variables that might have been relevant such as place of residence and multiple causes of death (i.e., associated causes) were not included because this information was not available for all of the years under study. These 3 approaches have been previously validated for Mexico by using a random sample of known cases taken from the Mexican mortality database of 2004 (Shahraz et al. 2008). Another proposed method for reallocating nonspecifically coded deaths is a consensus approach by an expert panel (Naghavi et al. 2010); this approach was not employed in this study.

For this analysis, the ICD-10 codes considered as appropriately coded RT deaths were as follows: V01–V04 (0.1, 0.9); V09 (0.2, 0.3, 0.9); V12–V15 (0.3–0.9); V19.4–V19.6, V19.9, V20–V28 (0.3–0.9); V29–V79 (0.4–0.9); V80.3–V80.9, V81.1, V81.9, V82.1, V82.9, V83–V86 (0.0–0.3); V87.0–V87.8, V89.2, V89.9. Although suggested by some authors (Naghavi et al. 2010), we did not include deaths attributed to long-term consequences of RT injuries. Instead, we followed the World Health Organization’s (WHO) recommendation of considering only deaths occurring within the first 30 days after a collision (WHO 2009).

For the calculation of mortality rates, we used mid-year Mexican population estimates from the Consejo Nacional de Población (National Population Council; CONAPO) by sex and age groups from the years under study. All rates are expressed per 100,000 population. To model mortality trends, a Poisson regression analysis was employed using the raw data and the 3 adjusted estimates using bootstrap standard errors with 10,000 iterations.

RESULTS

From 1999 to 2009, there were a total of 5,338,322 registered deaths, of which 402,847 were coded as unintentional injuries. Of those, 169,932 were RT deaths. Between 1 and 2 percent of the RT deaths registered in Mexico during the period under study were coded as “R” (ill-defined/unknown), and in 2 to 3 percent of deaths the intent was not determined. In 19 to 28 percent of unintentional deaths, the external cause of death was not specified, and in 18 to 45 percent of RT deaths the type of road user was not specified.

Over the 10-year period, the proportion of deaths in most of the nonspecific code categories remained the same; notably, the assignment of deaths to codes V87–V892 and V899 increased by 19.3 percent across the decade. Without any adjustment, estimates from the Poisson regression model showed that the crude road traffic mortality rate in Mexico has increased by an average of 1.41 percent each year (95% confidence interval [CI]: 0.7, 2.1) during this period, increasing from 14.38 per 100,000 in 1999 to 16.60 per 100,000 in 2009 (Figure 1). There was a slight but notable dip in the crude mortality rate (to ~14 per 100,000) in 2007 that did not appear in any of the adjustment approaches. This is due to a higher number of nonspecific codes that appeared that year (there was a 32.9%–44.6% RTI underestimation in that year, which is on the higher end of the average, as shown in Table II).

The different adjustment techniques show an underestimation of RT mortality between 18 and 45 percent, depending on the method employed and the year under study (Table II). We compared the relative percentage of underestimation by adjustment technique (Figure 2) for different indicators of RT mortality. For example, the raw data show that approximately 42.16
percent (95% CI: 41.65, 42.67) of unintentional injury deaths are due to road traffic crashes (Figure 2). When the proportional adjustment method is applied, this number increases to 53.68 percent (95% CI: 53.18, 54.18).

In general, all 3 approaches yield similar revised RT mortality figures. Of note, the discrepancy between approaches was highest from 2004 to 2007. This is likely due to the higher numbers and differences in the relative distribution of nonspecific codes that occurred during that time period.

Because the World Health Organization’s global burden of disease studies traditionally use the proportional approach, it provides the best numbers for comparison across studies/estimates. In this analysis, the proportional approach results in slightly higher estimates of mortality and a positive trend that showed an average of 1.15 percent increase in traffic fatalities each year (95% CI: 0.9, 2.1) during this period. This approach showed a maximum rate of 22 per 100,000 in 2009 (Figure 1).

