Evaluation of first information reports of Delhi police for injury surveillance: Data extraction tool development & validation

Sajjan Singh Yadav1,2, Phil Edwards3 & John Porter2

1Department of Expenditure, Ministry of Finance, Government of India, New Delhi, India & Departments of 2Global Health & Development & 3Population Health, London School of Hygiene & Tropical Medicine, London, England

Received February 20, 2020

**Background & objectives:** Policymakers and health professionals need to know the distribution, patterns, trends and risk factors of injury occurrence to develop strategies that reduce the incidence of injuries. The first information report (FIR) of Indian police is one potential source of this information. The aims of this study were to identify the minimum data set (MDS) recommended for injury surveillance, to develop a tool for data extraction from FIRs, to evaluate whether FIRs contain this MDS and to assess the inter-rater reliability of the tool.

**Methods:** This was a cross-sectional study of incidents reported to Delhi Police in 2017. A systematic literature search was conducted to identify the MDS recommended for injury surveillance. A tool was designed for extraction of data, and its inter-rater reliability was assessed using Cohen’s kappa and the percentage availability of each MDS data item in the FIRs, was calculated.
**Results:** The literature review identified 24 reports that recommended 12 MDS for injury surveillance. The FIRs contained complete information on the following five MDS: sex/gender (100%), date of injury (100%), time of injury (100%), place of injurious event (100%) and intent (100%). For the following seven MDS, information was not complete: name (93.1%), age (67.2%), occupation (32.8%), residence (86.2%), activity of the injured person (86.2%), cause of the injury (93.1%) and nature of the injury (41.4%). The inter-rater reliability of the data extraction tool was found to be almost perfect.

**Interpretation & conclusions:** Information on injuries can be reliably extracted from FIRs. Although FIRs do not always contain complete information on the MDS, if missing data are imputed, these could form the basis of an injury surveillance system. However, use of FIRs for injury surveillance could be limited by the representativeness of injuries ascertained by FIRs to the population. FIRs thus have the potential to become an important component of an integrated injury surveillance system.

**Key words** First information reports - injury surveillance - inter-rater reliability - minimum data sets - police records

Unintentional injuries kill more than five million people each year globally and cause many millions to live with disability. Low- and middle-income countries (LMICs) report 90 per cent of global injury-related deaths. Injury is an important contributor to disease burden in India and is one of the leading causes of death for all ages. Injuries have a definitive causative pattern and mechanism and thus are both predictable and preventable. Policymakers and public health professionals need to know the distribution, patterns, trends and risk factors of injury occurrence. These data can help in developing public health strategies that reduce the incidence of injuries. An injury surveillance system can provide these data.

Injury surveillance systems are often based on hospital records. Such surveillance
systems tend not to ascertain all injury events and may be biased towards more severe injuries\textsuperscript{7,10}. Hospital attendance for non-fatal injuries is low, especially in LMICs; documentation of injuries is generally poor and information about the circumstances of injury is often lacking\textsuperscript{7,11}. Studies from India report that some non-fatal and minor road traffic injuries and fatalities occurring after a crash go unreported\textsuperscript{12}. Police records can be considered as one of the potential data sources for injury surveillance\textsuperscript{6,7,12}. In India, the information received by the police pertaining to a crime, including an accident, is to be recorded in a prescribed format, known as the first information report (FIR)\textsuperscript{13}. The FIRs could be a potential data source for an injury surveillance system\textsuperscript{14}. This study was undertaken to identify the minimum data set (MDS) recommended for injury surveillance, to develop a tool for the extraction of MDS data from FIRs, to evaluate whether FIRs contain this MDS and to assess the inter-rater reliability of the data extraction tool.

\textbf{<H1>Material & Methods}

This was a cross-sectional study conducted in Delhi, India, and was based on FIRs of accidents registered from January 1 to December 31, 2017. Injuries included in this study did not include psychological harms. A list of all accident FIRs was obtained from Delhi police, and FIR documents were downloaded from the Delhi Police website\textsuperscript{15}.

