Accident Analysis and Prevention 40 (2008) 1822–1829

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Estimating the distribution of external causes in hospital data from injury diagnosis

Kavi Bhalla[a,*], Saeid Shahraz[a], Mohsen Naghavi[b], Rafael Lozano[b,c], Christopher Murray[b]

[a] Harvard University Initiative for Global Health, 104 Mt Auburn Street, Cambridge, MA 02138, USA
[b] Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA
[c] Mexico Ministry of Health, Cuauhtemoc, DF 06600, Mexico

ARTICLE INFO

Article history:
Received 29 November 2007
Received in revised form 1 June 2008
Accepted 8 July 2008

Keywords:
Injury surveillance
Hospital discharge
Missing data
Non-fatal injuries
Bayesian inference
ICD

ABSTRACT

Hospital discharge datasets are a key source for estimating the incidence of non-fatal injuries. While hospital records usually document injury diagnosis (e.g. traumatic brain injury, femur fracture, etc.) accurately, they often contain poor quality information on external causes (e.g. road traffic crashes, falls, fires, etc.), if such data is recorded at all. However, estimating incidence by external causes is essential for designing effective prevention strategies. Thus, we developed a method for estimating the number of hospital admissions due to each external cause based on injury diagnosis. We start with a prior probability distribution of external causes for each case (based on victim age and sex) and use Bayesian inference to update the probabilities based on the victim's injury diagnoses. We validate the method on a trial dataset in which both external causes and injury diagnoses are known and demonstrate application to two problems: redistribution of cases classified to ill-defined external causes in one hospital data system; and, estimation of external causes in another hospital data system that only records nature of injuries. In comparison with age–sex proportional distribution (the method usually employed), we found the Bayesian method to be a significant improvement for generating estimates of incidence for many external causes (e.g. fires, drownings, poisonings). But the method, performed poorly in distinguishing between falls and road traffic injuries, both of which are characterized by similar injury codes in our datasets. While such stop gap methods can help derive additional information, hospitals need to incorporate accurate external cause coding in routine record keeping.

1. Introduction

Reliable injury statistics are essential for describing the public health burden of injuries, setting priorities, developing safety policies and benchmarking achievements. While deaths are often accurately documented, information about the consequences of non-fatal events is incomplete and/or of poor quality even in countries where detailed mortality statistics are routinely documented (Ameratunga et al., 2006; Annest et al., 2008). Unfortunately, mortality statistics are a poor surrogate for injury morbidity data because the leading causes of non-fatal injury hospital visits and injury deaths differ substantially (Annest et al., 2008; Bergen et al., 2008; Bhalla et al., 2008). Thus methods for accurately estimating incidence from existing sources of information about non-fatal injuries (such as emergency room records, hospital records and injury surveys) are needed.

Hospital discharge records can be a valuable source of estimating the incidence of severe non-fatal injuries (Lawrence et al., 2007; Langlois et al., 1995). Thus, the increasing computerization of hospital record keeping worldwide is presenting new opportunities to estimate the incidence and burden of non-fatal injuries. From the perspective of injury surveillance for developing prevention strategies, the most useful information in these records is the external cause (E code) of the event. However, in most countries, hospitals discharge datasets either do not include E codes, or when they do, the records are plagued with underreporting and the use of ill-defined or vague coding categories. The reasons and circumstances vary: for instance, in New Zealand private hospitals do not usually track E codes (Conner et al., 2003); hospital records in some states in the US do not assign a separate field for E codes so that patients with multiple injuries or diagnosis (e.g. the elderly) are often missing E codes (Finkelstein et al., 2006; Marganitt et al., 1990); sufficient information about external causes exists in the medical records but are not coded in discharge records (Langlois et al., 1995), perhaps because E codes are low priority as they are not needed for reimbursements (Marganitt et al., 1990); and the set of

* Corresponding author. Tel.: +1 617 496 6903; fax: +1 617 495 8231.
E-mail address: kavi_bhalla@harvard.edu (K. Bhalla).

