



Errors in accident data, its types, causes and methods of rectification-analysis of the literature



Ashar Ahmed*, Ahmad Farhan Mohd Sadullah, Ahmad Shukri Yahya

School of Civil Engineering, Universiti Sains Malaysia, Pulau Pinang, Malaysia

ARTICLE INFO

Keywords:

Under-reporting
Road traffic injury
Capture-recapture
Crash mapping
Accident reporting form
Accident database

ABSTRACT

Most of the decisions taken to improve road safety are based on accident data, which makes it the back bone of any country's road safety system. Errors in this data will lead to misidentification of black spots and hazardous road segments, projection of false estimates pertinent to accidents and fatality rates, and detection of wrong parameters responsible for accident occurrence, thereby making the entire road safety exercise ineffective. Its extent varies from country to country depending upon various factors. Knowing the type of error in the accident data and the factors causing it enables the application of the correct method for its rectification. Therefore there is a need for a systematic literature review that addresses the topic at a global level. This paper fulfils the above research gap by providing a synthesis of literature for the different types of errors found in the accident data of 46 countries across the six regions of the world. The errors are classified and discussed with respect to each type and analysed with respect to income level; assessment with regard to the magnitude for each type is provided; followed by the different causes that result in their occurrence, and the various methods used to address each type of error. Among high-income countries the extent of error in reporting slight, severe, non-fatal and fatal injury accidents varied between 39–82%, 16–52%, 12–84%, and 0–31% respectively. For middle-income countries the error for the same categories varied between 93–98%, 32.5–96%, 34–99% and 0.5–89.5% respectively. The only four studies available for low-income countries showed that the error in reporting non-fatal and fatal accidents varied between 69–80% and 0–61% respectively. The logistic relation of error in accident data reporting, dichotomised at 50%, indicated that as the income level of a country increases the probability of having less error in accident data also increases. Average error in recording information related to the variables in the categories of location, victim's information, vehicle's information, and environment was 27%, 37%, 16% and 19% respectively. Among the causes identified for errors in accident data reporting, Policing System was found to be the most important. Overall 26 causes of errors in accident data were discussed out of which 12 were related to reporting and 14 were related to recording. "Capture-Recapture" was the most widely used method among the 11 different

methods: that can be used for the rectification of under-reporting. There were 12 studies pertinent to the rectification of accident location and almost all of them utilised a Geographical Information System (GIS) platform coupled with a matching algorithm to estimate the correct location. It is recommended that the policing system should be reformed and public awareness should be created to help reduce errors in accident data.

1. Introduction

When an accident occurs, it is either reported by the victim or recorded by the person in-charge from an institution/organization responsible for its monitoring. The cluster of these records is known as the accident data. Most of the decisions taken to improve road safety are based on this data, which makes it the back bone of any country's road safety system. The current decade has been declared as the "Decade of Action for Road Safety (2011–2020)" (Sleet et al., 2011) by the United

Nations, and the target is to reduce the death toll by 50% in low and middle-income countries (Bliss and Breen, 2012). To monitor the progress in this regard the first pre-requisite is high-quality reliable data, without which any observed changes would be uninterpretable (Chokotho et al., 2013). Errors in the accident data will make the patterns of accident more difficult to discern and effect of interventions less precisely known (Hauer and Hakkert, 1988), ultimately compromising the safety of the citizens commuting through the road. These errors lead to misidentification of black spots and hazardous road

* Corresponding author.

E-mail addresses: ashar.ue17@gmail.com, aahmed@neduet.edu.pk (A. Ahmed), cefrhn@usm.my (A.F.M. Sadullah), shukriyahaya@gmail.com (A.S. Yahya).

<http://dx.doi.org/10.1016/j.aap.2017.07.018>

Received 31 August 2016; Received in revised form 28 February 2017; Accepted 12 July 2017
Available online 29 July 2017

0001-4575/ © 2017 Elsevier Ltd. All rights reserved.

segments, projection of false estimates pertinent to accidents and fatality rates, and detection of wrong parameters responsible for accident occurrence, thereby making the entire road safety exercise ineffective. The inaccuracy of road accident injury data affects the prioritization of funding and resources; setting targets for interventions in areas of higher risk at the aggregate and disaggregate levels such as locations (urban/rural), age group, road user type, and specific injury severity; calculation of road accident injury cost; cost/benefit ratios for evaluation of interventions aimed at reducing road accident injuries (Watson et al., 2015); and inferences made on the effect of road safety measures on the injury severity risk (Abay, 2015).

Before proceeding with the any kind of analysis it is extremely important to know the type of error in the accident data and the factors causing it so that the correct solution could be identified for the exact problem. Any mistake occurring at the micro level, such as the source from where the data is being recorded and transmitted, cannot be addressed by making improvements at the macro level, such as advanced database management system. The circumstances vary from country to country. Different regions of the world have different kinds of errors based on their socio-economic, demographic and political situation. Similarly the methods used to rectify the errors in the accident data of one country might not be applicable to the other. Given the above argument, the cause behind a particular type of error defines the method to be used for its rectification. But there has been no work done to encapsulate the literature on the topic methodically.

Considering the importance of the issue and the lack of a systematic literature review that addresses it, there is an urgent need of a study that summarizes the discrepancies in accident data from different countries in all the six regions of the world (WHO, 2016). This paper fulfils the above research gap by providing an evidence based review of the different types of errors found in the accident data of 46 countries, their causes and the methods used for their rectification. The errors are classified and discussed with respect to each type and analysed with respect to geographic region and income level; assessment with regard to the magnitude for each type is provided; followed by the causes that result in their occurrence, and the methods used to address each type of error.

2. Errors in accident data

In the past, errors in accident data have not been the subject of extensive research; however, scientists have paid attention to this important area of road safety in the last few decades. As documented by Hauer and Hakkert (1988), one of the first formal reports related to errors in accident data came from the United States of America in 1966. Since then several reports and articles were published that covered different aspects of inaccuracies in accident data. Maximum number of studies came from Europe, followed by other regions of the world and the most recent were countries from Africa. Most of the researches related to errors in accident data are from High income countries followed by middle income and the least are from low income countries, as shown in Table 1. This indicates that the reliability of accident data in low income countries is questionable as the efforts put into the validity of road accident data are miniscule. The least number of reports are from the region of Africa, which are only three, and they are also very recent. The reviews of the studies from different countries across the world imply that discrepancies exist in the accident database of all countries because such data can never be completely free from errors. However the extent varies from country to county depending upon the legislation, method of accident data recording, process of transferring data from accident site to the centralized database, available resources, political stability, public awareness and overall Policing system.

3. Types of errors

The errors in accident data can be broadly classified into two types.

Table 1

Name of countries/territories in each region with errors in accident data and the year the study got published.

Region	Country	Income Level*	Year Published
Americas	USA	High	1966, 1971, 1977, 1983, 1999, 2003, 2004, 2005, 2007, 2009, 2013
	Canada	High	1977, 1983, 2009
	Nicaragua	Middle	2004
Europe	UK	High	1973, 1995, 1999, 2009, 2010, 2014, 2015
	Netherland	High	1984, 1999, 2010, 2014
	Germany	High	1984, 1999
	La Réunion-France	High	1999
	Denmark	High	1999, 2014, 2015
	Slovenia	High	2001, 2014
	France	High	1999, 2006, 2007, 2010, 2014
	Switzerland	High	1999
	Norway	High	1999
	Sweden	High	1999
	Greece	High	2009, 2010, 2014
	Czech Republic	High	2010, 2014
	Hungary	Middle	2010, 2014
	Spain	High	2010, 2014
	Italy	High	2013
Russia	High	2013	
Portugal	High	2015	
Croatia	High	2016	
Ireland	High	2016	
Western Pacific	Australia	High	1979, 1992, 1993, 1994, 2001, 2006, 2015
	Hong Kong-China	Middle	1985, 2006
	New Zealand	High	2001, 2012, 2013
	Japan	High	2001
	China	Middle	2003, 2009, 2011, 2013
	Cambodia	Low	2005
	Laos	Middle	2005
	Philippines	Middle	2005
	Singapore	High	2005
	Vietnam	Middle	2005, 2011
	Brunei	Middle	2005
	Malaysia	Middle	2007, 2012, 2013, 2014, 2015
	Korea	High	2015
Eastern Mediterranean	Saudi Arabia	High	1996, 2003
	Pakistan	Middle	1998, 2010, 2011
	UAE	High	2004, 2012
South-East Asia	Sri Lanka	Middle	1984, 2013
	India	Middle	1985, 2008, 2015
	Indonesia	Middle	2005
	Myanmar	Low	2005
	Thailand	Middle	2005
Africa	Malawi	Low	2012
	South Africa	Middle	2013
	Ethiopia	Low	2014

*As per WHO (2015).

One is the error in reporting and the other is the error in recording, as shown in Fig. 1. The error in accident data reporting results into faulty accident rates pertinent to a city, locality, territory or country. They occur due to under-reporting or over-reporting of accidents. The error in accident data recording results into ambiguities related to factors responsible for the occurrence of an accident, such as the road, the environment, the driver, the vehicle, the location, the control of the facility where the accident occurred and classification of injury severity. They occur due to incomplete or inaccurate recording of information for the accidents. Different studies have examined different aspects of errors in reporting and recording. The errors in reporting are

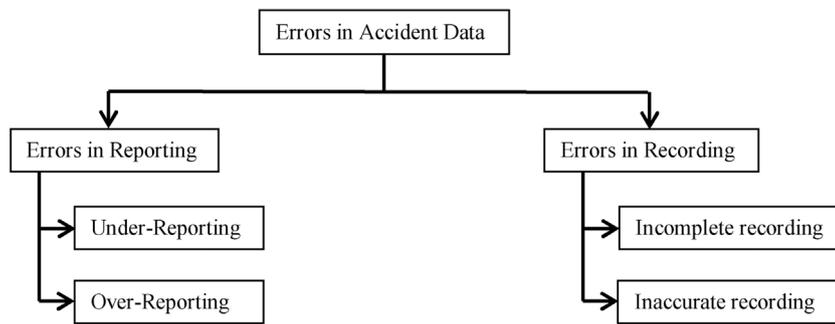


Fig. 1. Types of errors in accident data.

measured with respect to type of accident which is commonly classified into four categories which are fatal, severe injury, slight injury (Janstrup et al., 2014) and property damage only (PDO) (Blincoe et al., 2002). Other categories include non-fatal (Nakahara and Wakai, 2001), very slight injury (Elvik and Mysen, 1999), mild, moderate and critical injury accidents (Rosman, 2001). The errors in recording are measured with respect to location of accident, victims' information, vehicles' information and the information pertinent to the road and environment. The location of accident could be in the form of geographic coordinates (Miler et al., 2016), local road map coordinates (Imprialou et al., 2014; Qin et al., 2013), route number (Burns et al., 2014), link-node (Imprialou et al., 2015), district (Austin, 1995a), and name of road/street/highway (Razzak and Luby, 1998; Dutta et al., 2007). The victims' information include number of victims, their gender, age, injury severity, type (driver/occupant/pillion/pedestrian), safety belt usage, helmet usage, alcohol and drug use (Howard et al., 1979; Austin, 1995b; Razzak and Luby, 1998; Li et al., 1999; Giles, 2001; Chokotho et al., 2013). The vehicles' information include number of vehicles, their registration number, type, make, colour, extent of damage and trajectory (Howard et al., 1979; Shinar et al., 1983; Razzak and Luby, 1998; Nguyen et al., 2011). The road and environment is a broad category which contain information related to date of accident, time of accident, speed limit, controls upon road, intersection type, road grade, curvature (vertical/horizontal), road features, road surface type (seal/unseal), road surface condition (wet/dry), obstruction, foreign substance on road, type of accident, cause of accident, weather, lighting and traffic conditions (Howard et al., 1979; Shinar et al., 1983; Nguyen et al., 2011).

3.1. Errors in accident data reporting

Various indices to measure road safety such as number of fatalities per 100,000 population, number of fatalities per billion vehicle kilometres, and number of fatalities per 10,000 registered vehicles (IRTAD, 2015), depend upon the proportion of the total number of accidents reported and the accuracy with which this proportion is known (Hauer and Hakkert, 1988). Error in reporting, either from the authorities responsible for accident data collection or from the general public, who is required to report an accident as per the country's legislation, results into wrong estimation of road safety indices. The socio-economic conditions of a particular country also affect the extent of reporting of each type of accident (Abegaz et al., 2014). Therefore, an analysis of error in accident data with respect to income level of a country was conducted. The classification of each country in a particular income group is based on WHO (2015). The statistics present the type of injury more prone to errors and provide a chronological comparison among various countries along with the magnitude of each study.

3.1.1. Status in high-income countries

The high-income countries for which the status of error in accident data reporting is available are mostly from the regions of Americas, Europe and Western Pacific. The common classification of accidents

prevalent in these studies is slight, severe and fatal, while some have classified as non-fatal and all injury accidents. It was documented by Smith (1966, cited by Hauer and Hakkert, 1988) that there was no error in reporting of fatal injury accidents in California, USA, while very little error existed in the reporting of injury accidents, which was 7%, however the property damage only accidents were under-reported by 62%. A study by Scott and Carroll (1971) indicated that 65% of all injury accidents do not get reported while slight injury and property damage only are least reported. In Canada, the situation was also similar to USA, where the in-patients and out-patients were under-reported by 3.4% and 24.4% respectively, while all cases of fatal injury accidents were recorded by the police (Bourbeau and Laberge-Nadeau, 1977). Among the European countries the initial accounts of error in accident data were documented by Bull and Roberts (1973), who claimed that the slight and severe injury accidents in UK were 67% and 16% under-reported. There have been cases in which the accident data was found to be over-reported. One such unique case of over-reporting by the police came from Australia in the Western Pacific region, and was documented by O'Connor (1992). In his study the numbers of admissions in the hospital due to road injuries were compared with that of Police reported admissions. A comparison was made between different states and it was found that only in the state of Victoria the numbers of admissions were 9.5% over-reported in the Police data.