This study was approved by the London School of Hygiene and Tropical Medicine (LSHTM), London, UK, Observational Research Ethics Committee vide LSHTM Ethics Reference number 15992 dated November 26, 2018.

\textbf{<H2>Data extraction tool development:} A systematic search of the published and grey literature was conducted to identify MDS recommended for injury surveillance. A tool was designed for the extraction of data from FIRs. Standard classifications and codes recommended
by the WHO and other international guidelines were used in the tool\textsuperscript{16,17}. The tool was reviewed
by two injury experts and then by three professionals, each qualified at least as Masters in
Public Health\textsuperscript{18}. In a third round of development, the tool was applied by four data extractors
who were subsequently invited for a focus group discussion. A set of instructions for data
extractors when using the tool was also prepared.

\textbf{Sample size and sampling:} A random sample of 50 FIRs was selected from all 8638 FIRs
pertaining to accidents reported in Delhi in 2017: all the 8638 FIRs were serially numbered. A
list of 50 random numbers in the range of 1 to 8638 was generated from the website
random.org. FIRs having serial numbers corresponding to these random numbers were selected
for inclusion in the study. The sample size for the inter-rater reliability study was based on
published recommendations\textsuperscript{19-21}. A sample size of 50 was sufficient to allow us to estimate the
percentage availability of data items with reasonable precision (\textit{i.e.} within 13\% of the true
percentage with 95\% confidence). Data were extracted from 50 FIRs using the data extraction
tool. The percentage availability for each MDS data item with respect to each of the 58 persons
reported with injuries in these 50 FIRs was calculated.

\textbf{Estimation of inter-rater reliability:} To assess inter-rater reliability, data extraction was
first conducted by the lead author and then by one of the professionals who had tested the tool.
Cohen’s kappa coefficient (\(\kappa\)) was calculated as the measure of inter-rater reliability\textsuperscript{22,23}.
Cohen’s kappa gives a quantitative measure of the magnitude of agreement between observers
after taking into account any agreement due to chance alone. Cohen’s kappa was calculated
using the following formula\textsuperscript{23}:

\[
\text{Cohen’s kappa (K)} = \frac{(P_o - P_e)}{(1 - P_e)}
\]

Where, \(P_o=\)Proportion of observed agreement; \(P_e=\)Proportion of agreement by chance alone

The kappa values ranged from \(-1\) to 1, where 1 is perfect agreement and 0 is no agreement
beyond what would be expected by chance. Kappa values <0 indicate no agreement or poor agreement\textsuperscript{24}. We interpreted the estimates of Cohen’s kappa using the standard for strength of agreement provided by Landis and Koch\textsuperscript{24}. Accordingly, a kappa value of 0 indicates poor agreement; 0.01-0.20 indicates slight agreement; 0.21-0.40 indicates fair agreement; 0.41-0.60 indicates moderate agreement; 0.61-0.80 indicates substantial agreement and 0.81-1.0 indicates almost perfect agreement\textsuperscript{19}. After obtaining the values of kappa, its precision was quantified by calculating 95 per cent confidence intervals for each kappa value\textsuperscript{23}.

\textbf{Results}

\textit{Dataset requirements for an injury surveillance system:} The literature search yielded 24 studies including 13 national and international guidelines and data standards. These included three sets of WHO guidelines\textsuperscript{6,17,25}. The WHO guidelines recommend the following eight data items which must be collected for injury surveillance: (i) person identifier, (ii) age of the injured person, (iii) sex of the injured person, (iv) intent, (v) place of injury, (vi) nature of activity when the injury happened, (vii) cause of injury, and (viii) nature of injury\textsuperscript{6,17,25}. These eight data items were included as the MDS in our tool. In addition, if a data item was recommended as an MDS data item by the majority of the remaining 10 guidelines, it was also included in the tool.\textsuperscript{10,26-34} This yielded four further data items namely (ix) date of injury, (x) time of injury, (xi) occupation of the injured person, and (xii) residence of the injured person.

