HOW MANY ROAD TRAFFIC INJURIES IN IRELAND?

Jack Short and Brian Caulfield

Trinity College Dublin

Abstract

Fatalities are the principal indicator of how dangerous roads are. Injuries and material damage get less attention but are a very significant share of the total social costs of collisions. This paper reports on work to understand better the size and nature of the injury problem due to road traffic collisions in Ireland. Data from police and hospitals are used to derive injury estimates using data linkage and capture –recapture methods. Initial results are presented and suggest that the number of traffic injuries could exceed the official number by a factor of 2 with higher factors for particular classes of casualty like serious injuries, cyclists or motorcyclists. Limitations to the methods used are examined and suggestions made for improvements in data collection and for further analysis.

Introduction

Since2005 Ireland has reduced road fatalities by more than a half and by this measure its roads are now among the world's safest. The fatality targets in the 2007-2012 road safety strategy were achieved well in advance. A new strategy was adopted in March [1] with the aim further to reduce fatalities to 25 per million by 2020, around 124 deaths. As in other Countries, the new strategy also committed to concentrate more on injuries.

Unlike fatalities, which are consistently defined and measured internationally (most countries now apply the accepted definition of death within 30 days of the crash) injuries are not well defined and their measurement has been a long standing problem. Injuries data matter, first because injuries have a large social cost – in many Countries the social costs of injuries is greater than the social cost of fatalities -and second, because safety policy faces explicit challenges to reduce injuries.

This paper first discusses the sources of data on collisions and injuries in Ireland. Then the main sources –police and hospital data are examined and compared. Data matching techniques are applied to these data to assess the degree of overlap. Capture-recapture theory is used to provide indications of the possible number of injuries in road crashes. The paper ends with some conclusions and suggestions for further work.

Data Sources

The official road crash data in Ireland are collected by the police and transmitted to the Road Safety Authority for analysis and publication. The second, though little used source is hospital admissions data which contains demographic and medical information on people hospitalised as a result of transport accidents. Other actual or potential sources of data on injuries are the Injuries Board, Insurance Companies, Accident and Emergency departments, general practitioners and Ambulance or fire brigade services. This paper concentrates on the police and hospital data.

Police data

Police recorded collisions are the basis for Irelands official data on road crashes. Injury accidents must be reported to the police in some Countries but there is no legal obligation to do so in Ireland. The circumstances under which collision and personal data are collected at the scene of a crash are neither straightforward nor a priority for the police. Subsequent data

transmission- orally at a later stage to a call centre in Castlebar- and coding are further potential sources of error. Despite the editing and cross checking before transmission by the police to RSA (which can be several months later) and the subsequent editing by RSA, the data base that is available is far from ideal, and non responses, inconsistencies and overuse of residual coding categories reduce the data's value. This said, the data base is an invaluable source that could be significantly improved. From 2014 some data transmission will be automated, and this provides an opportunity to review all the arrangements for the collection coding editing and transmission of the data. Their timeliness, accuracy and relevance will also be improved by clear assignment of responsibilities and costs between the agencies involved.

One specific data issue is the definition of injury and in particular serious injury. The definition and application varies widely between Countries.[2] As an illustration, the ratio of injuries to fatalities varies from about 150 to 20 across the OECD. In Ireland this ratio is 39, at the lower end of the scale (it is 150 in Northern Ireland). The number of serious injuries is around double the number of fatalities, a comparatively very low figure. The definition of serious injuries used by Police in Ireland is the same as in UK and New Zealand.[1] There may be some verification with hospitals but it is not systematic or required. In Ireland, the application of the definition of serious injury is a clear problem as the number actually attending hospital as in-patients exceeds the number classified as seriously injured ,though the definition implies that all hospital in-patients are seriously injured. The police data on serious injuries may be consistent but the stated definition is not applied.

In future a medical definition of serious injury will be used. Following reports and recommendations from OECD [3] and EU Ireland will move to adopt a medical definition based on the Abbreviated Injury Scale classification. This will allow consistent classification of hospital patients but will not by itself tell how many people are seriously injured.

Hospital Data

The Hospital In Patient Enquiry (HIPE) data provides information on hospital discharges. The hospitals forward the data electronically.to the ESRI. A unit there verifies and codes the data. In contrast with police data there are few non responses or blanks and the coding rules and practices are set out in a formal written way. The higher level of statistical quality is partially explained by the more controlled circumstances for the collection of data but also by the resources put into its checking, coding and editing.