The status of error in accident data among high-income countries is graphically presented in Fig. 2. The level of error related to slight, severe and fatal injuries is arranged in a chronological order. The two studies from USA showed a decrease in the error pertinent to slight injury accidents from 67% in the early 1970s (Scott and Carroll, 1971) to 39% in the mid 2000s (Sciortino et al., 2005), however, the former estimated the overall error for various states across the country while the later was limited to San Francisco County. The average level of error in the reporting of slight injury accidents among various countries was 50% in late 1980s (Hauer and Hakkert, 1988) which increased to 75% in late 1990s (Elvik and Mysen, 1999). It was found that the actual number of accidents were higher than those reported by the Police in Slovenia (Simončič, 2001). In Australia the level of error was reported to be 69% (Rosman, 2001). Among the European countries, there has not been any improvement over the period of time. In France and New Zealand the level of error was 75% (Amoros et al., 2007) and 59% (Wilson et al., 2012) respectively. In Denmark the level of error was 82% in 2014 (Janstrup et al., 2014) which reduced to 78% in 2015 (Abay, 2015). The most recent study was from Ireland where the value was 42.5% (Short and Caulfield, 2016). From Fig. 2, it can be observed that the level of error in accident data related to severe injury is less than that of slight injury. The maximum value was 52%, reported for Denmark (Janstrup et al., 2014), which subsequently reduced to 28% next year (Abay, 2015). Ireland had the least level of error which was 8.5% (Short and Caulfield, 2016). In earlier studies upto late 1980s (Bull and Roberts, 1973; Maas and Harris, 1984; Hauer and Hakkert, 1988), the level of error among high-income countries ranged between 16%–20% which subsequently increased to 30%–40% from the year 1992 to 2012 (O'Connor, 1992; Elvik and Mysen, 1999; Rosman, 2001,

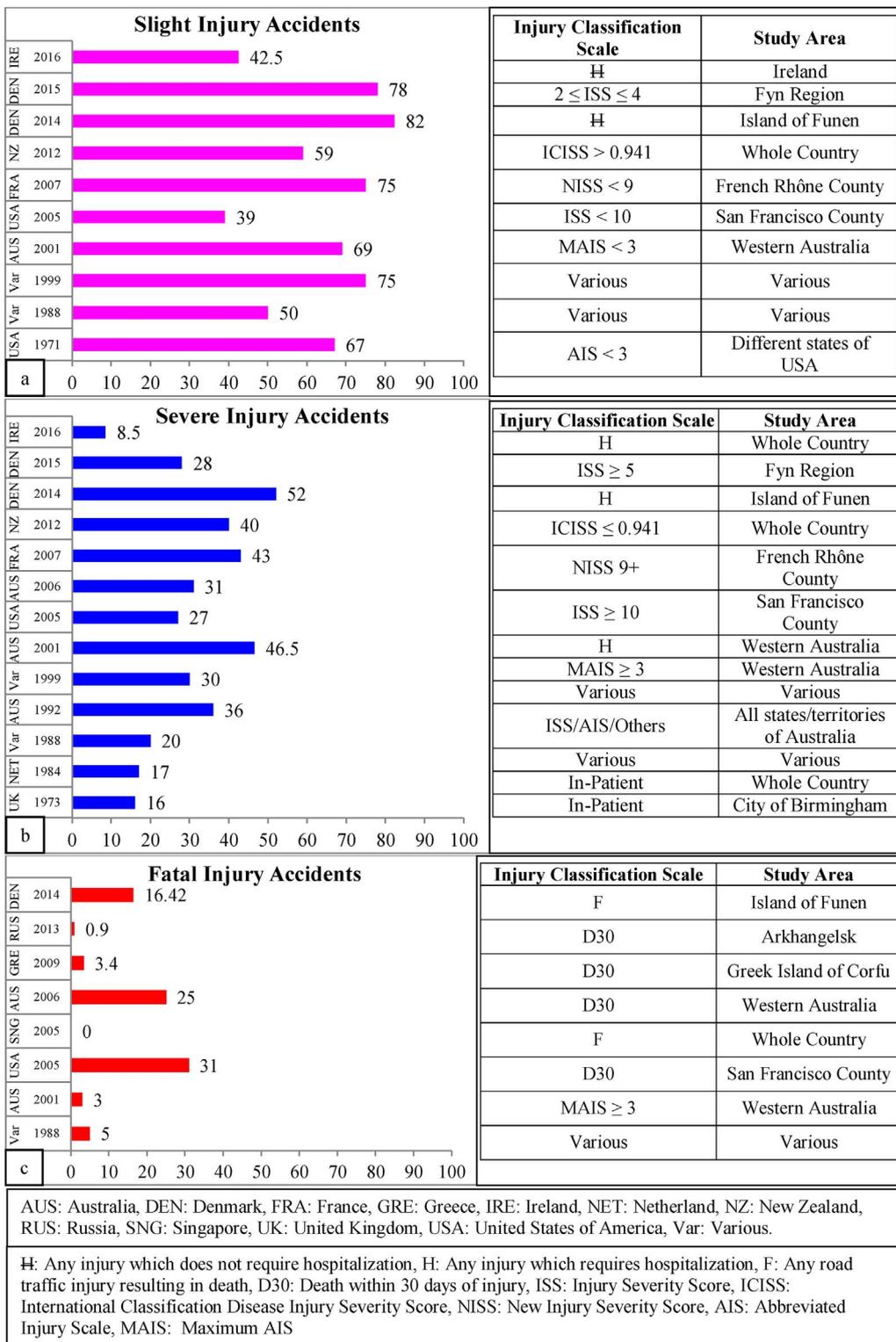


Fig. 2. (a) Percentage error in slight injury accident data reporting among high-income countries. (b) Percentage error in severe injury accident data reporting among high-income countries. (c) Percentage error in fatal injury accident data reporting among high-income countries.

Sciortino et al., 2005; Meuleners et al., 2006; Amoros et al., 2007; Wilson et al., 2012). Fatal injury accidents were reported with highest accuracy as the level of error was 0% in Singapore (Sigua and Palmiano, 2005), 0.9% in Russia (Kudryavtsev et al., 2013), and 3.4% in Greece (Petridou et al., 2009). For fatal injury accidents in San Francisco, USA, the level of error was 31% but the study was limited to pedestrians only (Sciortino et al., 2005). In Australia, the level of error in the road fatality data was reported to be 3% by Rosman (2001). This increased to 25%, as reported by Meuleners et al. (2006), but the former study

covered all types of accidents while the later was limited to accidents involving heavy vehicles only. In Denmark the value obtained after averaging the under-reporting level of fatal injury accident data for pedestrian, cyclists, moped, motorcyclist, car, bus and other victims of road crashes, was 16.42% (Janstrup et al., 2014). In the literature there was only one country, called the Slovak Republic, for which all types of accidents, including property damage only, were stated to be 100% reported (Derriks and Mak, 2007).

There were some studies which classified the data at the aggregate

Table 2
Error in non-fatal and all injury accident data reporting among high-income countries.

Country	Author(s) (Year of Publication)	Non-Fatal Injury	All Injury	Injury Classification Scale	Study Area
USA	Scott & Carroll (1971)		65%	Abbreviated Injury Scale (AIS); 1-Minor, 2-Moderate, 3-Severe(Not life-threatening), 4-Severe(Life-threatening), 5-Critical(Survival Uncertain)	Various States of USA
Australia	Rosman and Knuiiman (1994)		36%	Police Definition for each type of Injury and Abbreviated Injury Scale (AIS)	Western Australia
France	Aptel et al. (1999)	75%		Injury Severity Score; ISS < 4 = less severe injury; ISS ≥ 10 = Severe Injury	La Réunion
New Zealand	Alsop and Langley (2001)	37%		Abbreviated Injury Scale (AIS); AIS < 5 = Non-Severe Injury; AIS ≥ 5 = Severe Injury	Whole Country
Japan	Nakahara and Wakai (2001)		14%	Fatal and Non-Fatal traffic injury	Whole Country
USA	Sciortino et al. (2005)		21.7%	Injury Severity Score; ISS ≥ 10 = Severe Injury; ISS < 10 = Complaint of Pain, Visible Injury	San Francisco County
Singapore	Sigua and Palmiano (2005)	12%		Any road traffic injury resulting/not-resulting in death	Whole Country
France	Amoros et al. (2006)	62.3%		New Injury Severity Score	French Rhône County
Greece	Petridou et al. (2009)	84%		Injury Severity Score; ISS ≥ 5 = Severe Injury; 30-day Rule for fatality	Greek Island of Corfu
New Zealand	Tin et al. (2013)		93.5%	All kinds of injury	Whole Country
Australia	Watson et al. (2015)		66.66%	Abbreviated Injury Scale (AIS); 1-Minor, 2-Moderate, 3-Serious, 4-Severe, 5-Critical, 6-Maximum	State of Queensland
Portugal	Ferreira et al. (2015)	29%		Length of Hospital Stay, Maximum Abbreviated Injury Scale	Porto Metropolitan Area

level as non-fatal injury accidents (Aptel et al., 1999; Alsop and Langley, 2001; Sigua and Palmiano, 2005; Amoros et al., 2006; Petridou et al., 2009; Ferreira et al., 2015) and all injury accidents (Scott and Carroll, 1971; Rosman and Knuiiman, 1994; Nakahara and Wakai, 2001; Sciortino et al., 2005; Tin et al., 2013, Watson et al., 2015) as shown in Table 2. The former does not include injuries resulting in death while the later may include them as well. For non-fatal injury accidents the maximum error was found to be 84% in the Greek data (Petridou et al., 2009), while the minimum was found to be 12% in the Singaporean data (Sigua and Palmiano, 2005). The average error in non-fatal injury accident data among high-income countries was observed to be 50%. In the case of all injury accidents the maximum error was found to be 93.5% in the New Zealand data (Tin et al., 2013), while the minimum was found to be 14% in the Japanese data (Nakahara and Wakai, 2001). The average error in all injury accident data was observed to be 49%. This is very near to the average of non-fatal injury accidents. It shows that even among high-income countries, with latest technologies available for accident record keeping and database management, the errors are prevalent.

3.1.2. Status in middle-income countries

The middle-income countries for which the literature is available are mostly from the regions of South-East Asia and Western Pacific while one each is from Americas, Europe, Eastern Mediterranean and Africa. Majority of the studies pertinent to these countries have classified the accidents into fatal and non-fatal. The probable reason could be the unavailability of data segregated with respect to injury type such as slight and severe. Studies that highlighted the issue of error in accident data and reported it quantitatively among high-income countries started from early 1970s. For middle-income countries, such studies started in late 1990s. The oldest study reported was for Sri Lanka by Sayer and Hitchcock (1984, cited by Elvik and Mysen, 1999). They observed that the police data under-reported road traffic accidents by 28% as compared to hospital records. Severe injury accidents were found to under-reported by 96% in Pakistan (Razzak and Luby, 1998) and 32.5% in Vietnam (Nguyen et al., 2011). In India, one study stated the slight and severe accidents to be under-reported by 97.7% and 82.8% respectively (Dandona et al., 2008), while a subsequent study showed them to be under-reported by 93% and 78% respectively (Mohan et al., 2015). Furthermore no specific parameter was stated as erroneous in the Indian accident database, but it was discovered that discrepancies exist in accident statistics between National Crime Bureau and different City Traffic Police organizations (IRC, 2012). The first account of errors pertinent to accident data in Malaysia was documented by Ng et al. (2007), who stated that under-reporting of injuries is common. The issue was further investigated by Abdul Manan and Várhelyi (2012) in which a comparison of the RTI statistics was made between Sweden and Malaysia. It was found that in general the ratio of deaths to serious to minor injury accidents was 1:9:57 in Sweden while it was 1:1.4:4 in Malaysia, and in particular for motorcyclists the ratio of deaths to serious to minor injury accidents was 1:7:17 in Sweden (SIKA, 2009) while it was 1:1.4:2.6 in Malaysia (PDRM, 2008). The result is further supported when compared to the figures in the United States of America (Wilmot et al., 2005), where the ratio of fatal to injury only to property damage only crashes was 1:20:400 as calculated from Blincoe et al. (2002). Moreover, in an international study the officially reported road deaths in Malaysia by the Royal Malaysian Police were 6915 in 2013, while the estimated deaths were 7129 by WHO (2015).

A comparison between the errors in non-fatal and fatal injury accident data are presented in Fig. 3. For the purpose of graphical representation, the maximum of the values amongst slight and severe injury accidents were also plotted in the non-fatal injury category. The only two countries that had error less than 50% in their non-fatal injury accident data were Vietnam (Sigua and Palmiano, 2005; Nguyen et al., 2011) and Sri Lanka (Periyasamy et al., 2013). Hong Kong, which is

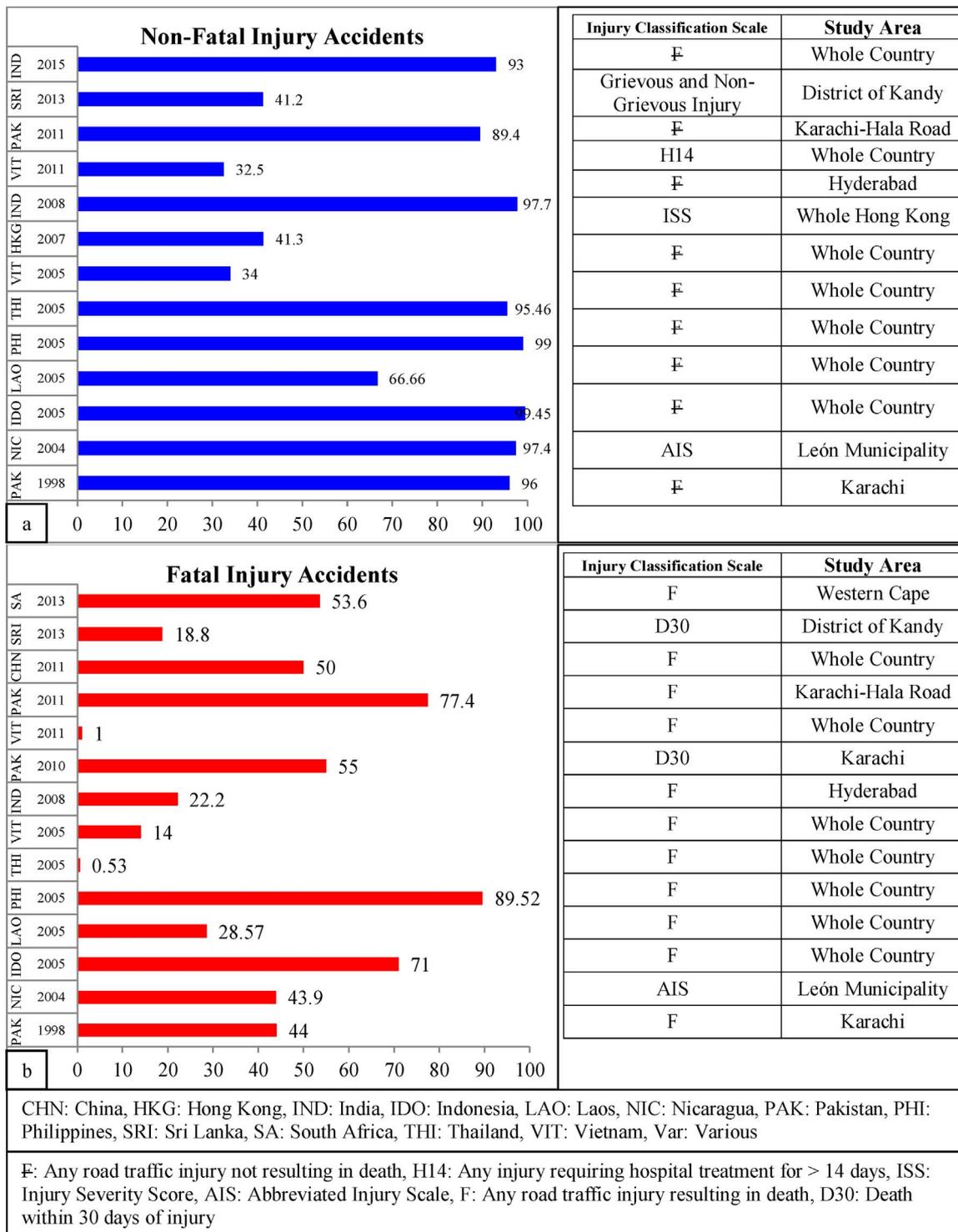


Fig. 3. (a) Percentage error in non-fatal injury accident data reporting among middle-income countries. (b) Percentage error in fatal injury accident data reporting among middle-income countries.

now a special administrative region of China, also had error less than 50% (Loo and Tsui, 2007) in the non-fatal injury accident data. An error of 66.66% was observed in Laos (Sigua and Palmiano, 2005). In two countries, which are Pakistan and India, very little improvement was observed in terms of reduction in error in the non-fatal injury accident data over the period of time. In Pakistan the level of error was 96% (Razzak and Luby, 1998) which reduced to 89.4% (Bhatti et al., 2011) while in India the level of error was 97.7% (Dandona et al., 2008) which reduced to 93% (Mohan et al., 2015). Majority of the middle-

income countries, like Nicaragua (Tercero and Andersson, 2004), Indonesia, Philippines and Thailand (Sigua and Palmiano, 2005), had substantial error in their non-fatal injury accident data ranging between 95%–99%.