\textit{Development of the data extraction tool:} Based on the recommendations of the two international injury experts, local terms used in the tool were replaced with internationally accepted ones. The recommendations of the three professionals in public health led to reorganization of some questions and to simplification of the language of some questions. Subsequent testing of the tool by four data extractors led to further improvements: questions that were unclear were modified. More response codes were added to some questions, for
example, a response code 100 was added to indicate that a question was not relevant. The inter-rater reliability of the tool was estimated: kappa values for agreement for extraction of the MDS items between the two raters were found to be between 0.40 and 1.0, indicating between substantial and almost perfect agreement. Subsequent focus group discussion with the four data extractors highlighted a lack of clarity between some response options, too many response options for some questions, choice of multiple possible response options for a few questions and a lack of information about the type of health facility and legal status of colony as reasons for low agreement. These issues were addressed in the revised tool.

**Availability of information on minimum data set:** Data extracted from the random sample of 50 FIRs indicated that a total of 58 persons were injured in these 50 incidents reported to the Delhi Police. Results on the percentage availability of information on the 12 MDS items with respect to these 58 victims are presented in Table I. The FIRs contained complete information on 5 of the 12 MDS items namely sex/gender, date, time, place and intent. For the following four items, information was above 80 per cent complete: name (93.1%), residence (86.2%), cause of injury (93.1%) and activity (86.2%) of the injured person; for the following three items, information was above 30 per cent complete: age (67.2%), occupation (32.8%) and nature of the injury (41.4%).

The percentage availability of data varied between fatal and non-fatal injuries. Of the variables for which information was not complete, the percentage availability of data was higher in cases of fatal injuries for age, occupation, mechanism/external cause of injury, nature of injury and nature of activity (Table-II). The percentage availability of data was higher for non-fatal injuries in case of unique identification and residence of victim (Table-II). The differences in the percentage availability of data between fatal and non-fatal injuries are shown in Table II.

**Inter-rater reliability:** Cohen’s kappa values were found to range between 0.87 and 1,
indicating almost perfect agreement between the two data extractors when extracting data for
the MDS. The kappa coefficients calculated for the MDS data items along with a number of
response options in the tool, standard errors and 95 per cent confidence intervals are presented
in Table III.

**Discussion**

Twelve data items were identified which formed the MDS for injury surveillance. The
FIRs were found to contain complete information on five of the 12 MDS data items. For seven
MDS items, information was less complete. Missingness of data was substantial (i.e. >10%) for five data items and ‘slight’ for two data items.

The reliability of data extraction from the FIRs as assessed using Cohen’s kappa was
found to be ‘almost perfect’. Studies from other countries have reported on the completeness
of police records as a source of data on road traffic and other injuries, but none of those have
evaluated the suitability of police records for an injury surveillance system. No study has
so far reported on the use of police records to ascertain construction-site injuries. The
systematic extraction of data from police records using a data extraction tool and an assessment
of the tool’s reliability were done in the present study. In India, where data on injuries are not
routinely published, this study opens a new area of research on injury epidemiology. The use
of Cohen’s kappa for estimating inter-rater agreement, requires certain conditions to be met:
cases rated must be independent of each other, the raters must work independently of each
other, rating categories must be mutually exclusive and exhaustive, the sample of cases used
in the reliability data should be a random sample and the data extractors used for inter-rater
reliability are not persons who are difficult to find. All these conditions were met in this study.

One limitation of the study lied in the inherent problem of police records as a source of
injury information including underreporting, low reliability, bias towards fatal and severe
injuries and recording of information without going into the veracity of claims. Another limitation of the study was that the data extraction tool was reviewed by public health experts and not by someone from police department and experts in medicine dealing with trauma and injuries.