Using the proportional adjustment approach, Figure 3 shows the adjusted road traffic mortality rate disaggregated by age group. By far, the highest mortality burden occurs in the elderly (60+ age group), with a mortality rate of more than 40 per 100,000 population. In 2009, this figure peaked at 46.7 per 100,000 population. This compares to an unadjusted RT mortality in the 60+ age group of 30.7 per 100,000.

Disaggregating the data by sex and road user category showed that the road traffic mortality rate in males is nearly 3-fold higher than in females; this increased rate in males is true for all road user categories (Table III). In addition, the increased mortality burden in pedestrians is striking, constituting the main road user category affected in Mexico, with an average mortality rate of 10.3 per 100,000 population in this time period; male pedestrians had a mortality rate of 16.1 per 100,000 population. This is also higher than the unadjusted male pedestrian mortality rate of 8.0 per 100,000. Worth noting is the substantial increase in mortality in both male motorcyclists and car occupants; the mortality rate increased 332.2 percent in the former and 30.6 percent in the latter during this period.

### Table II

<table>
<thead>
<tr>
<th>Approach</th>
<th>Maximum (%)</th>
<th>Minimum (%)</th>
<th>Average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportional method</td>
<td>44.63</td>
<td>29.46</td>
<td>34.17</td>
</tr>
<tr>
<td>Multiple imputation method</td>
<td>37.87</td>
<td>23.90</td>
<td>30.09</td>
</tr>
<tr>
<td>Regression method</td>
<td>37.88</td>
<td>17.89</td>
<td>27.18</td>
</tr>
</tbody>
</table>

**Figure 2** Adjusted proportional burden of road traffic deaths in Mexico, 1999 to 2009: (A) percentage of total deaths due to injuries; (B) percentage of total injury deaths that are from unintentional injuries; (C) percentage of unintentional injury deaths that are due to transport; and (D) percentage of transport injuries that are road traffic deaths.
DISCUSSION

Comparing multiple adjustment techniques for reallocating nonspecifically coded deaths, we estimated that the road traffic mortality rate in Mexico could be as high as 21.4 to 22.1 per 100,000 population, compared to the rate 16.6 per 100,000 yielded from the raw ICD-10–coded data. Adjustment techniques also highlight what has been shown in previous studies in Mexico; the older age group (60+ years) and pedestrians are particularly vulnerable to road traffic deaths (Hijar et al. 2001; Hijar, Trostle, and Bronfman 2003; Hijar, Vazquez-Vela, and Arreola-Risa 2003).

ICD-coded road traffic deaths in Mexico have a number of quality and specificity problems. Although R codes were not commonly employed and intent was defined in most of the cases (~2% nonspecifically coded), greater efforts are necessary in terms of defining external causes and specifying victims’ modes of transportation. Clearly, there are limitations to any method used to estimate missing data. Therefore, the goal should be to reduce the frequency of the utilization of these codes to less than 20 percent in order to conform to international standards.

This study documented that the already high burden of road traffic deaths in Mexico reported using crude figures are potentially underestimated by up an average of 27.2 to 34.2 percent. This would represent a substantial change in the road traffic mortality burden in the country. Future studies should evaluate regional and city-level variations in coding in addition to any other data collection inconsistencies that result in missing road traffic deaths. By doing so, policy makers could detect and target specific states/municipalities that face particular challenges in terms of filling out death certificates and/or coding mortality causes.

Comparing the 3 adjustment techniques, the proportional adjustment method yielded the highest rates for nearly all years and risk categories. The fact that all adjusted results are statistically different from the unadjusted (crude) data highlights the importance of implementing corrections to improve the quality of existing information systems. In general terms, all 3 approaches are robust; the results that each yields are similar. However, it seems that although multiple imputation and regression approaches take into account more information available from the database (Shahraz et al. 2008), which is preferable, they could be difficult for health policy makers and municipal public health officials to implement and replicate on a regular basis. On the other hand, the proportional method would be easier to be systematized and potentially easier to explain to and be utilized by nontechnical audiences.