The error in fatal injury accident data among middle-income countries was substantially higher than high-income countries. The maximum observed value was 89.52%, as shown in Fig. 3(b), while it was 31% for high-income countries as shown in Fig. 2(c). This indicates that the system of recording and maintaining accident data for road

fatalities among middle-income countries is quite poor. Among the 11 countries, for which the status is shown in Fig. 3(b), five countries had error greater than or equal to 50%. They include Indonesia (Sigua and Palmiano, 2005), Philippines (Sigua and Palmiano, 2005), Pakistan (Muhammad Usman Lateef, 2010, Bhatti et al., 2011), China (Hu et al., 2011) and South Africa (Chokotho et al., 2013). The rest of the studies had error ranging between 18%–44% except Thailand and Vietnam. They include Nicaragua (Tercero and Andersson, 2004), Laos (Sigua and Palmiano, 2005), India (Dandona et al., 2008), and Sri Lanka (Periyasamy et al., 2013). An increasing trend was observed for Pakistan where the error was initially reported to be 44% (Razzak and Luby, 1998) which increased to 55% (Muhammad Usman Lateef, 2010) and then 77.4% (Bhatti et al., 2011). In China, the reported road deaths in official statistics were found to be 40% less than the actual number estimated by the Beijing Traffic Engineering Research Institute (Kareem, 2003). A decreasing trend was observed for Vietnam where the error was initially reported to be 14% (Sigua and Palmiano, 2005) which decreased to 1% (Nguyen et al., 2011). The minimum observed error was 0.53% and 1% for Thailand (Sigua and Palmiano, 2005) and Vietnam (Nguyen et al., 2011) respectively.

3.1.3. Status in low-income countries

There are very few low-income countries for which the statuses of error in accident data are available. They are from the regions of Africa, South-East Asia and Western Pacific. As compared to middle-income countries the situation in low-income countries is slightly better in terms of error in accident data. The maximum observed error in non-fatal injury accident data was 80%. This is lower than the maximum value observed for the same category for middle-income countries, which was 99.45%. It is also lower than the maximum value for error in slight injury accident data observed for high-income countries, which was 82%. The minimum error observed in non-fatal injury accident data was for Cambodia (Sigua and Palmiano, 2005), as shown in Table 3. Despite being a low-income country, Myanmar had 0% error in fatal-injury accident data, while none amongst the middle-income countries had 0% error. The maximum error observed in fatal-injury accident data was for Malawi (Samuel et al., 2012), which is still less than the error level in Indonesia, Philippines, and Pakistan, who are middle-income countries. The most recent study from any low income African country was from Ethiopia (Abegaz et al., 2014), where the level of error in non-fatal and fatal injury accident data was found to be 76% and 41% respectively. This is lower than the error observed in many middle-income countries such as China, India, Indonesia, Nicaragua, Pakistan, Philippines, South Africa and Thailand.

3.1.4. Comparative analysis

Odds ratio was estimated for the percentage of error in accident data, due to under-reporting, with respect to the income level of the country. The response variable, which is the percentage of error, was dichotomized by categorizing it as less than 50% and greater than or equal to 50% while the predictor variable, which is the income level of the country, was classified into two groups; high income and middle and low income. A cross classification of the predictor and the response variables is presented in Tables 4 and 5. The high income countries were taken as the reference category against which the chance of having high percentage of error in accident data was measured for the

Table 4
Cross classification of error of non-fatal accidents dichotomized at 50%.

Under-Reporting Error in Non-Fatal Accident Data	Income Level		Total
	Low & Middle (1)	High (0)	
> = 50% (1)	12	7	19
< 50% (0)	4	12	16
Total	16	19	35

Note: The number in brackets represent the code given to the particular category.

Table 5
Cross classification of error of fatal accidents dichotomized at 50%.

Under-Reporting Error in Fatal Accident Data	Income Level		Total
	Low & Middle (1)	High (0)	
> = 50% (1)	7	0	7
< 50% (0)	12	9	21
Total	19	9	28

Note: The number in brackets represent the code given to the particular category.

low and middle income countries. Therefore, high income countries were coded as “0” while low and middle income countries were coded as “1”. Similarly the error in accident data less than 50% was coded as “0” while greater than or equal to 50% was coded as “1”. The values in Table 4 shows that the low and middle countries which had error > = 50% and < 50% were 12 and 4 respectively. The high income countries which had error > = 50% and < 50% were 7 and 12 respectively. All together there were a total of 35 datapoints used in cross classification analysis of non-fatal accidents. Likewise, the values in Table 3 shows that the low and middle countries which had error > = 50% and < 50% were 7 and 12 respectively. The high income countries which had error > = 50% and < 50% were 0 and 9 respectively. All together there were a total of 28 datapoints used in cross classification analysis of fatal accidents.

The odds ratio for “Error in Non-Fatal Accident Data” obtained by dividing the error values for low and middle income countries with that of high income countries, which is (12/4)/(7/12), was found to be 5.14. This shows that if the income level of the country is high then the chance of having less error (that is < 50%) in accident data pertinent to Non-Fatal accidents would be 5 times more as compared to low and middle income countries. The model summary shows that the predictor variable was found to be significant at 95% confidence interval (Standard error: 0.748) and the Nagelkerke R-Square was 0.187. The odds ratio for “Error in Fatal Accident Data” could not be calculated because the relationship obtained by dividing the error values for low and middle income countries with that of high income countries, which is (7/12)/(0/9), is undefined. However, it was noted that among all high income countries, there was not a single one in which the error in data pertinent to fatal accidents was > = 50%. Furthermore, all high income countries had a very well maintained database for fatal accidents with average error of 0.1% only. Since, the error in accident data has been dichotomized at 50%, therefore the logistic relationship between the income level of a country and its corresponding probability to have error < 50% can be represented in the form of an S-curve as

Table 3
Error in non-fatal and fatal injury accident data reporting among low-income countries.

Country	Author(s) (Year of Publication)	Non-Fatal Injury	Fatal Injury	Injury Classification Scale	Study Area
Cambodia	Sigua and Palmiano (2005)	69%	19%	Any road traffic injury resulting/not-resulting in death	Whole Country
Myanmar	Sigua and Palmiano (2005)	80%	0%	Any road traffic injury resulting/not-resulting in death	Whole Country
Malawi	Samuel et al. (2012)	–	61%	Any road traffic injury resulting in death	Lilongwe District
Ethiopia	Abegaz et al. (2014)	76.3%	40.85%	Any road traffic injury resulting/not-resulting in death	Addis Ababa-Hawassa Highway

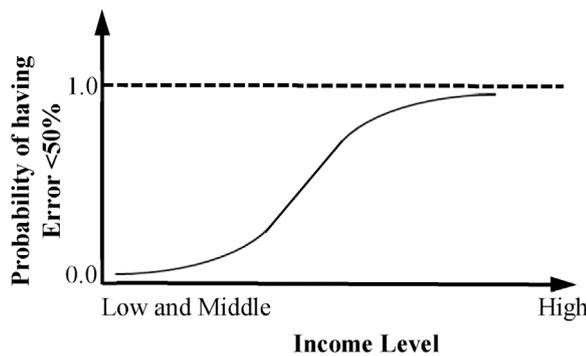


Fig. 4. The logistic relationship between error in reporting non-fatal accidents and income level of a country.

shown in Fig. 4.

The logistic relationship shown in Fig. 4 indicates that as the income level of a country increases the probability of having less error in accident data also increases. Hence, for low and middle income countries the chance of having less error in accident data is low. There are several causes for the high error in the accident data pertinent to low and middle income countries. They have been explained in detail in Section 4.

The overall analysis of all four types of accidents, which are fatal, severe injury, slight injury, and property damage only; indicated that the extent of error varies with the level of accident severity. The statistics are presented in Table 6. The highest number of data points available are for fatal injury accidents and then for severe and slight injury. This is in line with previous studies (Hauer and Hakkert, 1988; Elvik and Mysen, 1999; Derriks and Mak, 2007). For slight injury accident data, the maximum error found was for India (Dandona et al., 2008) while the minimum error found was for USA (Sciortino et al., 2005). For severe injury accident data, the maximum and minimum error found were for Pakistan (Razzak and Luby, 1998) and Ireland (Short and Caulfield, 2016) respectively. For fatal injury accident data, the maximum error found was for Philippines while the minimum error found was for Brunei, Myanmar and Singapore (Sigua and Palmiano, 2005). There was only one study which mentioned the average level of reporting for property damage only accidents (Elvik and Mysen, 1999). From it, the error in data due to under-reporting was calculated to be 75%, which is higher than the mean error found in the accident data for severe, slight and fatal injury accidents. Linear regression analysis indicated that as severity increases the under-reporting decreases. Hence, it was established that the extent of under-reporting is directly related to the level of accident severity. The relationship is represented in the form of a ‘Severity versus Under-Reporting Diagram’ as shown in Fig. 5. Maximum under-reporting is for the accident with the least severity, which is property damage only, because in many countries these accidents do not have to be reported (Elvik and Mysen, 1999). Minimum under-reporting is for the accident with the highest severity, which is fatal injury, because even in a low income country like Myanmar, these accidents are 100% reported (Sigua and Palmiano, 2005).

Table 6
Descriptive statistics of percentage error with respect to injury severity.

	Slight Injury	Severe Injury	Fatal Injury
Mean	69	40	28
Standard Deviation	18	24	26
Maximum	98	96	89.5
Minimum	39	8.5	0
n	12	17	28

Based on the findings of Elvik and Mysen (1999) the error in data due to under-reporting of property damage only accidents was calculated to be 75%.

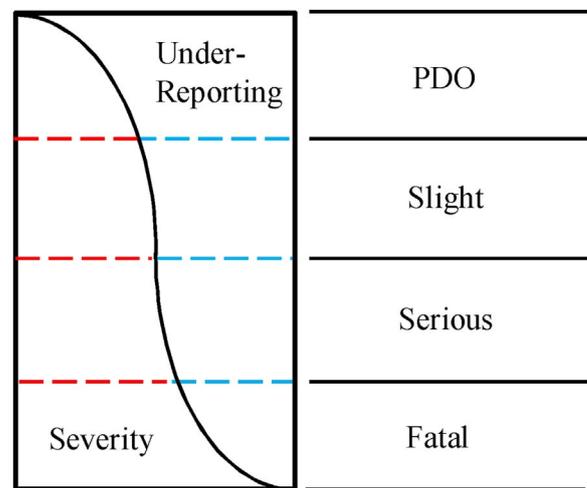


Fig. 5. Severity versus Under-Reporting Diagram.

3.2. Errors in accident data recording

Error in accident data recording results into ambiguities related to all factors responsible for the occurrence of an accident, such as the road, the environment, the driver, the vehicle, the location and the control of the facility where the accident occurred. In-correct recording of victim’s injury severity, status of driving under influence, seating type and all other related information also comes under the banner of erroneous data recording. Different studies have examined different aspects of it. Most of them were from high-income countries. To the best of author’s knowledge, no study has investigated the status of error in accident data recording in a low-income country.

3.2.1. Status in high-income countries

Documentation of errors related to accident data recording had generally been done by the countries in the American, European and Western Pacific regions, as most high-income countries lie in them. There were only three studies among high-income countries which covered all the four categories of error in recording. All three of them were conducted about three decades ago. The first one was by Howard et al. (1979), in which a comparison was made between the accidents reports filled by a survey team with the reports prepared by the police in South Australia. A total of 23 items were identified that had discrepancies in recording. Out of them seven were related to the vehicle; which were type, make, colour, registration number, number of vehicles, vehicle movement 1 and vehicle movement 2, three were related to the victim/driver; which were sex, severity and pedestrian movement, while the rest were related to the environment; which were time of accident, police attendance, speed limit, road features, road grade, road surface (seal/unseal), road condition (wet/dry), controls upon road, controls erected, type of accident, weather, lighting and traffic conditions. The average calculated for the error of all items in each category is presented in Table 7. Their findings were similar to a previous study conducted in Indiana, USA, by Shinar and Treat (1977) in which it was stated that the informativeness of the police reports were practically nil with respect to driver and vehicle characteristics while very little information was provided for the vehicle defects, environmental or road deficiencies, and presence of different human conditions and states.

The second study which covered all the four categories of error in recording was by Shinar et al. (1983). Unlike Howard et al. (1979), they found that the police data were the most reliable for location, which was erroneous for only 0.8% of the cases, and five other variables which were date, day of week, number of drivers, passengers and vehicles. However, the police data was found to be least reliable for the items in the other three categories such as road grade, its horizontal character

Table 7
Error in accident data recording among high-income countries.