0001-4575/$ – see front matter
doi:10.1016/j.aap.2008.07.002

Table 1
Data sources

Dataset	Description	Injury cases	Use in current analysis
SAEH[a]	2004 Subsistema Automatizado de Egresos Hospitalarios hospital discharge database (contains both E codes and N codes)	115,589 (Unintentional)	(1) Validation: used for generating Trainer and Test data. (2) Application redistribute ill-defined external causes
IMSS[b]	2004 Institute Mexicano del Seguro Social hospital discharge database (does not contain E codes)	176,185	Estimate external cause distribution of all cases
Validation-Trainer	Random sample of cases from SAEH database with known E codes	48,022	Bayes: estimate prior, conditional and marginal probabilities
Validation-Test	Random sample from SAEH	48,022	Test performance of methodology

[a] The SAEH-2004 contained records for a total of 1,795,796 patients (37% male) for all hospital visits. Unintentional injury cases were 6% of all records; 67% male; and had the following age distribution: 29.7% were ≤14 years, 60.5% were 15–64 years, and 9.8% were ≥65 years.

[b] The IMSS-2004 contained records for a total of 2,584,247 patients (33% male) for all hospital visits. Injury cases were 7% of all records; 19.1% were ≤14 years, 63.0% were 15–64 years, and 17.9% were ≥65 years.

cases with missing E codes may be biased because of difficulty in establishing intent (Winn et al., 1995).

On the other hand, hospital discharge records contain much better information about the nature of injuries (N codes) sustained by the victim (Lawrence et al., 2007). Thus, analytical methods that map N codes to the underlying E codes can allow estimation of the distribution of hospital visits by external causes enhancing the usefulness of hospital records for setting priorities and planning injury prevention programs. We develop such a tool to estimate the external cause probability distribution of each case by Bayesian updating of a prior (guess) distribution based on the set of injuries sustained by the victim. We successfully validate this method using a hospital dataset in which both external causes and the nature of injuries sustained are known. Next, we apply it to predict external causes for cases coded using ill-defined E codes in one hospital dataset, and estimate the external cause distribution of cases in another hospital dataset that does not track E codes.

2. Methods

2.1. Data sources

We used two patient-level datasets containing hospital discharge records in Mexico coded using International Classification of Diseases (ICD)-10 codes (see Table 1). The first, 2004 Subsistema Automatizado de Egresos Hospitalarios (SAEH), which is a routine hospital discharge database operated by the General Direction of Health Information of the Ministry of Health (MOH) for all MOH hospitals. It contains information on reasons for hospitalization, diagnostic and treatment procedures and results, patient care, mortality, reasons for discharge, and days at the hospital for all causes (injury and non-injury). For injury cases, the database contains ICD-10 codes describing the external cause of injuries in addition to six natures of injury diagnoses for all trauma cases. The second hospital discharge database covered all of the Instituto Mexicano del Seguro Social (IMSS) hospitals for the year 2004 and included most of the same patient-level data as the SAEH database with one critical difference—for injury-related hospital visits, the database contained three injury diagnoses but no information about external causes.

We developed and validated our methodology using the SAEH dataset because it contained information about both, the nature of injury diagnoses as well as the external causes of the injury cases. We constructed a training dataset and a trial dataset for the validation as follows. We started by only retaining cases that were due to unintentional injuries. Next, we excluded any cases classified to ill-defined or unknown ICD-10 external cause codes. Finally, we split the resulting dataset into two halves, calling the first the training dataset, and the second the trial dataset. Finally, we set the E codes in the trial dataset to missing, saving the original values in another variable for validation.

2.2. Reclassification of injuries and external causes

Four digit ICD coding results in an unwieldy number of N codes and E codes. We reduced the number of E codes by aggregating to the categories shown in Table 2. Hereafter, we refer to these as the “external cause categories” or “external causes” for short. Note that Table 2 includes three ill-defined categories, unknown road traffic injuries, unknown transport injuries, and unknown accidents, which were excluded from the validation (trainer and trial) datasets.