Missing data increase the risk of bias and may undermine the validity of research results. However, the problem of missing data is ubiquitous and unavoidable in epidemiological research. Even in developed countries, electronic health records were reported to have considerable missing data. The problem of missing data is well recognized in health surveillance systems and has been dealt with by various methods. In injury surveillance, imperfect data may still be a valuable source of information; work should therefore continue on improving the quality of these data. The challenge of missing data can be addressed using multiple imputation or full information maximum likelihood methods. Thus, although FIRs do not contain complete information on all the 12 MDS data items, they could still usefully form the basis of an injury surveillance system, provided that any missing data are imputed. Information on the MDS data item ‘nature of injury’ which was captured only in 41.4 per cent cases, may also be supplemented from other documents such as hospital records or the police record-named charge sheet which has a medico-legal report attached with it. Efforts could also be made to improve capturing of data in FIRs by training of police personnel.

FIRs may not ascertain all injuries, and the percentage of injuries ascertained by police records may be less than that ascertained by hospital records, as has been found in the UK. However, health records in India are either manual or are in disparate computer systems without interoperability or cross-sharing. FIRs are presently a better source for obtaining countrywide information on injuries because of the availability of FIRs from all over India in a centralized, web-based system named Crime and Criminal Tracking Network and Systems.
The system is now operational in 94.8 per cent of police stations in India and, 5,176,457 FIRs were registered using the CCTNS\(^4\). Moreover, although not all injuries are reported to the police, the total number of injuries can be estimated from FIRs in 2019 using methods such as capture recapture\(^4\). This will make the problem of injuries more visible to policymakers and may trigger an appropriate policy response.

As per the Sections of the Indian Penal Code dealing with unintentional injuries, only acts of negligence causing injuries to other persons are considered a criminal offence and FIRs may be registered for such acts. Any self-sustained unintentional injuries (which are not required to be reported to the police) will therefore be outside the scope of any such injury surveillance system. Moreover, access to police to register an FIR is affected by a person’s socioeconomic class and place of residence in an urban or a rural area. This can make FIRs less representative of all injuries and limit their use in an injury surveillance system.

In conclusion, information on injuries can be reliably extracted from FIRs using a data extraction tool designed in this study. Although FIRs do not always contain complete information on all the 12 data items in the MDS for injury surveillance, these may still usefully form the basis of an Indian injury surveillance system, provided that any missing data are imputed. We acknowledge the limitation that FIRs do not capture all unintentional injuries. However, in the absence of any other comprehensive data source, efforts can be made to improve the quality of data extraction and deal with missing data to make FIRs better suited for injury surveillance.

\(<H2>Acknowledgment:\</H2> Authors thank Delhi Police for granting permission and providing data for the study. We also thank Dr S. Ankush Sanghai, Ranjit Ambad, Vishnu Udan, Chintan
Parmar and Vaibhav from the Department of Health, Dadra and Nagar Haveli, for testing the data extraction tool and providing valuable feedback.

**Financial support & sponsorship:** This study was conducted as a part of the DrPH (Doctor of Public Health) programme of the London School of Hygiene and Tropical Medicine of the first author (Dr Sajjan S. Yadav), and his fee for the programme was partially paid by the Government of India.

**Conflicts of Interest:** None.

**References**


17. ICECI Coordination and Maintenance Group. *International classification of external causes of injuries*. Ver. 1.2. 2004. Published for the ICECI Coordination and Maintenance Group by Consumer Safety Institute Amsterdam, Netherlands and AIHW National Injury Surveillance Unit Flinders University, Adelaide, Australia.


*For correspondence:* Dr Sajjan Singh Yadav, House No. 1, Vinay Marg, Chanakya Puri, New Delhi 110 021, India