Country	Author(s) (Year of Publication)	Error in Recording			
		Location	Victim's Information	Vehicle's Information	Environment
Australia	Howard et al. (1979)	26%	20%	8%	12%
USA	Shinar et al. (1983)	0.8%	21%	15%	18%
Various	Hauer and Hakkert (1988)	19%	7.30%	7.90%	8.87%
UK	Austin (1995a)	10%			
UK	Austin (1995b)		9.4%		
USA (Hawaii)	Li et al. (1999)		9%		
Australia	Giles (2001)		33%		
USA	Farmer (2003)		49%		
USA	Schiff and Cummings (2004)		17.55%		
USA	Dutta et al. (2007)	40.5%			
Greece	Petridou et al. (2009)		20.3%		
Czech Republic	Broughton et al. (2010)		68%		
France			1%		
Greece			98%		
Netherlands			59%		
Spain			65%		
UK			73%		
UAE	Hawas et al. (2012)	16.5%			
Italy	Montella et al. (2013)	36%			
USA	Qin et al. (2013)	28.28%			
UK	Imprialou et al. (2014)	7.3%			
Canada	Burns et al. (2014)	50%			
UK	Imprialou et al. (2015)	26.6%			
Croatia	Miler et al. (2016)	33.5%			
Average		24.54%	38.8%	11.5%	15%

and surface composition, speed limit, vehicle model, driver age and injury severity. The highest error was in the implication of view obstructions, found to be 97%. The average level of error in items related to victim's information, vehicle's information and environmental/roadway factors was calculated to be 21%, 15% and 18% respectively, as presented in Table 5. A comparative analysis of error found in the accident data of various countries, all high income, was done by Hauer and Hakkert (1988). From their study, the average error related to each category was calculated. Similar to the findings of Howard et al. (1979), the highest level of error was observed in accident location. The average error in the other three categories was less than the ones calculated for Howard et al. (1979) and Shinar et al. (1983).

Most studies that reported error in recording victim's information focused primarily on the mis-classification of victim's injuries. In this area one of the earliest studies done in UK was by Austin (1995b), who reported that the level of error due to omissions and misidentifications in the police reported accident data was 0.4% for gender, 12.2% for severity and 15.7% for age. Upon comparing the values calculated from Giles (2001) with Howard et al. (1979), the level of error related to victim's information in Australia was found to increase over the period of time. Unlike Australia, in USA, the level of error related to victim's information varied over the period of time, as evident from the values presented in Table 7. Among European countries, considerable difference was observed in error related to victim's information, as the least value was found to be 1%, for France, and the highest value was 98%, for Greece (Broughton et al., 2010). It is to be noted that level of error reported for Greece was 20.3% by Petridou et al. (2009) and 98% by Broughton et al. (2010). Although both studies were related to injury severity, however the reason behind the difference in level of error could be the method of estimation used by each.

In road safety studies targeted towards the improvement of road-related features, location errors were identified as the most important (PIARC, 2003). It is so for the reason that each accident is considered as a geographic event because it is tied to a unique location (Thill, 2000). Furthermore the correctness of accident location is necessary to identify any hazardous segments in a network so that effective engineering countermeasures could be designed (Karlaftis and Golias, 2002) and

spatial distribution of safety risk across the network could be analysed for the introduction of targeted and specialized crash avoidance measures (Steenberghen et al., 2004). In Honolulu, Hawaii; it was found that the quality of information related to accident location within police crash reports is poor, because they were generally referenced in terms of distance to or from an intersection, mile markers, nearest entry/exit ramp, street names/addresses, and other reference points such as landmarks or locally known geographical features (Kim and Levine, 1996; Levine and Kim, 1998). There was no formal system for accident location referencing in Abu Dhabi, UAE, before January 1st 2003 and it was acknowledged as the fundamental deficiency in their accident records. Due to this the place of accident was often described with the help of a landmark or a road name, and if the road was very long, there was no adequate way to estimate where along the road the accident actually occurred (Khan et al., 2004). Al-Ghamdi (1996) pointed out that the Police records in Saudi Arabia contain incomplete, unclear and incorrect information, the most important among whom is the inaccurate accident location (Al-Ghamdi, 2003).

The earlier studies pertinent to mistakes in accident location only checked whether the road name has been mentioned in the accident reports or not, but the advent of GPS marked a new era in determining the extent of error in this category. Although commercial GPS devices, such as "TomTom GO" became common in 2004 (Corporate.tomtom.com, 2016) but still in countries like USA, they were not installed in every police vehicle (Dutta et al., 2007). When the use of GPS did become common, other problems arise such as the coordinates being located outside the city boundaries or in places where there was no road network. It was mentioned in a research by Montella et al. (2013) that in Italy, 36% of crashes that occurred in rural area had missing linear reference and the accuracy level of accident location in all crashes was only upto 1 km. Addressing the same problem, it was reported by Qin et al. (2013) that 2.75% of the accidents in Wisconsin did not had complete location information, 14.54% had missing critical location information and 28.28% of the total state wide crashes could not be mapped uniquely. In UK, in spite of having Ordnance Survey Grid map coordinates, the error in accident location was reported to be 7.3%, after the application of the algorithm used for obtaining precise

Table 8
Error in accident data recording among middle-income countries.

Country	Author(s) (Year of Publication)	Error in Recording			
		Location	Victim's Information	Vehicle's Information	Environment
Pakistan	Razzak and Luby (1998)	0%	10%	1.45%	
Hong Kong-China	Loo (2006)	27.50%			
Hong Kong-China	Tsui et al. 2009		19.60%		
Hungary	Broughton et al. (2010)		43%		
Vietnam	Nguyen et al. (2011)		52.50%	47.50%	38.75%
China	Wang et al. (2013)	88.10%			
South Africa	Chokotho et al. (2013)		58%		
Average		38.53%	36.62%	24.48%	

accident location, by Imprialou et al. (2014). Furthermore the names of the roads were not available for all minor road crashes in the city of London and Greater London, which is approximately 26% of the crash database (Imprialou et al., 2015). Half of the crash records lacked co-ordinates in the municipality of Amos, Quebec, Canada (Burns et al., 2014), while in Croatia, the co-ordinates which represent the true location of an accident were found to be incorrect in 33.5% of the cases (Miler et al., 2016). These studies are a proof that despite the state of the art technology available in the high-income and highly industrialized countries the issue of location error still exists.

3.2.2. Status in middle-income countries

The number of studies which presented an account of the status of error in accident recording among middle-income countries were few. To the best of author's knowledge no study has been done to determine the level of error in recording location, vehicular or environmental attributes of accidents in low-income countries. As mentioned in Table 5, the category with the highest average error among high-income countries is victim's information, followed by accident location. But in middle-income countries it is vice versa, as shown in Table 8. There was no study among middle-income countries which has reported the level of error for all the four categories related to accident data recording. Furthermore, only one research documented the level of error for the category of environment; therefore, its average could not be calculated.

The primary difference in the status of location error between high-income countries and low-income countries is the way it is measured. In most high-income countries the difference between the actual and the reported co-ordinates of accident locations is reported as the level of error. In most middle-income countries the difference between the actual and the reported names of road on which the accident occurred is considered as the level of error. Hence, Razzak and Luby (1998) reported 0% error in recording accident location in Karachi, Pakistan. Loo (2006) reported 27.5% error in Hong Kong. Similar issue related to location errors were also stated in two studies pertinent to unsignalized intersections in Malaysia, in which it was pointed out that the Police data do not always refer to the exact location and instead they provide some indication of a typical pattern about the occurrence of the accident within the vicinity of a particular intersection (Abdul Manan and Várhelyi, 2015; Abdul Manan, 2014). Wang et al. (2013) stated that 88.1% of the crashes in Harbin, China could not be located to a point or an area. In another study pertinent to China, discrepancy in recording accident location along with other fields such as intersection type, road surface conditions and incomplete filling of accident reports was stated but not quantified (Wang et al., 2011). On an average the accident locations in middle-income countries are 38.53% erroneous. This value is 24.54% for high-income countries, as mentioned in Table 5.

Most studies which documented victim's information reported the error due to mis-classification of injury severity (Tsui et al., 2009; Broughton et al., 2010; Nguyen et al., 2011). Mistakes in the classification may include overestimation as well as underestimation of injury

severity. Other discrepancies in this category involved name of the victim (Razzak and Luby, 1998), age and identity numbers not being recorded (Chokotho et al., 2013). On an average these errors accounted for 36.62% of the victim's information available in the accident data of middle-income countries. Surprisingly, it is lower than the average error among in high-income countries. This also implicates that irrespective of the income level, mis-classification of injury severity by the Police is common in most countries around the world.

Mistakes in recording vehicle's information were related to the registration number not being known (Razzak and Luby, 1998) or the extent of damage not being assessed correctly (Nguyen et al., 2011). In comparison, errors in recording vehicle's information in high-income countries were related to the model year, type, colour and maneuver (Howard et al., 1979; Shinar et al., 1983; Hauer and Hakkert, 1988). The average error for this category was found to be low for high-income countries in relation to middle-income countries. But the studies carried out in this context are very old and there are possibilities that other types of errors might have been introduced in the accident data of high-income countries over the period of time. Similar to the category of vehicle's information, studies in high-income countries related to the category of road and environment were also primitive. The average error for this category in the accident data of high-income countries was 15%. This is lower than 38.75% (Nguyen et al., 2011), as calculated from the only study available for middle-income countries. But it was related to the cause of accident and traffic density only. Apart from these two, other factors like road grade, its horizontal character and surface composition, controls upon road, controls erected, speed limit, accident type, light and weather conditions were also evaluated in the studies from high-income countries (Howard et al., 1979; Shinar et al., 1983; Hauer and Hakkert, 1988).

3.2.3. Comparative analysis

The highest level of error was observed in the recording of the data related to the victim's information. It is followed by the level of error in recording accident location and environment, while the category least prone to error is the vehicle's information. The descriptive statistics of all four categories is presented in Table 9. Among them its only victim's information which is related to human beings whiles the rest deal with non-living physical objects. This increases the possibility of human

Table 9
Descriptive statistics of percentage error in recording with respect to its sub-types.

	Location	Victim's Information	Vehicle's Information	Environment
Mean	27	37	16	19
Minimum	0	1	1.45	8.87
Maximum	88.1	98	47.5	38.75
Standard Deviation	21.34	26.53	16.33	11.64
n	15	14	5	4

error induced both from the recorder’s as well as the victim’s side. Furthermore, it is easy to judge the type and extent of vehicle’s damage as compared to the type and severity of victims’ injuries. After victim’s information, accident location was the category most prone to errors. The high level of error in recording accident location is very harmful in terms of its effect on the safety of the entire road network. The consequences of incorrect recording of accident location or mis-identification of a certain roadway facility such as; stop controlled versus priority controlled intersections, curve versus straight sections, and divided versus undivided carriageways, would lead to the prediction of false black spots or non-hazardous road sections. Due to its importance, the number of studies which reported error in accident location was the highest. They were followed by the number of studies presenting an account of the victim’s information. The least number of studies were those which documented the level of error related to vehicular and environmental attributes. The most recent of them was from Vietnam, which is a middle income country. A study done by *Ljung Aust et al. (2012)*, in Norway, stated that investigators under-report information related to the cause of accident. They investigate from a legislative point of view, focusing more on reporting the factors that prove the guilt of a particular party involved, rather than presenting the complete scenario of the location and the road environment. This creates a bias in the reporting. Therefore, there is a need for studies among high-income countries, for the evaluation of the accuracy in recording vehicular and environmental attributes in accident data.

4. Causes of errors

4.1. Causes of errors in accident data reporting

The primary cause behind errors in accident data reporting is the “Policing System”. This system covers everything from the occurrence of an individual accident to the dissemination of national accident statistics. The structure of a typical policing system is shown in *Fig. 6*. Irrespective of the socioeconomic and geographic conditions, it is responsible for most errors occurring in the accident data of several countries around the world, as presented in *Fig. 7*. The system is based on passive recording, that is, the accidents had to be generally reported to the police, and hence by their nature the police statistics are incomplete (*Maas and Harris, 1984*). The error initiates from the advent

of an individual accident. Not all accidents are required to be reported, as the requirements vary with respect to the extent of damage (*Scott and Carroll, 1971*) and type of accident such as driver only injured and no other vehicle involved (*Bull and Roberts, 1973*). Even if the police are informed it is not necessary that they will always come to the scene as it depends upon their priorities with respect to other matters which demand their immediate attention. In Western Australia, the accidents attended by the police account for only 25% of the total reported accidents (*Rosman, 2001*). Similarly in Netherlands, police is present in only 25% of all road accidents and 45% of all accidents which involved injuries, and in Sweden their presence is reliant on the distance between a police station and the accident site (*Maas and Harris, 1984, Derriks and Mak, 2007*). Since in urban areas police stations are more abundant as compared to rural locations, therefore, this variation in landuse affects the extent of accidents recorded by the police. In India, the accidents are under-reported by 10–20% in urban areas and 50% in rural areas (*Mohan et al., 2015*). Also in Australia, the accidents in remote and rural areas are more prone to under-reporting as compared to urban areas (*Watson et al., 2015*). The proximity of the local police stations in Denmark affect the reporting propensity of the drivers (*Abay, 2015*).

The third level in the policing system where the error gets introduced is the step where the accident is recorded by the officer. Arrival of the police at the accident scene is not always followed by a formal registration. Hence all accidents which are attended by the police are not recorded because the persons involved preferred to coordinate the accidents themselves (*Derriks and Mak, 2007*). Other reasons include increased time pressure on police (*Alsop and Langley, 2001*), reluctance of police to record non-fatal traffic injury accidents and not following the patient for 30 days (*Bhatti et al., 2011*). In low-income countries like Malawi, police reports are more likely filed when the accidents involved more injuries, deaths and property damage (*Samuel et al., 2012*). The flaw in the policing system of most low and middle-income countries, enables the perpetrator to bribe and get away and the victim to receive no attention or even got harassed in case of reporting (*Bhatti and Salmi, 2012*). When the accident is finally recorded by the police, the next step is to digitize the manually recorded information. At this stage the errors get introduced due to reasons such as the person who fills the form at the accident site is not the same who enters the data into a data base (*Derriks and Mak, 2007*), or the information being lost

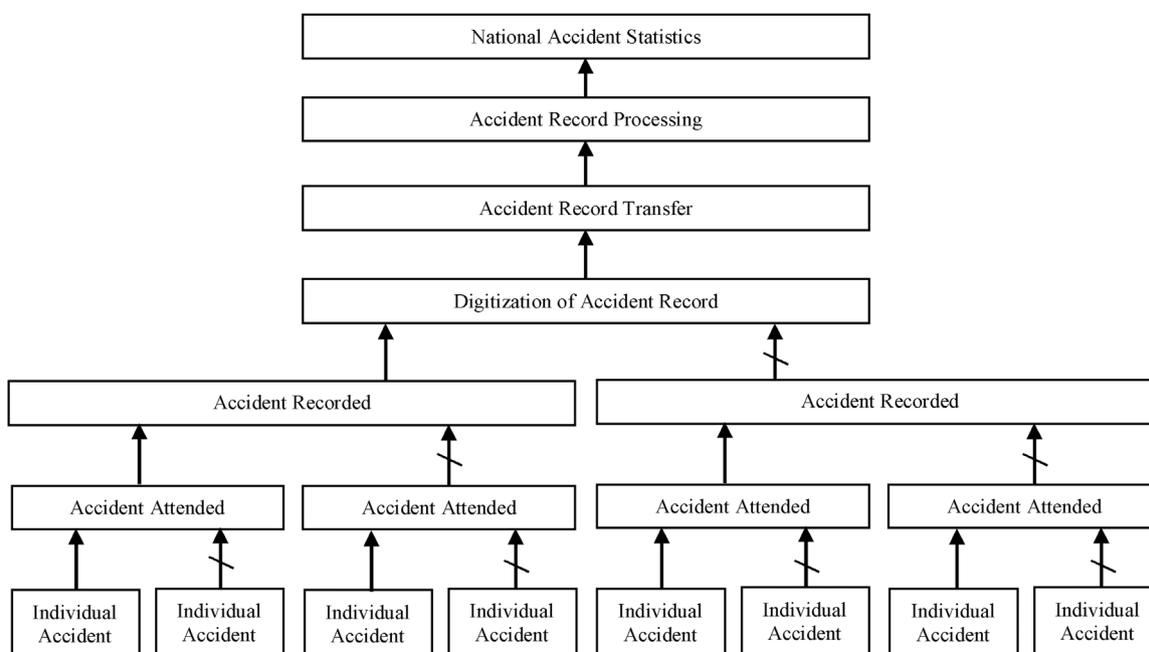


Fig. 6. The Typical Policing System (Legend: \rightarrow action not performed, \rightarrow action performed).