We reduced the number of N codes by using the ICD-10 injury mortality diagnosis (IMD) matrix (Fingerhut and Warner, 2006) with minor modifications. The IMD matrix organizes ICD-10 N codes by 16 nature of injury diagnosis categories for 17 body regions (272 cells). This matrix excluded a few ICD-10 N codes that we classified to the existing categories or to a new category, Medical Complications. (Appendix A shows a tabulation of all injuries in the SAEH dataset classified using the ICD-10 IMD matrix)

Table 2
External cause classification for unintentional injuries

External cause category	ICD 10 E-codes
RTI-pedestrian	V01–V04, V06, V09
RTI-bike	V10–V19
RTI-two wheeler	V20–V29
RTI-three wheeler	V30–V39
RTI-car	V40–V49
RTI-van	V50–V59
RTI-truck	V60–V69
RTI-bus	V70–V79
RTI-others including animal riders	V80.3–V80.5, V81.0–V81.1, V82.0–V82.1, V82.8, V82.9, V83.0–V83.3, V84.0–V84.3, V85.0–V85.3, V86.0–V86.3
RTI-unknown	V87–V89
Transport—unknown	V99
Transport non-RTI	Remainder of V
Poisons, venomous animals and plants	X40–X49, X20–X29, W53–W60
Falls	W00–W19
Drownings	W65–W74
Fires and hot substances	X00–X19
Firearm	W32–W34
Other unintentional	W20–W31, W35–W52, W61–W64, W75–W99, X20–X39, X50–X58
Accident—unknown	X59

Many hospital admissions have multiple nature of injury diagnoses. Since the pattern of multiple injuries could potentially contain important information for predicting the external cause, we used the following heuristic approach for categorizing case-level injuries.

1. Over 3000 IMD-classified injuries and injury combinations occur in the MOH dataset. However, the vast majority of these are rare outcomes (i.e. occur with low frequency in the dataset). Thus, we identified the injury and injury combinations that accounted for 90% of the cases and assigned these to our list of possible injury outcomes. This list was entirely made of single injuries and double injury combinations.
2. We reduced the remainder single and multiple injury categories by counting the number of body regions involved and the types of injuries in each category. If more than one body region and/or more than one injury type was found in the combination, it was reassigned to the cell corresponding to multiple body regions and multiple injury types in the IMD matrix.
3. Finally, we recomputed the frequency of occurrence of this reduced list of outcomes, and retained those that accounted for 99% of the cases, and classified the rest to a residual category. In all, this resulted in 112 possible injury outcomes.

2.3. Bayesian probability updating

We use Bayesian inference to infer the external cause probability distribution based on the injuries sustained. Bayesian inference relies on additional available information about an event to update a prior belief about the probability distribution. We use the age–sex distribution of the trainer dataset as the prior probability distribution for cases in each age–sex group of the trial dataset. We