e-mail: sajjan.yadav@lshtm.ac.uk
Tables

**Table I.** Percentage availability for each minimum data set data item for 58 injured persons

<table>
<thead>
<tr>
<th>Data item</th>
<th>Number of injured persons for which information was available</th>
<th>Percentage availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique identification/name of the injured person</td>
<td>54</td>
<td>93.1</td>
</tr>
<tr>
<td>Age</td>
<td>39</td>
<td>67.2</td>
</tr>
<tr>
<td>Sex/gender</td>
<td>58</td>
<td>100</td>
</tr>
<tr>
<td>Occupation</td>
<td>19</td>
<td>32.8</td>
</tr>
<tr>
<td>Residence</td>
<td>50</td>
<td>86.2</td>
</tr>
<tr>
<td>Date of injury</td>
<td>58</td>
<td>100</td>
</tr>
<tr>
<td>Time of Injury</td>
<td>58</td>
<td>100</td>
</tr>
<tr>
<td>Place of injurious event</td>
<td>58</td>
<td>100</td>
</tr>
<tr>
<td>Intent</td>
<td>58</td>
<td>100</td>
</tr>
<tr>
<td>Mechanism/external cause of injury</td>
<td>54</td>
<td>93.1</td>
</tr>
<tr>
<td>Nature of injury</td>
<td>24</td>
<td>41.4</td>
</tr>
<tr>
<td>Nature of activity</td>
<td>40</td>
<td>86.2</td>
</tr>
<tr>
<td>Data item</td>
<td>Injuries reported in 50 FIRs</td>
<td>Information found available in FIRs</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td></td>
<td>Fatal</td>
<td>Non-fatal</td>
</tr>
<tr>
<td>Unique identification</td>
<td>3</td>
<td>55</td>
</tr>
<tr>
<td>Age</td>
<td>3</td>
<td>55</td>
</tr>
<tr>
<td>Sex/gender</td>
<td>3</td>
<td>55</td>
</tr>
<tr>
<td>Occupation</td>
<td>3</td>
<td>55</td>
</tr>
<tr>
<td>Residence</td>
<td>3</td>
<td>55</td>
</tr>
<tr>
<td>Date of injury</td>
<td>3</td>
<td>55</td>
</tr>
<tr>
<td>Time of injury</td>
<td>3</td>
<td>55</td>
</tr>
<tr>
<td>Place of injurious event</td>
<td>3</td>
<td>55</td>
</tr>
<tr>
<td>Intent</td>
<td>3</td>
<td>55</td>
</tr>
<tr>
<td>Mechanism/external cause of injury</td>
<td>3</td>
<td>55</td>
</tr>
<tr>
<td>Nature of injury</td>
<td>3</td>
<td>55</td>
</tr>
<tr>
<td>Nature of activity</td>
<td>3</td>
<td>55</td>
</tr>
</tbody>
</table>

FIRs, first information reports
Table III. Agreement between two data extractors in extracting minimum data set data items from fifty first information reports

<table>
<thead>
<tr>
<th>Data item</th>
<th>Number of injured persons</th>
<th>Number of response options in the data extraction tool</th>
<th>Cohen’s kappa value</th>
<th>SE</th>
<th>95% CIs for the kappa estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique identification</td>
<td>58</td>
<td>2</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>Age</td>
<td>58</td>
<td>2</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>Sex/gender</td>
<td>58</td>
<td>3</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>Occupation</td>
<td>58</td>
<td>19</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>Residence</td>
<td>58</td>
<td>2</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>Date of injury</td>
<td>58</td>
<td>2</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>Time of injury</td>
<td>58</td>
<td>2</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>Place of injurious event</td>
<td>58</td>
<td>2</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>Intent</td>
<td>58</td>
<td>2</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>Mechanism/external cause of injury</td>
<td>58</td>
<td>15</td>
<td>0.87</td>
<td>0.01</td>
<td>0.86-0.88</td>
</tr>
<tr>
<td>Nature of injury (non-fatal)</td>
<td>58</td>
<td>19</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>Nature of injury (fatal)</td>
<td>58</td>
<td>19</td>
<td>1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>Nature of activity (non-fatal injury)</td>
<td>58</td>
<td>18</td>
<td>0.97</td>
<td>0.02</td>
<td>0.80-0.98</td>
</tr>
<tr>
<td>Nature of activity (fatal injury)</td>
<td>58</td>
<td>18</td>
<td>0.97</td>
<td>0.02</td>
<td>0.95-0.99</td>
</tr>
</tbody>
</table>

CIs, confidence intervals; SE, standard error