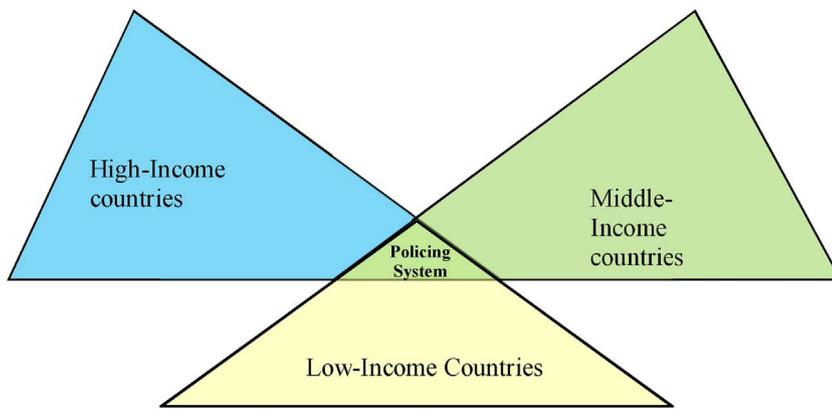


Fig. 7. The common cause of error in accident data reporting with respect to income level of country.

or altered during transfer through telephone to a call centre where it is entered into the police database (Short and Caulfield, 2016).

The raw accident data transferred to the central database is required to be processed to obtain different demographic and socio-economic indices. They could also be segregated with respect to injury severity, accident type, and roadway facility. Processing errors might result in wrong statistics pertinent to individual index. Sometimes bad maintenance of computers results into loss of accident records (Nguyen et al., 2011). The indices calculated after processing raw accident data accumulated from the whole country, formulates the national accident statistics, which is the last step in the overall policing system. The inclusion criteria arbitrarily applied by the police officers to include an accident in the traffic accident statistics results into erroneous accident reporting (Nakahara and Wakai, 2001). Since, the performance of police is judged on the basis of crime figures which include road traffic fatalities, therefore, they could be under-reported (Bhatti and Salmi, 2012). This phenomenon is also termed as “achievement-illness” due to which the authorities intentionally reduce the road accident figures, especially related to slight injury and property damage only accidents, to show their progress (Nguyen et al., 2011).

The road user introduces the rest of the errors in accident data reporting. One of the oldest reasons for not reporting an accident by the road users is the ignorance of law related to the accidents which are required to be reported to the police (Howard et al., 1979). Sometimes people intentionally do not comply with the rules regarding reporting of accidents to the authorities (Howard et al., 1979; Giles, 2001; Meuleners et al., 2006) and are not willing to call the police (Aptel et al., 1999). In other cases they are reluctant to summon the police (Sciortino et al., 2005) especially pedestrians (Amoros et al., 2006) and younger road users (Wilson et al., 2012). Teenagers don't often report because of lack of awareness about the need to report or their fears of being rebuked by their parents (Loo and Tsui, 2007). A similar case is of accidents related to cyclists aged less than 15 that do not involve any motor-vehicle may not be seen by parents as necessary to report to police (Watson et al., 2015).

The victims who don't get seriously injured don't report to the police (Amoros et al., 2006). Apart from non-serious injuries, victims who involved single vehicle accidents in New Zealand were less likely to report to the police (Wilson et al., 2012). Similar to New Zealand, in Denmark also the drivers who involved single vehicle crashes were less likely to report to the police, along with other factors such as peak hour, women and mid-aged driver (Abay, 2015). Sometimes the parties involved in the accident write an agreement for insurance purposes. Another reason was none of the victims thought of calling the police (Amoros et al., 2006), which is a unique cause of error in accident data reporting.

The issue of settlement between the parties without calling the police is a common cause of error in the accident data of most low and middle income countries. The known countries with this issue are India

(Dandona et al., 2008), Pakistan (Bhatti et al., 2011) and Ethiopia (Abegaz et al., 2014). In Vietnam, the common reasons behind error in accident data reporting are hiding the number of non-fatal and non-serious injury accidents, not sharing the data, lack of systemization in collection and storage, and inaccessibility (Nguyen et al., 2011). The six reasons for not reporting an accident to the police in India were: not interested, no other party to complain about, own mistake, settlement between the parties, not necessary, and hit and run case (Dandona et al., 2008). It was further found by Mohan et al. (2015) that the reasons for under-reporting of non-fatal crashes in India have little to do with injury severity, for example; extent of property damage, involvement of multiple parties. Similarly in Malawi, accidents with less property damage, deaths and injuries were less likely to be reported to the Police (Samuel et al., 2012). The 12 causes behind error in reporting of accidents in Bangladesh were: thinking police will take bribe, poverty, driver given warning, unskilled driver, driver was known, accident to save child, hit by domestic animal, minor injury, driver not responsible, self-accident, the driver ran way, and locally compromised (Ahmed et al., 2014). Similarly in Ethiopia, the victims who belonged to poor socioeconomic background negotiate with perpetrators and settle issue without police involvement. Furthermore, they go to the hospital without informing police (Abegaz et al., 2014).

The ongoing conflict in the Eastern Mediterranean region is another reason for erroneous accident data reporting. This region mostly comprises of low and middle-income countries such as: Afghanistan, Iraq, Libya, Pakistan, Palestine, Somalia, and Yemen, and the political cum militia conflicts influence the quality of the road fatalities reported thus producing errors in their accident data (Bhatti and Salmi, 2012).

In a retrospect the common causes behind error in accident data reporting are:

- Policing system
- Distance to the police station
- Patients not followed for 30 days
- Ignorance of law
- Limited legal requirements for reporting an accident
- Intentional non-compliance by particular road users
- Reluctance of people to summon the police and file a report
- An agreement between the parties
- No one seriously injured
- No one thought of calling police
- Hit and run case
- Political and militia conflicts

4.2. Causes of errors in accident data recording

In most countries around the world accident data is recorded by the Police. Mistakes in recording are inevitable because the Police have many duties to perform at the accident site (Austin, 1995a). They

record information on standard forms which are commonly known as the ‘Accident Reporting Forms’ (ARF). In Indiana, the informativeness of these forms were found to be practically nil with respect to driver and vehicle characteristics while very little information was provided for the vehicle defects, environmental or road deficiencies, and presence of different human conditions and states (Shinar and Treat, 1977). Of all the items analysed for the ARF in Australia, 60% had errors. Amongst them a very large error rate was found for the ‘Type of Location’. The cause was the difficulty of categorizing the actual locations with respect to the choices provided in the form (Howard et al., 1979). Other reasons identified for the high error rates in recording of various items of the ARF was the inadequate knowledge of the accident item definitions by the officers responsible for filling the form. Therefore, it was recommended that a training program for the police should be developed, definition of accidents that occur at intersections should be examined, the definitions used by the highways department should be circulated and the ARF should be revised (Howard, 1979). A similar observation was made by Mohan and Bawa (1985) who found that none of the Police officers in Delhi, India were trained for accident investigation. Moreover the ARF used by the Delhi Police was very exhaustive in terms of road and crash information. It contained 98 variables but the officers could not utilize it properly, probably due to lack of time and training.

It was argued by Evans and Courtney (1985) that the ‘cause of accident’ was improperly recorded by police officers in Hong Kong, China, as 52.58% of the reported bus accidents fell into the vague category “other improper or illegal action” which does not identify any precise reason for the accident occurrence. This is because suitable options were not provided in the ARF to properly categorize the cause of accident, thus making the form inadequate. In the ARF prevalent in Abu Dhabi, UAE, there are many fields available but all information is not recorded which results into incomplete and inaccurate data (Khan et al., 2004). Identical findings were reported by Nguyen et al. (2011) who stated that the Vietnamese ARF comprised of 66 fields out of which less than or equal to 50% were incomplete within half of the ARFs collected for the study while the other half contained more than 90%. Location of accident site, weather conditions, and road conditions were amongst the emptiest data fields. In UAE, the six fields of the ARF in the order of most to least difficult to be filled were: injury type, crash type, cause of accident, drug status, landuse, and GPS data (Hawas et al., 2012). The five causes behind the difficulty in filling the ARF of UAE in the order of most to least difficult were: it does not include figures to understand clearly, it does not illustrate, personnel do not have proper training to fill up the form or database, it has many categories, it has less categories (Hawas et al., 2012). In Malaysia, the motorcycle accident fatality data collected from the police was not found to be homogeneous in terms of collision type. The reason provided for this error was the inconsistencies in data entry and illegible description in the ARFs (Abdul Manan et al., 2013).

The cause of error in recording the correct age of the victim in UK was that the Police officers either estimated it or asked the victims about their age, instead of matching it from their date of birth while the mistakes in the classification of serious and slight injuries was attributed to the lack of specific training given to the Police officers in this area (Austin, 1995b). Inaccuracies in coding, particularly of accident severity have also been documented in Australia (Ozanne-Smith and Haworth, 1993). Assessment of injuries by police officers in Ireland had been reported to be subjective and is based on observation and experience (Short and Caulfield, 2016). Similar to UK, in USA also providing false information to the police is one of the causes of error in accident data recording. Two such studies, one conducted in Hawaii (Li et al., 1999) and the other on the ‘National Accident Sampling System Crashworthiness Data System (CDS)’ (Schiff and Cummings, 2004), reported this discrepancy in their accident data. Both were associated with the use of seat belts or safety belts. The studies stated that the victims sometimes falsely claim to the police that they were wearing

seat belt to avoid a fine. From the road users end providing false information is thus an important cause of error.

Time constraint and lack of resources available to the police are also the reason behind accident data recording error in USA, which hinders the gathering of information related to crash causes, crash severity, and injury mechanisms (Farmer, 2003). Lack of scientific measures for the evaluation of various details on accident site along with difference in the police that initially collects data and the one that investigates in detail later was reported as the causes of error in Vietnam (Nguyen et al., 2011). Similarly in other countries around the world errors get introduced in the accident data due to the difference in person who fills out the form at the crash scene and the one who enters the data into a digital system in the Police station or an office after the crash (Loo, 2006; Derriks and Mak, 2007).

Another argument behind the errors in recording by the Police officers are their lack of attention towards road safety as they do not always realize the importance of accident data collection and associate it with the burden of paperwork, listlessness in noting the details of the accident such as cause of the accident or the specifics of the accident location, and mistakes in data entry from the accident reporting form to the database (Derriks and Mak, 2007). One more probable reason for errors in the accident data recorded by the police is their concentration towards the general safety and security of the public rather than collecting accident data. Based on the above argument many classical as well as modern studies suggested that the design of the ARF should be improved and the police officers should be trained and motivated accordingly, to achieve reduction in recording errors (Shinar et al., 1983; Evans and Courtney, 1985; Chung and Chang, 2015).

Spelling errors, alternative road names, incorrect or insufficient location information, missing WISLR (Wisconsin Information System of Local Roads) link/nodes were considered as the reasons behind error in mapping accidents to their location in Wisconsin, USA (Dutta et al., 2007). A later study reported complex interchanges, incomplete or no valid information related to highway/street name, distance from milepost and incorrect geometry such as missing legs/approaches to be responsible for the location error in Wisconsin, USA (Qin et al., 2013). Similarly the six causes accountable for the failure to pinpoint accidents accurately in Indiana, USA were: missing of directional prefix, spelling error, missing or incorrect street suffix, mile marker/interchange/parking lot, alternate name and officer’s error (Tarko et al., 2009). Inconsistencies between crash mapping network and actual road network such as incorrect road name or type were also considered as the cause of error in mapping accidents in UK (Imprialou et al., 2014).

Although the advent of GPS and its application in road safety has improved the way accidents are located but incomplete, inaccurate recording or no recording of GPS coordinates, lack of awareness among police officers and use of a web mapping service instead of a GPS device have become the sources of error for GPS based accident location in various countries around the world such as UAE (Hawas et al., 2012), Italy (Montella et al., 2013), UK (Imprialou et al., 2014) and Croatia (Miler et al., 2016). Other causes of error in accident data recording include police not enquiring about the actual length of the hospital stay, as reported for Hong Kong China (Tsui et al., 2009), non-existence of any systematic communication between police and hospital which resulted into the established definitions of injury not being applied by the police, such as the case in Portugal (Ferreira et al., 2015), and human error (Tarko et al., 2009; Chung and Chang, 2015).

In a retrospect the common causes behind error in accident data recording are:

- Lack of knowledge of the items in the ‘Accident Reporting Form’ by the Police officers
- Lack of training of the Police officers in collection and recording of accident data use of the ARF
- Lack of time and resources available to the Police
- Police not enquiring about the actual length of the hospital stay

- Difference in officers/personnel recording accident information on site and transferring to the accident database
- The ‘Accident Reporting Form’ being very exhaustive
- The ‘Accident Reporting Form’ lacked figures and illustrations
- Asking information from the victims and not verifying it from their licences
- Established definitions of injury severity not being applied by the Police
- Another person processing the crash reports at the police station or an office
- Incomplete, inaccurate recording or no recording of GPS coordinates or use of a web mapping service instead of a GPS device
- Spelling errors, alternative road name, partially missing or incorrect road name and geometry
- False information provided by the victims
- Human error

5. Methods of rectification

5.1. Methods of error estimation and rectification related to road traffic injuries

Apart from the general recommendations to motivate and train the police and other authorities responsible for accident data collection and record keeping, techniques for the correction of errors in the data that has already been collected was also devised by scientists and engineers. A list of these methods is presented in Table 10 along with their respective error type.