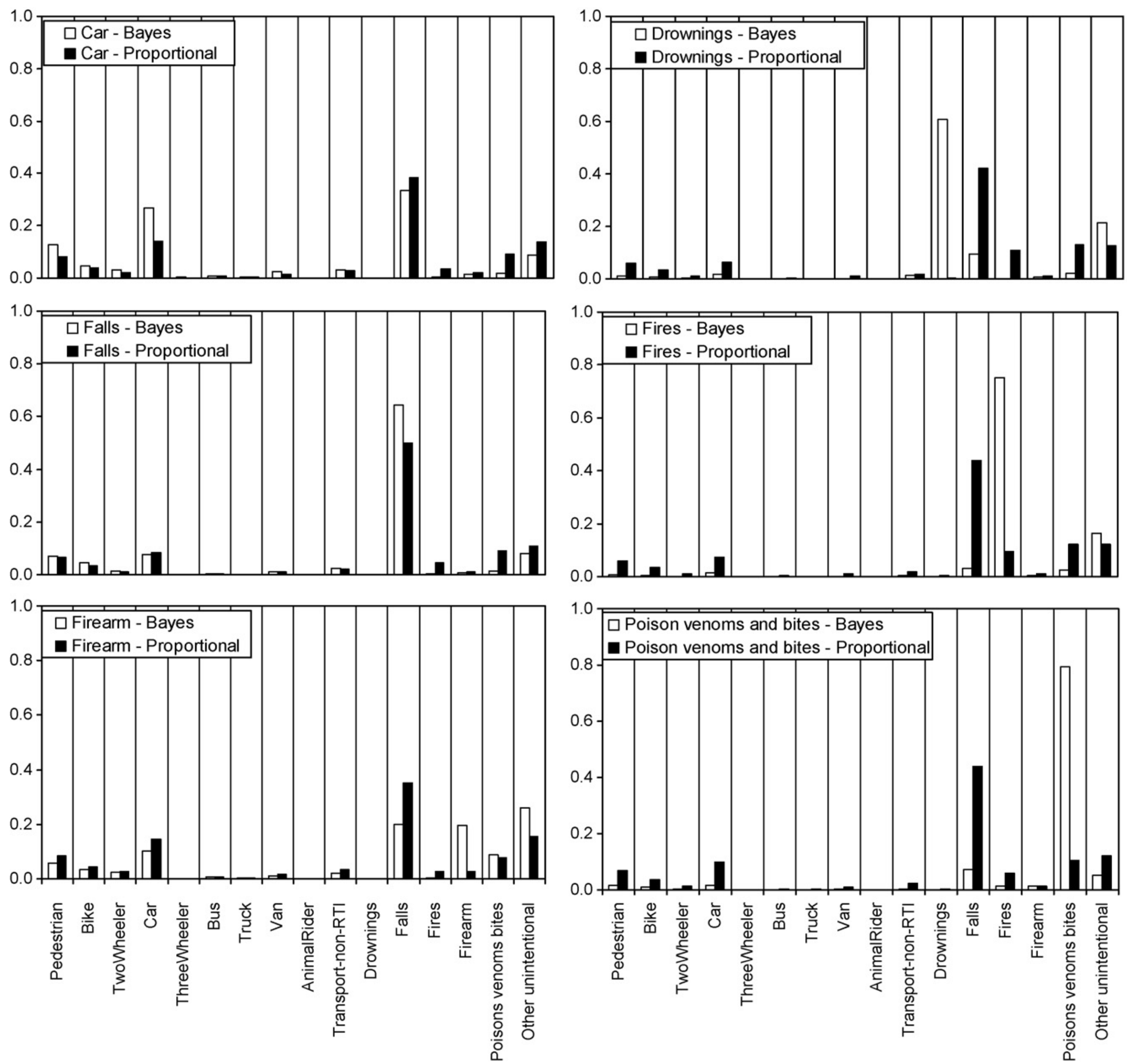

Fig. 1. Validation study: fraction of cases in the trial dataset due to each external cause that were assigned to different external causes by the algorithm. For e.g., among fire cases in the trial dataset, 75% of the set was correctly assigned to fires and 16% was incorrectly assigned to other unintentional injuries.

update this distribution based on the probability distribution of N codes. Thus, for instance, if $p(\text{external\_cause}_i)$ is the probability of the null-hypothesis (prior probability) that a case was due to the external cause i inferred before new evidence is available, $p(\text{injury}_j/\text{external\_cause}_i)$ is the conditional probability that injury j was observed for an event of external cause i, and $p(\text{injury}_j)$ is the marginal probability of injury j, then the posterior probability that the case was due to the external cause i is

$$p\left(external.cause_i \middle| injury_j\right) = \frac{p\left(injury_j \middle| external.cause_i\right)p(external.cause_i)}{p(injury_j)} \quad (1)$$

We obtain the prior probability distribution, conditional probability and posterior probability distribution from the frequency of external cause and injury occurrence in the trainer dataset and apply Eq. (1) to estimate the distribution of cases due to different external causes in the trial dataset and in the application (IMSS) dataset.

It should be noted that Eq. (1) does not uniquely assign each case to an external cause but instead provides a probability distribution. However individual identification is unnecessary for our goal of estimating the total number of cases due to external causes, which we accomplish this by summing the probabilities for each external cause over all cases.

3. Results

Fig. 1 shows the results of the validation study. In order to gauge the performance of the methodology, we have plotted the distribution of each of six external categories in the trial dataset predicted by the Bayesian algorithm. Thus, for e.g., the Bayesian algorithm assigned 75% of the set of known fire cases in the trial dataset to fires. The remaining was assigned primarily to other unintentional injuries (16%) and, in much smaller quantities to falls and poisons, venoms and bites (2% each). Similarly 79% of the set of poisons, venoms and bites, 64% of falls, and 61% of drownings were correctly predicted by the algorithm.