Injury coding follows the International Classification of Diseases (ICD)[4] and includes information on transport and traffic injuries. The four digit ICD coding in the hospital data allows the exclusion of off road and non-traffic accidents. However, some of the ICD codes are not completely unambiguous and some specific categories that might have been excluded were actually retained in the subsequent analysis. An agreed list should be developed from the ICD codes and cross checks made from the original files. The work here should be considered as preliminary.

The hospital data records episodes of care and the same person can have more than one episode for the same incident. These are not readily identifiable. Day patients, those attending for treatment but not staying overnight, are identified separately. They have been excluded from the following analysis as otherwise they would appear several times in the base.

---

[1] "The definition of "serious injury" is an injury for which the person is detained in hospital as an 'in-patient', or any of the following injuries whether or not detained in hospital: fractures, concussion, internal injuries, crushings, severe cuts and lacerations, severe general shock requiring medical treatment."

.Accident and Emergency cases are not included unless they lead to an episode of care and if an accident victim is brought to A and E and not committed to hospital he will not appear in the hospital data. The number here is unknown as is the number who go to a doctor.

Police and Hospital Data; Some Broad Comparisons

In this section hospital and police data are compared on the basis of the adjustments made to the hospital data to exclude off road and non-traffic accidents. Table 1, provides an overview of the main aggregates over the period from 2005 to 2011.

Table 1. Police and Hospital Road Accident Data, 2005-2011.

<table>
<tr><th>Source</th><th colspan="2">Description</th><th>Number</th></tr>
<tr><td rowspan="4">Police Reported</td><td colspan="2">Serious Injuries</td><td>5,295</td></tr>
<tr><td colspan="2">Minor Injuries</td><td>53,498</td></tr>
<tr><td colspan="2">No Injuries</td><td>38,046</td></tr>
<tr><td colspan="2">Went to Hospital</td><td>53,414</td></tr>
<tr><td rowspan="4">Hospital data</td><td colspan="2">Total Injured in Traffic</td><td>31,148</td></tr>
<tr><td rowspan="3">Length of Stay</td><td>Less than 24 hours</td><td>13,031</td></tr>
<tr><td>1 to 6 days</td><td>11,502</td></tr>
<tr><td>More than 6 days</td><td>6,615</td></tr>
</table>

Sources: RSA and ESRI

Over the period police recorded over 58,000 people as injured in road collisions. Hospitals admitted around 31,000 as in-patients from traffic collisions. Police recorded over 5,000 as seriously injured while the number of people hospitalised was 5 times larger. The number of people retained in hospital for more than 24 hours (one of the other possible definitions of serious injury) was about three times larger at 18,000. The number in hospital for over six days is 6,600. This is the most extreme (in the sense of duration of stay in hospital) definition of severity that has been used internationally and is 25% larger than police- reported number of serious injuries.

The police record whether injured persons went to hospital. This number is almost the same as the number of minor injuries. But the police do not necessarily know if the person was retained as an in-patient; as it possible that the person went only to A and E and was released. This number is over 60% more than the number actually recorded in hospital. Finally, the police also record those uninjured in the injury collisions they attend and there were 38,000 people involved here.

In demographic terms, the two data sets show similar, though not identical, structures with a predominance of men and of younger age groups. Hospital data shows a higher proportion of men (65 % against 62%) and higher proportions in the under 14 years age group (15% against 8%). Most striking in both data sets is the predominance of younger people. The two sources have over 25% each in the age group 15-24 compared to 13% of the population as a whole. The individual ages between 17 and 30 are most numerous in both datasets, starkly underlining how much the road injury problem affects young people.

Table 2 below shows trends between 2005 and 2011. The table shows significant declines over the period for all indicators The more rapid decline in fatalities and police- reported serious injuries than in collisions or minor injuries indicates in a broad sense how system safety is not improving as much as the headline indicators. The main conclusion is that collisions have become less severe. This is shown also by the doubling of the ratio of minor injuries to serious injuries from around 7 to 14.

Note too the more rapid decline in police reported serious injuries than in long stay hospitalisation (defined as remaining in hospital for more than 6 days),. This could indicate a change in the way police record serious injuries but this is difficult to test at this stage. In general, the hospital data show smaller declines than the police indicators.