Table 10
List of rectification or estimation methods along with their respective error type.

Method of Rectification/Error Estimation	Type of error	Country	Author (Year of Publication)
Capture-Recapture	Under-reporting of severe injuries and fatalities	Pakistan	Razzak and Luby (1998)
	Under-reporting of non-fatal injuries	France (La Re’union)	Aptel et al. (1999)
	Under-reporting of fatalities and non-fatal injuries	Nicaragua	Tercero and Andersson (2004)
	Under-reporting of non-fatal injuries	Australia	Meuleners et al. (2006)
	Under-reporting of non-fatal injuries	France	Amoros et al. (2007)
	Under-reporting of fatal accidents	Canada	Trepanier (2009)
	Under-reporting of fatalities	Pakistan	Lateef (2010)
	Under-reporting of fatalities	Malawi	Samuel et al. (2012)
	Under-reporting of all types of injuries	New Zealand	Tin et al. (2013)
	Under-reporting of fatalities	South Africa	Chokotho et al. (2013)
	Under-reporting of fatalities and non-fatal injuries	Ethiopia	Abegaz et al. (2014)
	Under-reporting of fatalities, slight and severe injuries	Denmark	Janstrup et al. (2014)
	Comparison with Health Sector Data	Under-reporting of fatalities, slight and severe injuries	UK
Under-reporting of severe injuries		Netherland	Maas and Harris (1984)
Under-reporting of severe injuries		Australia	O’Connor (1992)
Under-reporting of fatalities and non-fatal injuries		Various	Sigua and Palmiano (2005)
Under-reporting of fatalities		China	Hu et al. (2011)
Probabilistic Linkage	Under-reporting of slight and severe injuries	Denmark	Abay (2015)
	Under-reporting of all types of injuries	Australia	Rosman and Knuiman (1994)
	Under-reporting of fatalities, slight, severe, and all injuries	USA	Sciortino et al. (2005)
	Under-reporting of non-fatal injuries	France	Amoros et al. (2006)
	Under-reporting of non-fatal injuries	New Zealand	Alsop and Langley (2001)
Mixed Deterministic and Probabilistic Linkage	Under-reporting of fatalities, slight and severe injuries	Australia	Rosman (2001)
	Under-reporting of slight and severe injuries	New Zealand	Wilson et al. (2012)
Matching with Health Sector Data	Under-reporting of non-fatal injuries	Portugal	Ferreira et al. (2015)
	Under-reporting of fatalities, slight and severe injuries	Australia	Watson et al. (2015)
	Under-reporting of slight and severe injuries	Ireland	Short and Caulfield (2016)
Cross-Sectional Population-Based Study	Under-reporting of non-fatal injuries	Hong Kong-China	Loo and Tsui (2007)
	Under-reporting of fatalities and non-fatal injuries	Pakistan	Bhatti et al. (2011)
Driver Surveys	Under-reporting of fatalities, slight and severe injuries	India	Dandona et al. (2008)
	Under-reporting of fatalities and non-fatal injuries	Sri Lanka	Periyasamy et al. (2013)
Comparison with fire department and insurance data	Under-reporting of all types of injuries	USA	Scott and Carroll (1971)
	Under-reporting of all types of injuries	Japan	Nakahara and Wakai (2001)
Regression Analysis	Under-reporting of fatalities and non-fatal injuries	Greece	Petridou et al. (2009)
	Under-reporting of fatalities and severe injuries	Vietnam	Nguyen et al. (2011)
Data Accuracy Index	Under-reporting of fatalities	Russia	Kudryavtsev et al. (2013)

5.1.1. Capture-recapture method

Most widely used method to rectify inaccuracies in road crash reporting and estimate the true numbers of accidents is the ‘Capture-Recapture’ method. This is also evident from the pie chart shown in Fig. 8. It is based on the methodology proposed by Cook et al. (1967) for the estimation of deaths and injuries due to road traffic accidents. It is used for the identification of the number of accidents missing in different records, such as the police or the hospital accident record, by estimating the probable number of accidents for a given population. It is a very famous tool to estimate the level of error, due to under-reporting, in the accident database and has been used in many studies such as: Razzak and Luby (1998) for the assessment of deaths and injuries due to road traffic accidents in Pakistan, Aptel et al. (1999) used it to estimate the total number of injured people in road accidents in La Réunion (a French island), Tercero and Andersson (2004) provided an estimate of the incidence of traffic related fatalities and non-fatal injuries with the help of it in Nicaragua. In Western Australia Meuleners et al. (2006) used this method to establish that the number of fatalities and serious injuries pertinent to heavy vehicle drivers were under-estimated. Amoros et al. (2007) argued that non-fatal road casualties are under-reported in the French Rhone County and reported that the police ascertainment rate was at most 57%, using the Capture-Recapture analysis. This method was used by Trépanier et al. (2009) who estimated that the road accident database in Quebec, Canada accounts for about two third of the fatal accidents that were related to the transportation of hazardous material. Lateef (2010), Samuel et al. (2012) and Chokotho et al. (2013) utilised this method for the determination of road traffic deaths in Pakistan, Malawi and South Africa respectively. Shortcomings in the bicycle crash injury data in New Zealand was

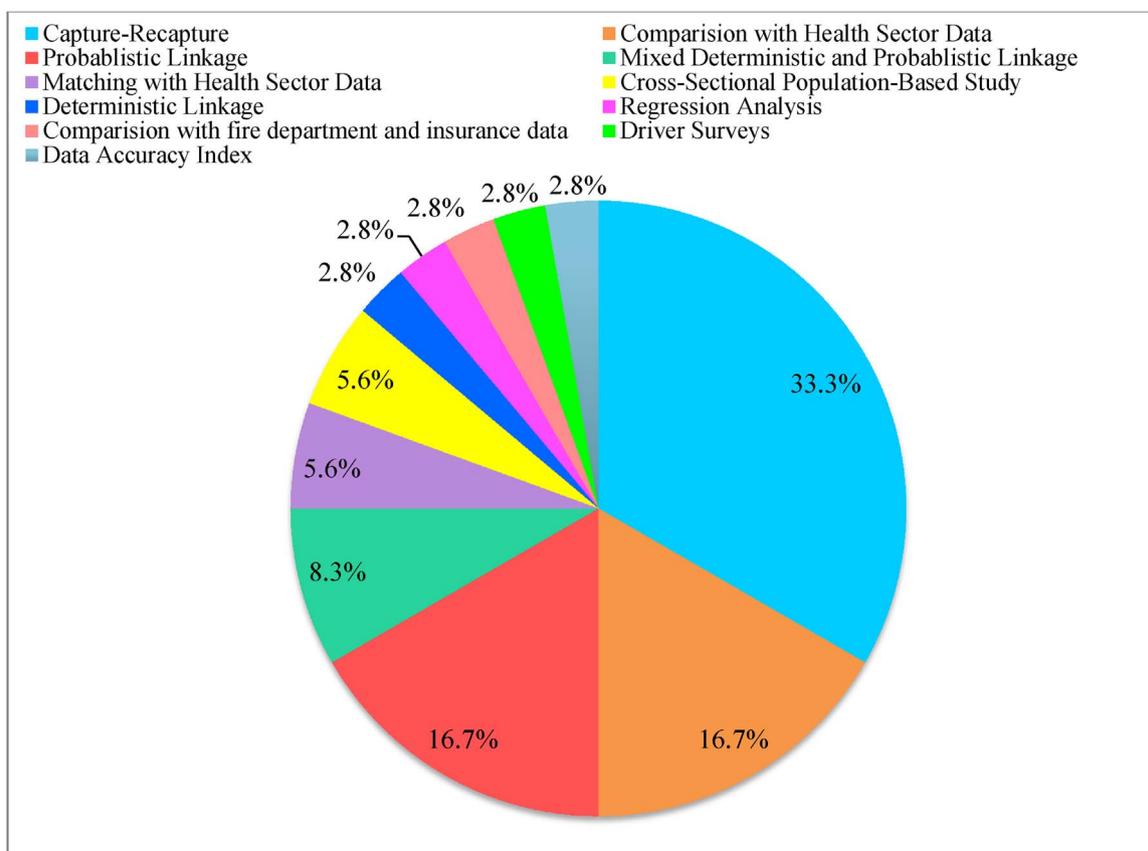


Fig. 8. Distribution of methods used in the error estimation or rectification of road traffic injuries.

evaluated by Tin et al. (2013) using capture-recapture method and comparison with self-reported injury data. Two recent studies utilised this method for the estimation of fatalities and injuries. One was pertinent to Ethiopia (Abegaz et al., 2014) and the other was related to Denmark (Janstrup et al., 2014).

5.1.2. Comparison with health sector data

It is a simpler method as compared to Capture-Recapture. It is easy to apply, therefore, more readily used for the analysis of slight and severe injury accidents along with fatalities, while Capture-Recapture was mostly utilized for the error estimation of fatal road traffic accidents. One of the classical examples of this method is a study from UK where the cases of road accidents which occurred in Birmingham, and attended the accident hospital were compared with the police records to check whether they were included in the official national statistics (Bull and Roberts, 1973). In Netherlands, the police data related to road accident in-patients was compared with hospital discharge data to establish the extent of under-reporting of severe injuries (Maas and Harris, 1984). Similarly for the estimation of under-reporting of severe injuries, state level comparisons between hospital admissions and police reported admissions were made in Australia (O'Connor, 1992). Health sector data was utilized to estimate true figures of road traffic fatalities and injuries in Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand and Vietnam (Sigua and Palmiano, 2005). Death certificates completed by physicians were compared with police records in China to evaluate the rate of death from road traffic injuries (Hu et al., 2011). This simple method of comparison with health sector data to estimate the level of error in police data is still in practice as Abay (2015) applied it to investigate the nature of reporting bias in police reported accident data in Denmark.

5.1.3. Linkage with health sector data

This method involves two or more data sets, usually the police and the hospital data, to be linked with each other using certain fields such as name of victim, age, sex, date and time of accident, severity, etcetera. Unlike Capture-Recapture analysis, it does not require population as one of the input parameters. Apart from simple comparisons this is a more scientific method to estimate the error in accident data, particularly under-reporting of road traffic fatalities and injuries. The linkage could be probabilistic, deterministic or a mixture of both.

5.1.3.1. Probabilistic linkage. To determine the extent of under-reporting in the Western Australian Police data, a Generalised Iterative Record Linkage System (GIRLS) was used by Rosman and Knuiman (1994) to link the police data with the hospital records. In a subsequent study another software known as “Automatch” was used accompanied by GIRLS to carry out the linking process (Rosman, 2001). Age, gender, date and time variables were used in USA to link the Statewide Integrated Traffic Reporting System (SWITRS) data specific to San Francisco with the hospital data (Sciortino et al., 2005). The same set of variables along with crash location and type of road user were used by Amoros et al. (2006) to link the police data, in Rhône county, France, with the road trauma registry data using a semi-automated record-linkage procedure. The “Automatch” software was also used by Alsop and Langley (2001) and Wilson et al. (2012) to probabilistically link the police records with the hospital data in New Zealand. Thirteen linkage variables which included: surname, first and second initials of the crash victim, geographic location, crash day and month, birth day, month and year, gender, age, and casualty type; were used by Alsop and Langley (2001) while only name, age, gender and date of crash were used by Wilson et al. (2012).

5.1.3.2. Deterministic linkage. Use of deterministic linkage techniques, to link police data with hospital data or other sources, is scarce but it is

Table 11
Methods used for the rectification of accident location.

Method of Rectification	Country	Author (Year of Publication)
Highway feature data Geographical Information System (GIS)	UK	Austin (1995)
GIS-based spatial data validation system	Hong Kong-China	Loo (2006)
Crash mapping algorithm	USA	Dutta et al. (2007)
Probabilistic assignment of crashes to roads	USA	Tarko et al. (2009)
Lagrangian relaxation-based solution algorithm	USA	Tegge and Ouyang (2009)
Angular difference between direction of vehicle and road segment	UK	Wang et al. (2009)
Crash-Mapping Automation Tool	USA	Qin et al. (2013)
Fuzzy logic based crash mapping	UK	Imprialou et al. (2014)
Geocoding using Google Maps Application Programming Interface (API)	Canada	Burns et al. (2014)
Vehicle Black Box (VBB)	Korea	Chung and Chang (2015)
Crash mapping using Multilevel Logistic Regression CM-MLOGIT	UK	Imprialou et al. (2015)
Computer Algorithm that utilizes string matching technique	Croatia	Miler et al. (2016)

a simpler technique as compared to probabilistic methods and has the capability to provide reliable and accurate results when interpreting small data files. This technique was applied by Petridou et al. (2009) to estimate the degree of under-reporting by linking the Greek road traffic police database with the Emergency Department Injury Surveillance System.

5.1.3.3. Mixed deterministic and probabilistic linkage. Data from multiple sources can be effectively linked with each other using both deterministic and probabilistic linkage technique. One such example is the study by Watson et al. (2015). They linked data from four sources to estimate the under-reporting by Police in Queensland, Australia using 'LinkageWiz' software. The same software was utilized by Short and Caulfield (2016) to link hospital, police and injury claims data. They selected age, gender, date, county and mode, as linkage variables. Another example is of Ferreira et al. (2015), in which the police data was linked with the hospital data, to estimate the level of under-reporting in police data as part of their study on the misclassification of injury severity by the police. Name, age, gender, date, and time, were the linkage variables (Amorim et al., 2014).

5.1.4. Matching with health sector data

Health sector data whether hospital records, injury surveillance records, ambulance records, trauma registry, mortuary logs, or emergency department records are commonly used to measure the discrepancies, shortcomings, and under-reporting of road traffic injuries and fatalities by the police. This is achieved by matching police and health sector records with the help of variables or fields common in both databases. It is a much simpler method as compared to linkage techniques and hence popular among middle-income countries. Examples are China and Pakistan. In China the Hong Kong Identity Card number was used to match the road crash victims' information in the police records with the hospital record (Loo and Tsui, 2007). The name of the victim, his age, gender, date and time of crash were used in Pakistan to match the information across three datasets which were police, hospital and ambulance records (Bhatti et al., 2011).

5.1.5. Cross-Sectional population-based study

In these kinds of studies interviews with road crash injury victims or their family members, in case of death of the victim, are carried out using a questionnaire. They are similar to Capture-Recapture method as they involve a given population, however a time frame is required to be specified. A population is selected and the interviewees are asked to recall any accident that occurred in the past during the study period, which is usually the preceding 12 months. Example is a study carried out in Hyderabad, India (Dandona et al., 2008) in which the recall period was 12 months for non-fatal road traffic injury cases and three years for fatal road traffic injury cases. The same recall period of 12 months was used to measure under-reporting in police data in Sri Lanka (Periyasamy et al., 2013). If the interviewee claimed that he/they

reported the casualty to the police, then it was cross-checked with the police records using name, age, address, crash registration number, date and place of accident occurrence, place of reporting, and severity of injury.