In comparison, the algorithm was less successful in predicting car occupant and firearm fatalities, with correct estimation of only 27% and 19% of cases respectively. In the case of car occupants, 51% of the cases were nevertheless redistributed within road traffic cases (12% were assigned to pedestrian), while the majority of the remainder (38%) were due to falls. This suggests that the types of injuries sustained do not carry sufficient information to differentiate between the different types of road traffic injuries and falls. In fact, when we compare the leading nature of injury codes for these external cause categories (Table 3), we find that the ICD-10 injury codes used in our datasets were similar. Similarly, it is likely that the loosely defined set of other unintentional injury cases contains many injuries that closely resemble firearm injuries.

Table 3
Comparison of the most common injuries among victims of falls and road traffic crashes (pedestrian and car occupants)

		Rankings by freq of occurrence		
		Falls	RTI-pedestrian	RTI-car
S069	Intracranial injury, unspecified	1	1	2
S828	Fractures of other parts of lower leg	2	6	11
S822	Fracture of shaft of tibia	3	2	5
S729	Fracture of femur, part unspecified	4	5	4
S720	Fracture of neck of femur	5	13	17
S527	Multiple fractures of forearm	6	>50	20
S424	Fracture of lower end of humerus	7	>50	41
S423	Fracture of shaft of humerus	8	7	8
S528	Fracture of other parts of forearm	9	32	31
S525	Fracture of lower end of radius	10	50	43

Fig. 1 also shows (in black) the results of applying proportional redistribution to the trial dataset based on the age–sex pattern of the validation dataset. It should be noted that the outcome from proportional distribution is different for each external category because of the difference in age–sex distribution of the categories in the trial dataset. Thus, since falls are more likely to occur in the elderly, the trial dataset contains many more elderly and proportional distribution is able to re-assign 50% of the cases back to falls. This is much than the fraction of car occupant crashes, which occur among a younger population, assigned to falls (38%).

Comparing the outcomes of the Bayesian methodology with age–sex proportional redistribution, we find that the Bayesian methodology significantly outperforms proportional distribution for each of the six external cause groups. The largest improvements are among drownings, fires, firearms, and poisons, venoms and bites. The differences are comparatively smaller when the

Table 4
Application: redistribution of unknowns in SAEH dataset and all injury admissions in IMSS dataset

	SAEH			IMSS-All	
	Knowns	Unknowns		Unknowns	
		Proportional	Bayes	Proportional	Bayes
RTI-pedestrian	6,987	1,414	1,522	12,790	10,333
RTI-bike	3,874	732	804	5,731	5,687
RTI-two wheeler	1,496	321	338	2,585	2,239
RTI-car	9,643	2,090	2,292	18,358	12,732
RTI-three wheeler	101	22	25	195	243
RTI-bus	431	91	113	839	842
RTI-truck	163	34	47	297	487
RTI-van	1,100	237	257	20,56	1,443
RTI-other	121	15	16	122	108
Transport-non-RTI	2,330	482	525	4,149	4,716
Drownings	179	28	33	184	221
Falls	43,405	8,296	8,591	82,806	87,711
Fires	4,435	806	464	6,281	4,717
Firearm	1,425	318	450	2,626	2,630
Poisons venoms bites	9,052	1,735	924	15,140	12,175
Other unintentional	11,825	2,404	2,622	20,360	28,233
Total	96,567	19,022	19,022	174,519	174,519

external cause is characterized by a well-defined age–sex pattern, for e.g. falls (elderly females) and road traffic crashes (young males).

Table 4 shows results from application of the methodology to predict the cases for which external cause was not known in the SAEH dataset and to all cases in the IMSS dataset that had at least one N code listed in the diagnosis. The fractions of cases in each category are compared with the fractions that would have been assigned using proportional distribution following the age–sex pattern of the validation dataset. The distribution of SAEH unknown cases predicted using the Bayesian algorithm is similar to that estimated using proportional redistribution. On the other hand, we find that in the IMSS dataset, the two methods result in different distributions. Bayesian redistribution predicts 24% fewer pedestrians, 44% fewer car occupants, and 6% more falls than proportional redistribution.

4. Discussion

Sound measurement of the public health burden of injuries is a key component of evidence-based policy making. Thus medical facility records can be an invaluable resource because they contain detailed injury descriptions needed for characterizing the burden of non-fatal events. Trauma registries, which are common in many high-income countries, are often designed keeping in mind the needs of injury surveillance. As a result, these have been extensively used for epidemiological research and for setting policy priorities and planning prevention activities (see, for e.g., Amoros et al., 2008; O'Connor, 2002; Welander et al., 1999). Many trauma registries pool their data into national or regional databanks (see, for e.g., NTDB; Champion et al., 1990; Tepas, 2004) making it possible to estimate population based injury rates after appropriately accounting for catchment areas (Alexandrescu et al., 2008). Thus it is essential to aid and encourage the development of trauma registries in developing countries.