Table 2. Police and hospital Injury data, 2005-2011

Year	Police Data					Hospital data		
	Collisions	**Killed**	**Went to Hospital**	**Minor Injury**	**Serious Injury**	**In hospital**	**Over 24 hours**	**Over 6 days**
2005	6,533	396	7,088	7,578	1,021	5,144	2,130	1,039
2006	6,018	365	7,142	7,435	907	4,946	1,970	1,121
2007	5,467	338	8,013	6,817	860	4,864	2,050	1,001
2008	6,736	279	8,905	8,686	835	4,572	1,842	1,000
2009	6,615	238	8,653	8,884	639	4,310	1,829	897
2010	5,780	212	7,277	7,492	561	3,861	1,689	810
2011	5,230	186	6,336	6,606	472	3,451	1,521	747
Total	42,379	2,014	53,414	53,498	5,295	31,148	13,031	6,615
% decline	*19.9*	*53.0*	*17.3*	*17.6*	*57.5*	*32.9*	*28.6*	*28.1*

Sources: RSA and HIPE

.Many countries count a serious injury as being one involving more than 24hours in hospital. By that definition the number of people hospitalised with serious injuries is more than 3 times the police reported figure. Using hospital stays over six days as a measure of severity shows a number that is 25% greater than the police reported number.

Table 3 compares the data by mode and shows some additional ratios.

Table 3. Police Recorded Injuries and Hospital Data by Mode, 2005-2011.

Category	Pedestrian	Cyclist	Motorcyclist	Car	Other	Total
Serious Injury (police) (S)	857	133	415	3,310	580	5,295
Minor Injury (police) (M)	5,938	2,000	2,470	37,379	5,711	53,498
Ratio (M/S)	6.9	15.1	6.0	11.3	9.8	10.1
Went to Hospital (Police) (HP)	5,540	1,634	2,538	35,734	5,323	50,769
Hospitalised (H)	4,028	3,158	3,943	15,634	4,385	31,148
H/S	4.7	23.7	9.5	4.7	7.6	5.9
H/M	.67	1.57	1.59	.42	.77	.58
HP/H	1.38	.51	.64	2.29	1.21	1.63

Sources: RSA and HIPE.

The table shows different patterns for the different modes and strongly suggests that there is a difference between modes in the reporting of collisions. For both cyclists and motorcyclists the numbers in hospital are greater than even the police reported number of minor injuries or

the number reported to go to hospital. This clearly indicates different likelihoods for the different modes of being included in police recorded data.

The data confirm the point made by Bedford et al [5] that police recorded serious injuries are much lower than the numbers of people admitted to hospital as in-patients. Bedford reported that the police understated serious injuries by a factor of 3 to 10 depending on the mode. But this finding needs to be nuanced as hospital data does not by itself demonstrate police underreporting of injuries, except for cyclists and motorcyclists. In fact, as the table shows, the police-reported number of injuries exceeds the number of in-patients. Bedford's work should be interpreted to be more about the inappropriateness or inaccuracy of the police recorded definition of serious injury than on the real extent of the problem. The next section begins the examination of this topic.

Data Matching

Comparing the data in the way shown above provides insights into the structure and composition of the two data sets. But it does not tell us whether they contain the same people or not. This is now attempted using the technique known as record linkage, which is the methodology of "bringing together corresponding records from two or more files or finding duplicates within files" [6]. Also called data or record matching, the technique originated in the public health area when different files for the same individuals were brought together using name, date of birth and other information. Where there are unique identifiers in the two data bases, data matching is a matter of merging files only. But this is usually not the case and the record matching literature and theory developed on the basis of a set of identifiers that are not unique and are subject to error.

Data matching can provide checks on data accuracy or can combine information from different sources in order to provide a fuller picture of an event or an individual. For trauma, the short term nature of care means that a single base will rarely provide enough information for analysis or understanding over time. And the main application has been and continues to be epidemiological studies. But it is also used in developing sampling frames, in deduplication of lists, in combining survey data with data from data bases, in building panels for surveys and in Census work to verify data from subpopulations

It began with geneticist Howard Newcombe[7] who had the key insights that allowed the technique to develop. He introduced specific terminology and the decision rules for deciding on matches and non matches. His key insight is related to the ratio of the proportion of matches in linked and unlinked pairs. The technique was put on a sound mathematical footing by Felligi and Sunter [8] in 1969. Clark[9] gives a helpful exposition using Bayesian reasoning. The technique is increasingly used in public health and epidemiology and allows analyses that would otherwise be impossible or extremely expensive. It combines ideas from statistics, computer systems and operations research and has been greatly facilitated by high speed computing though it began earlier. In 1959, Newcombe reported on the ability to link records using early FORTRAN programmes at a rate of about ten per minute. Twenty years later this had increased to 14,000 records per minute and now IBM claim to be able examine records at a rate in excess of a billion per minute.