5.1.6. Other methods of error estimation and rectification related to road traffic injuries

One of the earliest reports on accident data inaccuracies applied driver surveys to estimate the correct accident frequencies calculated through extrapolations of reported totals, using ratios of non-reporting (Scott and Carroll, 1971). They also used hospital records to obtain true estimates of the number and severity of accident injuries. Extrapolations of reported totals have also been applied in Vietnam to estimate the magnitude of traffic accidents and injuries through regression Analysis (Nguyen et al., 2011). The police data was compared with fire department and the "Marine and Fire Insurance Association of Japan" to clarify the magnitude of under-reporting of traffic injuries pertinent to children in Japan (Nakahara and Wakai, 2001). As a measure of error in traffic mortality data, a diagnostic test known as 'Data Accuracy Index' was formulated in Russia (Kudryavtsev et al., 2013). This was used to measure the reliability of police and health sector data.

5.2. Methods of rectification related to accident location

The largest number of rectification techniques devised is for accident location. This is because removal of location errors from accident data is necessary to ensure that the counts of number of accidents for a particular segment or location is correct (Wang et al., 2009). In general the methods deployed for the rectification of location errors involve matching of accident data with road network data. Unlike *weather* such as rainy, cloudy or sunny; *volume* such as low, medium or high and *visibility* such as day-time, night-time, fog or mist; that keep changing and thus, can not be matched with the information recorded at the time of accident, the variables commonly utilized to carry out the matching process were road name and type, district/county information, direction of travel, road and crash co-ordinates, landuse, land jurisdiction, and distance from a landmark. A list of these methods is presented in Table 11.

The first account of the application of any such matching technique is from UK. The objective was to remove errors pertinent to locational variables. A Geographical Information System (GIS) platform was used to check the inaccuracies. The highway features used to match an accident to its location on the road network were: road class, road number, district, speed limit, pedestrian crossing facilities, junction control, junction detail, and carriageway type and markings (Austin, 1995a). The procedure successfully identified the mistakes and the study concluded that if the highway features data is routinely entered into a GIS database then there will be no need for the police to record that information on their accident report forms.

Similarly a GIS platform was also used in Hong Kong, currently a

special administrative region of China, to carry out spatial validation of the police crash database (Loo, 2006). The procedure involved snapping the spatial information regarding location of a road crash to its nearest junction or road section and subsequently matching this information with the district board database to assign the correct district board. Unlike Austin (1995a), the spatial variables used in this study for matching purpose also utilized textual features. Identical to the findings of Austin (1995a), the study concluded that the process of filling crash report forms using pen-and-paper and then digitizing this information is error-prone. Therefore, the road crash spatial data should be validated before conducting any analysis.

Again a GIS based software known as WISLR was used in Wisconsin, USA by Dutta et al. (2007) to develop a pinpoint map of the accidents, which occurred on the local roads. This algorithm matched the roadway-route prefix, name, type, and suffix from the Wisconsin Information System of Local Roads (WISLR) with the on-street and at-street names in the Wisconsin crash database. The algorithm mapped more than two third of the road crashes, however, those which could not be mapped were due to spelling errors, unintelligible records, alternate road names and missing intersections in the GIS database.

Unlike Austin (1995a), Loo (2006) and Dutta et al. (2007), a probabilistic approach was adopted by Tarko et al. (2009) for assigning crashes to roads in Indiana, USA. Instead of using GPS co-ordinates; county ID, road administrator and road type were used for matching accidents with their respective locations. The conditional probability of match between a road crash record and road inventory record was calculated and the location with the highest likelihood was assigned. They concluded that insufficient information in the road inventory database was one of the causes for the decreased number of one to one matches. Their conclusion supported the findings of a previous study (Steenberghen et al., 2004) in which a similar approach was used to locate the accidents in Belgium. The Belgian study also stated the shortcoming of the technique by admitting that the location of accidents which occurred on local roads in the country side, where there is no particular address to refer to, could not be identified.

A new addition to the available methods was made by Wang et al. (2009) who used direction of travel to assign an accident to its location. Crash location co-ordinates were utilized to calculate a 'weighted score' that estimates the correctness of the segment assigned. Their work signifies the notion that even the availability of crash co-ordinates is no assurance that an accident could be correctly located on the map. With the help of this method the locations of about 98% of the accidents were rectified.

Accidents can be clustered and subsequently located on the basis of geometric attributes, thus, different clustering techniques were applied to crash data for location rectification along with several other applications. In hierarchical clustering, which is a type of clustering technique, the clusters are split into two sub-clusters (Estivill-Castro, 2002). The complex problem in accident data not only requires the clustering exercise to be performed on attributes with diverse units but also the clusters to be divided into three or more sub-clusters. Moreover in clustering of large datasets, with several levels of agglomeration or division, the resultant number of clusters will be numerous, thus the construction of a dendrogram could be computationally prohibitive (Jain et al., 1999). Therefore, different sets of algorithms coupled with suitable distance measures were developed to address the issue of incorrect accident location. Tegge and Ouyang (2009) have proposed a Lagrangian relaxation-based solution algorithm to estimate the most-probable location for each crash. The matching exercise revealed that 55.6% of the crashes were either located on the borderline of multiple sites or do not coincide with the existing road network. The probability estimate, which is a function of the distance measure, might not be suitable for complex road geometries such as roundabouts and multi-leg intersections.

Apart from the common variables used in previous studies, such as the document ID, the names of the municipality and the county in

which the accident occurred, new variables were used in the study by Qin et al. (2013). These variables included landuse like city, town or village, land jurisdiction like public or private and distance from the listed intersection location. However, the algorithm could map only 83% of all crashes. The reasons for this failure were the unavailability of valid names for the major/minor roads, misclassification of an intersection in the police report or the state road network database, missing number of legs of an intersection and complex interchanges. Similar to the shortcomings in the previous studies, the issue of complex geometry still remained unresolved.

Exhaustive effort was made by Imprialou et al. (2014) and Imprialou et al. (2015) to rectify the locations in the UK crash records database (STATS 19). They were able to achieve accuracy level of 98.9% and 97.1% using FLCM and CM-LOGIT respectively. But FLCM had an edge over other methods available in the literature as it can match accidents that occurred on complex geometric locations such as roundabouts.

Apart from the conventional 'matching' approach mentioned in all the studies previously discussed a data recording systems known as the 'Vehicle Black Box (VBB)' was used by Chung and Chang (2015) which recorded information such as location and time of accident and compared it with the information recorded by the Police. It was established that a spatial difference of 84.5 m and a temporal difference of 29 min exist between the information recorded by VBB and the data collected by Police through eyewitness's accounts or the statements given by the accident parties. Human error was considered as the cause behind the discrepancy in the Police record.

To assign co-ordinates to crash records in Quebec, Canada a methodology was developed by Burns et al. (2014). It simply required the textual address as the input variable. They found that despite obtaining a competitive match rate manual revision was necessary to ensure that the results returned by the 'geocoder' relate to the same intersection mentioned in the input data. Similar to the work done by Burns et al. (2014), a study conducted by Miler et al. (2016), to obtain the correct location of accidents in Croatia. It was found that the reason behind the erroneous recording of coordinates and street names in traffic accident records (TAR) was the lack of awareness among police officers and use of web services instead of GPS devices provided by the Police department.

6. Conclusion

Errors in accident data are inevitable because it is a passive system in which the incident is required to be reported. However, its extent varies from country to country depending upon various factors. To discern the pattern of errors in accident data they were classified with respect to each type and analysed with respect to geographic region and income level. The analysis revealed a stark relationship between the level of error and the income status concluding that the higher the income level the lower would be the probability of error. Furthermore, the level of reporting also varied with the injury severity. Accidents that involved any fatalities were the most reported while those which did not involve any injury, such as property damage only accidents, were the least reported.

The errors were classified into two major groups based on their nature. One was error in accident data reporting and the other was error in accident data recording. Errors in recording accidents were found to be more harmful than error in reporting because the later can be estimated using different techniques but the former cannot be rectified completely. Therefore, it is extremely important to motivate the personnel responsible for accident data collection to record all information as correctly as possible, specifically accident location and level of injury severity. The accident reporting forms should be improved taken into consideration the limitations of data collection on an active accident site. The definitions of all fields and items in the form should be clear and well circulated among all agencies pertinent to

accident data collection and record keeping. Furthermore it is recommended that modern tools should be used to improve the quality of data recorded at the accident scene (Topolšek et al., 2014). Examples of such tools are total stations, laser scanners, and photogrammetry. The police officers, in all countries in general and low and middle-income countries in particular, should be given trainings specific to accident data collection. Although the cause of errors in High income countries is slightly different than low and middle income countries but the solution to them is the same, which is to reform the Policing System. Hence to reduce errors in accident data around the world it is recommended that the efficiency, integrity, and affordability of the Police should be improved. The process of recording accident information on site should be automated to reduce human errors and the loss of information during transfer from manual reports to digital database. Awareness among public have to be created about reporting accidents to the police, particularly those which involve single vehicle, pedestrian or bicyclist, and slight or no injury.

Acknowledgments

The authors acknowledge the Ministry of Higher Education (MOHE) Malaysia for their FRGS Grant No. 203/PAWAM6071257. The funding source had no influence on the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. The authors have no competing interest, whether financial, professional or personal, with any individual or organisation.

References

- Abay, K.A., 2015. Investigating the nature and impact of reporting bias in road crash data. *Transp. Res. Part A: Policy Pract.* 71, 31–45.
- Abdul Manan, M.M., Várhelyi, A., 2012. Motorcycle fatalities in Malaysia. *IATSS Res.* 36, 30–39.
- Abdul Manan, M.M., Várhelyi, A., 2015. Motorcyclists' road safety related behavior at access points on primary roads in Malaysia—a case study. *Saf. Sci.* 77, 80–94.
- Abdul Manan, M.M., Jonsson, T., Várhelyi, A., 2013. Development of a safety performance function for motorcycle accident fatalities on Malaysian primary roads. *Saf. Sci.* 60, 13–20.
- Abdul Manan, M.M., 2014. Motorcycles entering from access points and merging with traffic on primary roads in Malaysia: behavioral and road environment influence on the occurrence of traffic conflicts. *Accid. Anal. Prev.* 70, 301–313.
- Abegaz, T., Berhane, Y., Worku, A., Assrat, A., Assefa, A., 2014. Road traffic deaths and injuries are under-reported in Ethiopia: a capture-recapture method. *PLoS One* 9, e103001.
- Ahmed, I., Ahmed, B., Hainin, M.R., 2014. Road traffic accident characteristics in dhaka, Bangladesh. *Jurnal Teknologi* 71, 75–82.
- Al-Ghamdi, A.S., 1996. Road accidents in Saudi Arabia: a comparative and analytical study. *Urban Transp. Environ.* 23, 231–253.
- Al-Ghamdi, A.S., 2003. Analysis of traffic accidents at urban intersections in Riyadh. *Accid. Anal. Prev.* 35, 717–724.
- Alsop, J., Langley, J., 2001. Under-reporting of motor vehicle traffic crash victims in New Zealand. *Accid. Anal. Prev.* 33, 353–359.
- Amorim, M., Ferreira, S., Couto, A., 2014. Linking police and hospital road accident records. *Transp. Res. Record: J. Transp. Res. Board* 2432, 10–16.
- Amoros, E., Martin, J.-L., Laumon, B., 2006. Under-reporting of road crash casualties in France. *Accid. Anal. Prev.* 38, 627–635.
- Amoros, E., Martin, J.-L., Laumon, B., 2007. Estimating non-fatal road casualties in a large French county, using the capture-recapture method. *Accid. Anal. Prev.* 39, 483–490.
- Aptel, I., Salmi, L.R., Masson, F., Bourdè, A., Henrion, G., Erny, P., 1999. Road accident statistics: discrepancies between police and hospital data in a French island. *Accid. Anal. Prev.* 31, 101–108.
- Austin, K., 1995a. The identification of mistakes in road accident records: part 1: locational variables. *Accid. Anal. Prev.* 27, 261–276.
- Austin, K., 1995b. The identification of mistakes in road accident records: part 2: casualty variables. *Accid. Anal. Prev.* 27, 277–282.
- Bhatti, J.A., Salmi, L.-R., 2012. Challenges in evaluating the decade of action for road safety in developing countries: a survey of traffic fatality reporting capacity in the eastern Mediterranean region. *Traffic Inj. Prev.* 13, 422–426.
- Bhatti, J.A., Razzak, J.A., Lagarde, E., Salmi, L.-R., 2011. Differences in police, ambulance, and emergency department reporting of traffic injuries on Karachi-Hala road, Pakistan. *BMC Res. Notes* 4, 1–7.
- Blincoe, L.J., Seay, A., Zaloshnja, E., Miller, T., Romano, E., Luchter, S., Spicer, R., 2002. The Economic Impact of Motor Vehicle Crashes, 2000. US Department of Transportation, National Highway Traffic Safety Administration, Washington, DC.
- Bliss, T., Breen, J., 2012. Meeting the management challenges of the decade of action for road safety. *IATSS Res.* 35, 48–55.
- Bourbeau, R.R., Laberge-Nadeau, C., 1977. Road injuries in Quebec (1974): nature, severity and direct costs. In: Huelke, D.F. (Ed.), 21 st Annual Meeting, American Association for Automotive Medicine. Arlington.
- Broughton, J., Keigan, M., Yannis, G., Evgenikos, P., Chaziris, A., Papadimitriou, E., Bos, N.M., Hoeglinger, S., Pérez, K., Amoros, E., Holló, P., Tecl, J., 2010. Estimation of the real number of road casualties in Europe. *Saf. Sci.* 48, 365–371.
- Bull, J.P., Roberts, B.J., 1973. Road accident statistics—a comparison of police and hospital information. *Accid. Anal. Prev.* 5, 45–53.
- Burns, S., Miranda-Moreno, L., Stipanovic, J., Saunier, N., Ismail, K., 2014. Accessible and practical geocoding method for traffic collision record mapping. *Transp. Res. Record: J. Transp. Res. Board* 2460, 39–46.
- Chokotho, L.C., Matzopoulos, R., Myers, J.E., 2013. Assessing quality of existing data sources on road traffic injuries (RTIs) and their utility in informing injury prevention in the western cape province, South Africa. *Traffic Inj. Prev.* 14, 267–273.
- Chung, Y., Chang, I., 2015. How accurate is accident data in road safety research?: an application of vehicle black box data regarding pedestrian-to-taxi accidents in Korea. *Accid. Anal. Prev.* 84, 1–8.
- Cook, L.M., Brower, L.P., Croze, H.J., 1967. The accuracy of a population estimation from multiple recapture data. *J. Anim. Ecol.* 36, 57–60.
- Dandona, R., Kumar, G.A., Ameer, M.A., Reddy, G.B., Dandona, L., 2008. Underreporting of road traffic injuries to the police: results from two data sources in urban India. *Inj. Prev.: J. Int. Soc. Child Adolesc. Inj. Prev.* 14, 360–365.
- Derriks, H.M., Mak, P.M., 2007. Underreporting of Road Traffic Casualties. IRTAD Special Report. IRTAD, Geneva.
- Dutta, A., Parker, S., Qin, X., Qiu, Z., Noyce, D., 2007. System for digitizing information on wisconsin's crash locations. *Transp. Res. Record: J. Transp. Res. Board* 256–264.
- Elvik, R., Mysen, A., 1999. Incomplete accident reporting: meta-analysis of studies made in 13 countries. *Transp. Res. Record: J. Transp. Res. Board* 1665, 133–140.
- Estivill-Castro, V., 2002. Why so many clustering algorithms: a position paper. *ACM SIGKDD Explor. Newslett.* 4, 65–75.
- Evans, W.A., Courtney, A.J., 1985. An analysis of accident data for franchised public buses in Hong Kong. *Accid. Anal. Prev.* 17, 355–366.
- Farmer, C.M., 2003. Reliability of police-reported information for determining crash and injury severity. *Traffic Inj. Prev.* 4, 38–44.
- Ferreira, S., Falcão, L., Couto, A., Amorim, M., 2015. The quality of the injury severity classification by the police: an important step for a reliable assessment. *Saf. Sci.* 79, 88–93.
- Giles, M.J., 2001. Data for the study of road crashes in Australia. *Aust. Econ. Rev.* 34, 222–230.
- Hauer, E., Hakkert, A., 1988. Extent and some implications of incomplete accident reporting. *Transp. Res. Rec.* 1185, 1–10.
- Hawas, Y.E., Khan, M.B., Maraqa, M.A., 2012. Analysis of the effectiveness of the road-crash database in the United Arab Emirates. *J. Transp. Saf. Secur.* 4, 225–243.
- Howard, B.V., Young, M.F., Ellis, J.P., 1979. Appraisal of the Existing Traffic Accident Data Collection and Recording System—South Australia. Dept. of Transport, Office of Road Safety, CR6, Canberra.
- Hu, G., Baker, T., Baker, S.P., 2011. Comparing road traffic mortality rates from police-reported data and death registration data in China. *Bull. World Health Organ.* 89, 41–45.
- IRC, 2012. Road Accident Recording Forms A-1 and A-4. Indian Roads Congress, New Delhi.
- IRTAD, 2015. Road Safety Annual Report 2015. OECD Publishing, Paris.
- Imprialou, M.-I.M., Quddus, M., Pitfield, D.E., 2014. High accuracy crash mapping using fuzzy logic. *Transp. Res. Part C: Emerg. Technol.* 42, 107–120.
- Imprialou, M.-I.M., Quddus, M., Pitfield, D.E., 2015. Multilevel logistic regression modeling for crash mapping in metropolitan areas. *Transp. Res. Record: J. Transp. Res. Board* 2514, 39–47.
- Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. *ACM Comput. Surv.* (CSUR) 31, 264–323.
- Janstrup, K.H., Hels, T., Kaplan, S., Sommer, H.M., Lauritsen, J., 2014. Understanding traffic crash under-reporting: linking police and medical records to individual and crash characteristics. *Transport Research Arena* 2014.
- Kareem, A., 2003. Review of global menace of road accidents with special reference to Malaysia—a social perspective. *Malays. J. Med. Sci.: MJMS* 10, 31–39.
- Karlaftis, M.G., Golias, I., 2002. Effects of road geometry and traffic volumes on rural roadway accident rates. *Accid. Anal. Prev.* 34, 357–365.
- Khan, M.A., Salim, A.L., Kathairi, A., Garib, A.M., 2004. A GIS based traffic accident data collection, referencing and analysis framework for Abu Dhabi. In: CODATU (Conference). (11th: 2004: Bucharest, Romania). pp. 292–302.
- Kim, K., Levine, N., 1996. Using GIS to improve highway safety. *Comput. Environ. Urban Syst.* 20, 289–302.
- Kudryavtsev, A.V., Kleshchinov, N., Ermolina, M., Lund, J., Grijbovski, A.M., Nilssen, O., Ytterstad, B., 2013. Road traffic fatalities in Arkhangelsk, Russia in 2005–2010: reliability of police and healthcare data. *Accid. Anal. Prev.* 53, 46–54.
- Levine, N., Kim, K.E., 1998. The location of motor vehicle crashes in Honolulu: a methodology for geocoding intersections. *Comput. Environ. Urban Syst.* 22, 557–576.
- Li, L., Kim, K., Nitz, L., 1999. Predictors of safety belt use among crash-involved drivers and front seat passengers: adjusting for over-reporting. *Accid. Anal. Prev.* 31, 631–638.
- Ljung Aust, M., Fagerlind, H., Sagberg, F., 2012. Fatal intersection crashes in Norway: patterns in contributing factors and data collection challenges. *Accid. Anal. Prev.* 45, 782–791.
- Loo, B.P.Y., Tsui, K.L., 2007. Factors affecting the likelihood of reporting road crashes resulting in medical treatment to the police. *Inj. Prev.* 13, 186–189.