On the other hand, the trend of rapid computerization of hospital administrative records in developing countries is also being driven by declining costs of database infrastructure. However, the quality of these records can be poor for injury surveillance because this was not their explicit purpose. Similar issues, although to a lesser extent, also plague hospital discharge data from high income countries as we have discussed earlier. Thus there is a need to develop analytical tools that can resolve data quality issues such as the absence of external cause codes, large amounts of missingness and cases assigned to unspecified or ill-defined codes. In this study, we present the application of a relatively simple statistical tool, Bayesian inference, for estimating the external causes of hospital admissions based on injuries sustained. We use the method to estimate the overall distribution of hospital admissions due to external causes in two Mexican hospital registries. In one dataset (SAEH) we used this methodology to identify the external causes to cases coded to unknown external causes, and in the other (IMSS), which did not record external causes at all, we predicted the distribution of external causes for the entire dataset.

Analysis of hospital datasets poses a challenge for several reasons. These datasets are often large (national hospital admissions can have millions of records), which precludes the use of some standard tools, such as the more sophisticated missing data methods that involve multiple imputations. In the framework of missingness methods, the external cause codes in hospital datasets would be considered missing values and mathematical models (typically, based on regression models) could be used for estimating these values based on the observed subject characteristics, which could theoretically include all variables included in the hospital dataset. In fact, prior to using Bayesian inference, we attempted developing a multinomial (polytomous) logistic regression model to estimate external causes based on injuries sustained by the victim. However, such regression models are computationally intensive because they require the use of iterative methods for finding the solution for a system of non-linear equations. Thus, in our application, we discovered that multinomial logistic regression was unwieldy for more than 10% of our training dataset. In comparison, the Bayesian inference method has the advantage of being computationally efficient—the analysis presented here runs to completion in a few minutes on a personal computer.

Our validation tests showed considerable success in estimating external causes that result from markedly different underlying injuries. The methodology is able to discriminate between fires, poisonings, drownings and poisons, venoms and bites. However, when the underlying injuries are similar, for e.g., falls and the various sub-categories of road traffic injuries, which are characterized by head and lower limb injuries, the method performed comparatively poorly. In addition, a listing of the most common injuries in these categories (Table 3) in the SAEH dataset revealed large numbers of unspecified injuries of the head and femur. Thus, poor coding of injury diagnosis may be part of the problem and we should expect better performance when the underlying datasets have better quality information about the predictor variables.

Dealing with multiple injuries is another aspect of the method that deserves further attention. We applied a heuristic algorithm to identify the most common injury patterns but there are other analytical approaches for handling multiple injuries. For instance, we explored one other method (results were similar and are not reported here): starting from the age–sex-based prior probabilities, we applied Bayesian inference iteratively, dealing with each IMD-classified injury independently. This requires the assumption that the occurrence of each additional injury is a statistically independent event. Although this assumption is unlikely to hold in reality, the performance of this technique was only marginally inferior to the results reported in this paper. This is because our hospital datasets were dominated by cases with single injuries and thus were affected little by fine tuning the analytical techniques for handling cases with multiple injuries.

Proportional distribution (by age and sex groups) of cases with ill-defined injury codes is usually the default method in such applications. Our validation results show that Bayesian inference was a significant improvement on proportional distribution (which was used to generate the prior probabilities) for all six of the external cause groups tested. Of course, proportional distribution should suffice if the only biases in a dataset were over representation of some age–sex groups. Bayesian inference method helps to reduce the effects of additional biases. Our application of the method to redistribute cases coded to ill-defined external causes in the SAEH dataset did not reveal a distribution significantly different from proportional. This suggests that the set of cases assigned to ill-defined external causes can be considered random conditional on the age and sex of the subject. On the other hand, when we train the analysis using SAEH data and apply to IMSS, the estimated distribution of external causes in the IMSS database was quite different from that predicted by age–sex proportional distribution. These findings indicate two important consequences. First, when estimating within a population (e.g., estimating unknowns within SAEH), age–sex proportional distribution may suffice—i.e., in our application, Bayesian inference provided little additional information. Second, the true power of Bayesian inference may lie in extrapolating to a different population. We used the SAEH age–sex specific probability of injuries by external causes as the prior guess for IMSS and then used Bayesian inference to correct these based on injury diagnoses. The change in distribution after correcting for injury diagnoses suggests that the injury profiles in the two populations (IMSS and SAEH) are different even within age–sex groups. This is to be expected because the socio-demographic characteristics of the populations that receive care by IMSS, which covers the formal sector of the Mexican economy, and SAEH, which provide health services to those outside the formal system, are likely quite different.