Our analysis has several limitations. First, we were limited by the variables available in the Mexican health sector database. The adjustments would all be more robust if additional demographic and crash-specific information had been included. Secondly, we did not attempt to quantify or account for the reasons for particular data being miscoded. It will be important for future studies to analyze which types of deaths are most commonly miscoded and identify ways to correct such problems. Finally, there are inherent limitations to these and all possible methods for identifying and recoding miscoded data. We chose to present these 3 methods because they are among the more commonly employed and intent was defined in most of the cases (~2% nonspecifically coded), greater efforts are necessary in terms of defining external causes and specifying victims’ modes of transportation. Clearly, there are limitations to any method used to estimate missing data. Therefore, the goal should be to reduce the frequency of the utilization of these codes to less than 20 percent in order to conform to international standards.

This study documented that the already high burden of road traffic deaths in Mexico reported using crude figures are potentially underestimated by up an average of 27.2 to 34.2 percent. This would represent a substantial change in the road traffic mortality burden in the country. Future studies should evaluate regional and city-level variations in coding in addition to any other data collection inconsistencies that result in missing road traffic deaths. By doing so, policy makers could detect and target specific states/municipalities that face particular challenges in terms of filling out death certificates and/or coding mortality causes.

Comparing the 3 adjustment techniques, the proportional adjustment method yielded the highest rates for nearly all years and risk categories. The fact that all adjusted results are statistically different from the unadjusted (crude) data highlights the importance of implementing corrections to improve the quality of existing information systems. In general terms, all 3 approaches are robust; the results that each yields are similar. However, it seems that although multiple imputation and regression approaches take into account more information available from the database (Shahraz et al. 2008), which is preferable, they could be difficult for health policy makers and municipal public health officials to implement and replicate on a regular basis. On the other hand, the proportional method would be easier to be systematized and potentially easier to explain to and be utilized by nontechnical audiences.

Our analysis has several limitations. First, we were limited by the variables available in the Mexican health sector database. The adjustments would all be more robust if additional demographic and crash-specific information had been included. Secondly, we did not attempt to quantify or account for the reasons for particular data being miscoded. It will be important for future studies to analyze which types of deaths are most commonly miscoded and identify ways to correct such problems. Finally, there are inherent limitations to these and all possible methods for identifying and recoding miscoded data. We chose to present these 3 methods because they are among the more widely used and robust. Certainly the best solution would be to identify ways to limit the number of nonspecifically coded deaths in the Mexican database.

The mortality rate estimated for 2006 (20.8 per 100,000) is actually quite similar to that which is officially reported in the regional report the status of the road safety in the Americas (Organización Panamericana de la Salud 2009). However, in that study, Mexican authorities stated that RT deaths were only those that occurred at the scene of a crash; for that reason these figures were multiplied by an “adjustment factor” of 1.30 taken from what had been observed in previous studies (WHO 2009). This was an assumption made in the consensus

<table>
<thead>
<tr>
<th>Road user category</th>
<th>Male Unadjusted</th>
<th>Male Adjusted</th>
<th>Female Unadjusted</th>
<th>Female Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pedestrians</td>
<td>8.0</td>
<td>16.1</td>
<td>2.3</td>
<td>4.5</td>
</tr>
<tr>
<td>Bicyclists</td>
<td>0.5</td>
<td>0.6</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Motorcyclists</td>
<td>0.6</td>
<td>1.3</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Car passengers</td>
<td>5.3</td>
<td>10.6</td>
<td>1.6</td>
<td>3.1</td>
</tr>
<tr>
<td>Other</td>
<td>9.5</td>
<td>3.3</td>
<td>2.6</td>
<td>0.9</td>
</tr>
<tr>
<td>All road traffic injuries</td>
<td>23.8</td>
<td>31.9</td>
<td>6.5</td>
<td>8.7</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

We thank Dr. David M. Bishai for his suggestions and comments on this analysis and Dr. Rafael Lozano for his advice on the implementation of the proportional approach. We also thank Daniel Vera-López for his invaluable help in processing data and Aarón Salinas-Rodríguez for his statistical advice. This work was conducted as part of the Road Safety in 10 Countries project funded by the Bloomberg Philanthropies.

REFERENCES


Stata 10.1 [computer program], College Station, Tex: Stata Press; 2009.