- Loo, B.P.Y., 2006. Validating crash locations for quantitative spatial analysis: a GIS-based approach. *Accid. Anal. Prev.* 38, 879–886.
- Maas, M.W., Harris, S., 1984. Police recording of road accident in-patients: investigation into the completeness, representativity and reliability of police records of hospitalized traffic victims. *Accid. Anal. Prev.* 16, 167–184.
- Meuleners, L.B., Lee, A.H., Cercarelli, L.R., Legge, M., 2006. Estimating crashes involving heavy vehicles in Western Australia, 1999–2000: a capture-recapture method. *Accid. Anal. Prev.* 38, 170–174.
- Miller, M., Todić, F., Ševrović, M., 2016. Extracting accurate location information from a highly inaccurate traffic accident dataset: a methodology based on a string matching technique. *Transp. Res. Part C: Emerg. Technol.* 68, 185–193.
- Mohan, D., Bawa, P.S., 1985. An analysis of road traffic fatalities in Delhi, India. *Accid. Anal. Prev.* 17, 33–45.
- Mohan, D., Tiwari, G., Bhalla, K., 2015. Road safety in India-Status Report. Transportation Research & Injury Prevention Programme (TRIPP), Indian Institute of Technology, Delhi.
- Montella, A., Andreassen, D., Tarko, A., Turner, S., Mauriello, F., Imbriani, L., Romero, M., 2013. Crash databases in Australasia, the European Union, and the United States. *Transp. Res. Rec.: J. Transp. Res. Board* 2386, 128–136.
- Nakahara, S., Wakai, S., 2001. Underreporting of traffic injuries involving children in Japan. *Inj. Prev.* 7, 242–244.
- Ng, C.P., Ma Soem, D.M., Rahman, H.A., Radin Sohadi, R.U., 2007. Research on the development of Malaysian road traffic injury surveillance system. *J. East. Asia Soc. Transp. Stud.* 7, 2728–2742.
- Nguyen, H.D.U.C., Duong Thi, M.H.O.A., Nguyen, T.H., Nguyen, N.B.A.O., 2011. Study on quality of existing traffic accident data in Vietnam. *Proceedings of the Eastern Asia Society for Transportation Studies*, 2011 374.
- O'Connor, P., 1992. Road injury information program. National Injury Surveillance Unit, Australian Institute of Health, Adelaide.
- Ozanne-Smith, J., Haworth, N., 1993. Development of data collection methodology: Crashes resulting in hospitalisation and casualty crashes not resulting in hospitalisation. CR 120. Monash University Accident Research Centre, Canberra.
- PDRM, 2008. Laporan Tahunan PDRM 2008 Kuala Lumpur. Royal Malaysia Police.
- PIARC, 2003. Road safety manual: recommendations from the World Road Association (PIARC). Switzerland], Route 2 Market.
- Periyasamy, N., Lynch, C.A., Dharmaratne, S.D., Nugegoda, D., Østbye, T., 2013. Under reporting of road traffic injuries in the district of Kandy, Sri Lanka. *BMJ Open* 3.
- Petridou, E.T., Yannis, G., Terzidis, A., Dessypris, N., Germeni, E., Evgenikos, P., Tselenti, N., Chaziris, A., Skalkidis, I., 2009. Linking emergency medical department and road traffic police casualty data: a tool in assessing the burden of injuries in less resourced countries. *Traffic Inj. Prev.* 10, 37–43.
- Qin, X., Parker, S., Liu, Y., Graettinger, A.J., Forde, S., 2013. Intelligent geocoding system to locate traffic crashes. *Accid. Anal. Prev.* 50, 1034–1041.
- Razzak, J.A., Luby, S.P., 1998. Estimating deaths and injuries due to road traffic accidents in Karachi, Pakistan, through the capture-recapture method. *Int. J. Epidemiol.* 27, 866–870.
- Rosman, D.L., Knuijan, M.W., 1994. A comparison of hospital and police road injury data. *Accid. Anal. Prev.* 26, 215–222.
- Rosman, D.L., 2001. The Western Australian road injury database (1987–1996): ten years of linked police, hospital and death records of road crashes and injuries. *Accid. Anal. Prev.* 33, 81–88.
- SIKA, 2009. Vägtrafikskador 2008. Staten Institut för Kommunikationsanalys.
- Samuel, J.C., Sankhulani, E., Qureshi, J.S., Baloyi, P., Thupi, C., Lee, C.N., Miller, W.C., Cairns, B.A., Charles, A.G., 2012. Under-reporting of road traffic mortality in developing countries: application of a capture-recapture statistical model to refine mortality estimates. *PLoS One* 7, e31091.
- Schiff, M.A., Cummings, P., 2004. Comparison of reporting of seat belt use by police and crash investigators: variation in agreement by injury severity. *Accid. Anal. Prev.* 36, 961–965.
- Sciortino, S., Vassar, M., Radetsky, M., Knudson, M.M., 2005. San Francisco pedestrian injury surveillance: mapping, under-reporting, and injury severity in police and hospital records. *Accid. Anal. Prev.* 37, 1102–1113.
- Scott, R.E., Carroll, P.S., 1971. Acquisition of Information on Exposure and on Non-Fatal Crashes. Volume II-Accident Data Inaccuracies. Washington, D.C. 20591.
- Shinar, D., Treat, J.R., 1977. Tri-level Study: Modification Task 3: Validity Assessment of Police-Reported Accident Data.
- Shinar, D., Treat, J.R., McDonald, S.T., 1983. The validity of police reported accident data. *Accid. Anal. Prev.* 15, 175–191.
- Short, J., Caulfield, B., 2016. Record linkage for road traffic injuries in Ireland using police hospital and injury claims data. *J. Safety Res.* 58, 1–14.
- Sigua, R.G., Palmiano, H.S.O., 2005. Assessment of road safety in the ASEAN region. In: *Proceedings of the Eastern Asia Society for Transportation Studies*. Bangkok.
- Simončič, M., 2001. Road accidents in Slovenia involving a pedestrian: cyclist or motorcyclist and a car. *Accid. Anal. Prev.* 33, 147–156.
- Sleet, D.A., Baldwin, G., Dellinger, A., Dinh-Zarr, B., 2011. The decade of action for global road safety. *J. Safety Res.* 42, 147–148.
- Steenberghen, T., Dufays, T., Thomas, I., Flahaut, B., 2004. Intra-urban location and clustering of road accidents using GIS: a Belgian example. *Int. J. Geog. Inf. Sci.* 18, 169–181.
- Tarko, A., Thomaz, J., Grant, D., 2009. Probabilistic determination of crash locations in a road network with imperfect data. *Transp. Res. Record: J. Transp. Res. Board* 76–84.
- Tegge, R., Ouyang, Y., 2009. Correcting erroneous crash locations in transportation safety analysis. *Accid. Anal. Prev.* 41, 202–209.
- Tercero, F., Andersson, R., 2004. Measuring transport injuries in a developing country: an application of the capture-recapture method. *Accid. Anal. Prev.* 36, 13–20.
- Thill, J.-C., 2000. Geographic information systems for transportation in perspective. *Transp. Res. Part C: Emerging Technol.* 8, 3–12.
- Tin, S.T., Woodward, A., Ameratunga, S., 2013. Completeness and accuracy of crash outcome data in a cohort of cyclists: a validation study. *BMC Public Health* 13, 1.
- Topolšek, D., Herbjaj, E.A., Sternad, M., 2014. The accuracy analysis of measurement tools for traffic accident investigation. *J. Transp. Technol.* 4, 84–92.
- Trépanier, M., Leroux, M.-H., DE Marcellis-Warin, N., 2009. Cross-analysis of hazmat road accidents using multiple databases. *Accid. Anal. Prev.* 41, 1192–1198.
- Tsui, K.L., SO, F.L., Sze, N.N., Wong, S.C., Leung, T.F., 2009. Misclassification of injury severity among road casualties in police reports. *Accid. Anal. Prev.* 41, 84–89.
- Usman Lateef, Muhammad, 2010. Estimation of fatalities due to road traffic crashes in Karachi, Pakistan, using capture-recapture method. *Asia-Pac. J. Public Health* 22, 332–341.
- WHO, 2015. Global status report on road safety 2015. Global status report on road safety. World Health Organization, Geneva 27.
- WHO, 2016. WHO regional offices [Online]. Geneva: WHO. Available <http://www.who.int/about/regions/en/> [Accessed 20th May 2016].
- Wang, C., Quddus, M.A., Ison, S.G., 2009. Impact of traffic congestion on road accidents: a spatial analysis of the M25 motorway in England. *Accid. Anal. Prev.* 41, 798–808.
- Wang, Y.-G., Chen, K.-M., Pei, Y.-L., Wang, Y., 2011. Integrating before and after crash features into measuring the effectiveness of intersection safety improvement project in Harbin. *Transport* 26, 111–120.
- Wang, Y., Zhang, C., Mao, C., 2013. Fatal motor vehicle crashes on road segments in Harbin: China: combining rates into contributory factors. *Transport* 28, 117–129.
- Watson, A., Watson, B., Vallmuur, K., 2015. Estimating under-reporting of road crash injuries to police using multiple linked data collections. *Accid. Anal. Prev.* 83, 18–25.
- Wilmot, C.G., Zhang, H., Fu, H., Jayadevan, A., Wolshon, B., Schneider, H., Sun, X., 2005. Statewide traffic safety study phase I: review of current traffic safety research, practice, analytical procedures and databases. Louisiana Transportation Research Center.
- Wilson, S.J., Begg, D.J., Samaranyaka, A., 2012. Validity of using linked hospital and police traffic crash records to analyse motorcycle injury crash characteristics. *Accid. Anal. Prev.* 49, 30–35.