5. Conclusions

Accurate data on fatal and non-fatal injuries are essential for the development of interventions. Hospital discharge datasets are often the only source for estimating the incidence of non-fatal injuries in a population. However, external causes, which are one of the most important variables for guiding injury prevention priorities, are often poorly coded in hospital discharge data. We have demonstrated one way of leveraging injury diagnosis (in addition to the victim's age and sex) for estimating external causes. While the method yielded encouraging results further improvements are possible. For instance, we ignored many other variables available in the dataset, including district of residence, urban/rural, insurance status, among others. Future work should focus on optimal use of all the information available on individual records.

Of course, these methods serve only as a stop gap solution and accurate recording of external causes by hospitals should be encouraged. In many high income countries, the need to improve external cause coding in hospital records is recognized and efforts have been made in this direction for several years. Yet, progress has only been modest this far (see Abellera et al., 2004 for a review of the experience of US states). Thus, alongside these efforts to improve hospital data quality, analytical tools need to be developed to make the best possible use of the existing data systems.

Acknowledgements

This work was supported by a grant from The World Bank Global Road Safety Facility. The two hospital discharge databases (IMSS and SAEH) were provided by the Mexican Ministry of Health.

Appendix A. Injuries in SAEH dataset classified using the ICD-10 IMD matrix (Fingerhut and Warner, 2006)

		Fracture	Dislocation	Internal organ injury	Open wound	Amputation	Blood vessel	Superficial and contusions	Crushing	Burn	Foreign body	Other effects of external causes	Poisoning	Toxic effects	Multiple injuries	Other specified	Unspecified	Medical Compli-cations	Total
Head and neck	TBI	3020		18,088	2600				19						82	333	2859		27,001
	Other head	1794	70		261	39	17	1230		21	342					172	613		4559
	Neck	221	19		203		3	70	2	4	73	1			2	67	31		696
	Head and neck, other	11	2		4			18	1	741									784
Spine upper back	Spinal cord			511															511
	Vertbral column	1076	156	19												1435			2686
Torso	Thorax	1378	5	853	435	6	21	461	3	51	85	1			19	51	482		3851
	Abdomen			640	964		27	239			48								1918
	Pelvis and lower back	1440	18	397	460	3	2	242	7	3	123								2695
	Abdomen, lower back and pelvis	23		77	25	11	13	64	14			1			1	99	1305		1633
	Other trunk	28			10			12	8	985	408				6	4	16		1477
Extremities	Upper extremities	25,225	2265		2702	1148	76	243	279	1172		1			45	923	459		34,538
	Hip	5159	511		28	3		103	2							14			5820
	Other lower extremities	21,756	472		2023	336	47	349	121	963		1			42	734	368		27,212
Unclassifiable by body region	Multiple body regions	123	1		179	12	4	6245	5	432		2			23	317	3797		11,140
	System wide											407	719	6845		325			8296
	Unspecified	131			1550		29	724		4400		1			67	143	675		7720
	Other (medical complica-tions)																	1559	1559
	Total	61,385	3519	20,585	11,444	1558	239	10,000	461	8772	1079	415	719	6845	287	4624	10,605	1559	144,096

Notes: ICD-10 codes that are not valid in the US were excluded by the ICD-10 IMD matrix. We assigned these ICD-10 codes to relevant cells of the IMD matrix. This also required the creation of a new category "Medical complications". Since many victims sustained multiple injuries, the number of injuries exceeds the number of victims listed in Table 1.