Its use in road safety research is relatively recent and serves different purposes. It has been used to make estimates of underreporting of particular kinds of accidents or injuries as in the Rhone region in France by Amoros [10]. It has allowed comparison of information on the medical consequences of accidents using the established ICD injury codes for hospital data combined with the on the spot assessment by the police. It is being used .to estimate costs, to identify associated factors in collisions and to estimate the number of severe casualties For example, the UK now uses the method to recalculate the social costs of crashes [11] and New Zealand uses it as a benchmark for the police as well as in the calculation of social costs. ITF[3] cites 16 Countries where the technique has been used in road safety. Despite its limitations, it can contribute to a better understanding of the injury problem. In addition,

the combination of information from police and hospitals is a valuable research resource on crashes and their consequences.

The terms “deterministic” “fuzzy” or “probabilistic” are used to describe the different kinds of record linkage Deterministic matching is where matches can be found directly by examination of the records using unique identifiers. Fuzzy matching is where scores are assigned for matches on different variables and the scores totalled to give a matching score. Pairs scoring above a threshold are deemed matches or links and those below another threshold non match. Scores between the thresholds can be verified clerically and assigned to one or other class. Probabilistic matching uses the mathematical theory to assign pairs as matched or unmatched or undecided. A mixture of fuzzy and probabilistic matching is used here.

Matching Police and Hospital data

The Irish data have no personal identifiers and matching is done on the basis of five variables, date of crash and hospital admission, age, sex, mode and County. None of these are perfect and even true matches can legitimately differ on them:

- Date: hospital admission can legitimately occur some days after the crash date. Here, admission one or two days after the crash is permitted.
- Age: the variables are not exact age (to protect anonymity HIPE date puts birthdate at the 15th of the month for all patients) so it is possible that age differs by one even when records are correctly filled in and truly matched.
- County: the county of collision and the county of residence (which are the variables available) do not necessarily have to be the same. While most crashes occur near home they can occur in neighbouring or other Counties .
- Mode: the classifications used are not precisely the same and there can therefore be genuinely matched pairs where the mode is not the same. This can occur mainly with the use of the residual “other” category.
- Sex: while it seems unlikely that there could be errors here but gender is not stated in 7% of the police file.

Before the matching can be carried out there is a significant amount of data preparation including;

- the police file which comprises collisions has to be converted to a file with one entry per person (incidentally, this is the way some other Countries present their road safety data and there are advantages compared to the collision approach used in Ireland);
- the variables to be compared have to be recoded so comparisons are possible;
- as described above the hospital data has to be corrected to include only road traffic accidents. This is not straightforward especially for residual codes as mentioned above;
- scores or weights need to set for matches on the different variables. These scores are loosely based on the likelihoods of matches on the variables so larger scores are given for date then age then county then mode then sex. Trials using different weights were made before a set of weights was settled on for a series of matching exercises;
- a computer programme (Linkage Wiz) was used This programme examines all combinations and selects the combination with the highest total weight.

Results of Data Matching

Table 4 summarises the results of the matching exercise between the file of police recorded injuries (over 58,000 records) compared with hospital in-patients over the years 2005-2011(31,000 records).

Scores are assigned to each match with lower scores for near matches (age differs by a year, for example). There are over 30 possible scores and a summary grouping of them is presented here for explanatory purposes.

A perfect match occurs when all five variables are matched on a record from the police file with one from the hospital file. " One variable unmatched" indicates that there was not a match on one variable (so, either age or date or sex or county or mode did not match). This brings together the matches in the highest score ranges. Similarly, weaker standards like two or three variables unmatched bring together the scores from pairs of records with increasingly unlikely chances of being true matches. This method of using Matching Standards has been applied in the literature [12]. It is convenient here as, at this early stage of analysis, it is difficult to be precise about whether matches with particular scores are true links.