References

Abellera, J., Annest, J.L., Conn, J.M., et al., 2004. How states are collecting and using cause of injury data: 2004 update to the 1997 report. Council of State and Territorial Epidemiologists, Atlanta, GA. Available at: <http://www.cste.org/pdffiles/newpdffiles/ECodeFinal3705.pdf>.

Alexandrescu, R., O'Brien, S.J., Lyons, R.A., Lecky, F.E., 2008. A proposed approach in defining population-based rates of major injury from a trauma registry dataset: delineation of hospital catchment areas (I). Trauma Audit and Research Network BMC Health Services Research 8, 80.

Ameratunga, S., Hijar, M., Norton, R., 2006. Road-traffic injuries: confronting disparities to address a global-health problem. The Lancet 367 (9521), 1533–1540.

Amoros, E., Martin, J.L., Lafont, S., Laumon, B., 2008. Actual incidences of road casualties, and their injury severity, modelled from police and hospital data. France Eur J Public Health 18, 360–365.

Annest, J.L., Fingerhut, L.A., Gallagher, S.S., Grossman, D.C., Hedegaard, H., Johnson, R.L., Kohn, M., Donna Pickett, Tomas, K.E., Trent, R.B., 2008. Strategies to improve external cause-of-injury coding in state based hospital discharge and emergency department data systems. Morbidity and Mortality Weekly Report 28, 57 (RR01).

Bergen, G., Chen, L., Warner, M., Fingerhut, L.A., 2008. Injury in the United States, 2007 Chartbook. US Department of Health and Human Services, CDC, National Center for Health Statistics, Hyattsville, MD.

Bhalla, K.B., Shahraz, S., Naghavi, M., Bartels, D., and Murray, C., 2008. Road traffic injuries in Iran, Harvard University Initiative for Global Health. Available from: <http://www.globalhealth.harvard.edu> (click on Research=> Road Traffic Injuries).

Champion, H.R., Copes, W.S., Sacco, W.J., Lawnick, M.M., Keast, S.L., Bain Jr., L.W., Flanagan, M.E., Frey, C.F., 1990. The major trauma outcome study: establishing national norms for trauma care. Journal of Trauma 30 (11), 1356–1365.

Conner, K., Langley, J., Tomaszewski, K.J., Conwell, Y., 2003. Injury hospitalization and risks for subsequent self-injury and suicide: a national study from New Zealand. American Journal of Public Health 93 (7), 1128–1131.

Fingerhut, L.A., Warner, M., 2006. The ICD-10 injury mortality diagnosis matrix. Injury Prevention 12, 24–29.

Finkelstein, E.A., Corso, P.S., Miller, T.R., Associate, 2006. The Incidence and Economic Burden of Injuries in the United States. Oxford University Press.

Langlois, J.A., Buechner, J.S., O'Connor, E.A., Nacar, E.Q., Smith, G.S., 1995. Improving the E coding of hospitalization for injury: do hospital records contain adequate documentation? American Journal of Public Health 85 (9), 1261–1265.

Lawrence, B.A., Miller, T.R., Weiss, H.B., Spicer, R.S., 2007. Issues in using state hospital discharge data in injury control research and surveillance. Accident Analysis and Prevention, 319–325.

Marganitt, B., MacKenzie, E.J., Smith, G.S., Damiano, A.M., 1990. Coding external causes of injury (E-Codes) in Maryland Hospital Discharges 1979–88: a statewide study to explore the uncoded population. American Journal of Public Health 10 (12), 1463–1466.

NTDB American College of Surgeons Committee of Trauma. National Trauma Data Bank. <www.facs.org/trauma/ntdb.html>.

O'Connor, P., 2002. Incidence and patterns of spinal cord injury in Australia. Accident Analysis and Prevention 34 (4), 405–415.

Tepas, J.J., 2004. The national pediatric trauma registry: a legacy of commitment to control of childhood injury. Seminars in Pediatric Surgery 13 (2), 126–132.

Welander, G., Ekman, R., Svanström, L., Schelp, L., Karlsson, A., 1999. Bicycle injuries in Western Sweden: a comparison between counties. Accident Analysis and Prevention 31 (1–2), 13–19.

Winn, D.G., Agran, P.F., Anderson, C.L., 1995. Sensitivity of hospitals' E-Coded data in identifying causes of children's violence-related injuries. Public Health Reports 110 (3), 277–281.