Table 4. Results of Matching Police and Hospital Data by Matching Standards

Matching Standard	Frequency	Percent of Hospital file	Cumulative percent
Perfect Match	3371	10.8	10.8
One Variable Unmatched	5950	19.1	29.9
Two Variables Unmatched	6166	19.8	48.7
Three Variables Unmatched	5227	16.8	65.5
Other Combinations	5231	16.8	82.3
Total	25945	(100% = 31148)	

The analysis shows that, with a Matching Standard of at most one variable unmatched, almost 30% of the records identified in the (smaller) hospital file are to be found in the police file. This may appear very low but work in other Countries shows results that are rather similar [3],[13].

From the presentation above it is not possible to say where the thresholds to decide on true matches or non matches should be drawn. The Matching Standard method used is helpful in setting a framework and in obtaining orders of magnitude. But it does not provide insights into where the limits could be drawn.

A more detailed examination of the individual records and an application of the probabilistic theory was used to provide further insight. The application uses the simple idea that if there is a match on a rarer characteristic (like a high age or a smaller County or a less common mode), the likelihood of a true match increases. Building on this idea, a more general use was made of probabilistic reasoning to refine the estimates above.

The model uses odds ratios as in Newcombe and Felligi and applies directly the fact that the Posterior odds of a match is equal to the prior odds times the likelihood ratios (see Clark [9]for an exposition) The likelihood ratios are the ratios of what are known as the M and U probabilities where M probabilities for the matching variables are the probabilities of a match on the given variable given the two records are a true match and the U probabilities are the probabilities of a match on a variable when the variables are not matched.

Using this method the table above was reworked giving higher weights to less likely matches (like a woman cyclist in Leitrim compared to the more common male 20 year old car driver in Dublin).

This recalculation gave the following results at a posterior probability level of .85. Here we see a very significant drop in the number of likely matches in the weaker Matching Standards

and the estimate of the number of matched records of just over 10,000, is less than 1/3 of the hospital register.

Table 5. Matching Frequency by Standard using Probability Model.

Matching Standard	Frequency	Percent of Hospital Records	Cumulative Percent
Perfect Match	3,371	10,8	10,8
One Variable Unmatched	4,929	15.8	26.6
Two Variables Unmatched	1,502	4.8	31.4
Three Variables Unmatched	401	1.3	32.7
Other Combinations	162	0.5	42.2
Total	10,365	42,2	

Source: own calculations using HIPE and RSA data

Similar calculations were made for different categories of injury and for the different modes. Table 6 below provides a summary showing perfect matches, weak matches (defined as two variables unmatched) as well as the number with the posterior probability exceeding .85

Table 6. Matching Results for Different Road User Groups, 2005-2011.

Road User Group	Data Source No. of Records		Matching Model No. of Matches			Proportion Matched (prob. model)	
	Police	Hospital	Perfect	Weak(Two Variables)	Prob. model	Police	Hospital
All Injuries	58,793	31,148	3,706	25,696	10,365	11.8	39.5
Serious	5,295	31,148	921	3,911	2,455	46.3	7.9
Serious	5,295	6,615(1)	423	1,466	1,238	23.3	18.7
Cyclists	2,133	3,158	107	824	571	18.0	12.2
Motorcyclists	2,885	3,943	350	1,812	1,366	47.3	34.6
Pedestrians	6,795	4,028	637	2,230	1,994	29.3	49.5
Car occupants	40,689	16,035	2,212	11,384	6,301	15.5	39.3

Source: Own calculations from HIPE and RSA data.
(1) Over 6 days in hospital

The table shows the number of matches, using different standards (perfect, weak and the probability model) for all injuries, for serious injuries(defined as a stay of over 6 days for hospital data)and for different modes. The final columns show the proportions matched for the probability model, with the proportions being shown as shares of each file. For understanding, it is better to concentrate on the higher of the two figures in the final two columns. In no case was there more than a 50% matching rate. Highest rates were for serious injuries and pedestrians and lowest for cyclists and car occupants. Deeper understanding of the implications of these data requires further work but it can be said that the matching exercise in general shows low matching rates for most categories of road user.

The calculations were also rerun with different posterior probability levels from .99 down to .70. The number of matches varies from around 7400 at 99% to about 12,000 at 70%. These change the proportions in the final columns of the table above but do not change the results

fundamentally. The conclusion is that the two sets of records have different compositions with relatively small overlaps. But this still does not tell us the total number of injuries. This is attempted in the next section using the method known as capture-recapture.

Capture-Recapture Methods and Results

The capture–recapture method originated in animal zoology, in order to estimate a population size. To estimate the number of fish in a pond, a first catch is made and the captured fish are tagged and released. Then a second catch is made and a count is made of the fish captured for the first and second time. From the frequencies of the three sub-groups (i.e. captured at the 1st catch only, captured at the second catch only, captured at both catches), it is possible to estimate the number of fish that were never caught, and hence the total number of fish in the pond.

The method has been used also in other fields including demography and health, in the estimation of the prevalence of medical conditions and the size of hidden populations like drug users. It has been applied more recently to road traffic injury, usually focusing on subgroups, such as children or teenagers, cyclists or truck drivers. ITF [3] gives a bibliography.

The basic principle is simple. If, in the two samples A and B,N_A and N_B individuals are captured and n are captured in both samples, the estimate of the true population size N is given by $N_A N_B/n$. The reasoning is that the proportion recaptured in the second sample is representative of the entire population and so $N_A/N = n/N_B$.

While the method is attractive and simple it is not uncontroversial. Seventeen recommendations have been made by Hook and Regal [14] to correctly implement it, present the analysis itself, and display the results. But other writers remain sceptical on the possibility to achieve these conditions. In the field of injury epidemiology, some authors are also critical like Jarvis [15], Morrison and Stone [16].

The capture-recapture approach is based on some key assumptions. The main ones are (see [14] and [3]) a closed population, perfect identification of subjects common to both registrations, independence between the registrations and homogeneity of capture by a given registration.

None of these are fully met for police and hospital data and particularly dubious is that on the perfect identification of common subjects. The analysis above demonstrates how it is difficult to be certain if two records are truly matched. The condition on independence between the registrations may not be met either as for example there is a greater chance of finding police recorded seriously injured people in the hospital file. The other conditions may be more closely met. The population is closed in the sense that we are considering accidents in the same time period and therefore all accidents have the possibility to be included. Finally, on the homogeneity, there is reasonable consistency in each data set with similar criteria applying in each case.

While it is clear that the conditions for application of the capture recapture method are not met, it is more difficult to make a judgement on the consequences of this. The true number injured could be more or less than the numbers indicated by the method. In particular, where there is strong dependence between the data sets the estimates can be too low and where there is weak dependence they can be too high.

Table 7 below shows the calculations that result from application of the method.Overlap is as defined using the probabilistic method described earlier. It is seen that the estimates for the total number of injuries are from 2 to 5 times the number shown in the police data. From the remark about independence above the estimate for total injuries is likely to be an upper limit given the weak dependence between the data sets. For the specific groups (cyclists motorcylists and pedestrians) the stronger dependence could indicate under estimation.

Table 7. Capture-Recapture Estimates of Injured Population, 2005-2011.

Category	Number		Overlap (O)	Total	Capture/Recapture
	Police (P)	Hospital (H)		(P+H-O)	(PH/O)
All Injuries	58,793	31,148	10,365	79,576	176,679
Serious Injuries	5,295	31,148	2,455	33,998	67,181
Serious Injuries	5,295	6,615	1,238	10,672	28,293
Cyclists	2,133	3,158	571	4,720	11,796
Motorcyclists	2,885	3,943	1,366	5,462	8,327
Pedestrians	6,795	4,028	1,994	8,829	13,726
Car Occupants	40,689	16,035	6,301	50,423	103,546

Source: Own calculations

At first sight, the estimates might appear to be highly unlikely and call for a basic discussion on traffic injuries and the assumptions of this model. The key assumption of the method is that the recorded cases are similar to the unrecorded ones. But where are the unrecorded ones? If someone is injured, especially seriously, shouldn't they turn up in hospital? The matching analysis above shows that some do and some don't. The analysis also shows that people with minor injuries or even no injuries can be matched with hospital in- patients. But how can we know if there are injured people who appear in neither hospital nor police data? If the two sources together were complete, then the total number of injuries would be the figure shown in the Total column (H+P-O). So this is a minimum estimate for the number of injuries in the different categories, but do we know if there are more?. We do not have data and only indications. Accident and Emergency Units obviously treat a lot of people and not all of them have been recorded by the police. There are certain kinds of collisions where police may not be called like certain single vehicle collisions, collisions involving alcohol or drugs, remote collisions and cycling and motorcycle accidents where the victims manage to get to medical assistance themselves. When the police attend a collision scene they may not be able record all injuries or may not transmit them all on. Other sources also provide some indications. For example, the Injuries Board has been processing over 15,000 claims annually from people who have been in a transport collision and who were not at fault [17]. These include up to 5,000 whiplash cases which may well not be in either of the two data sets. According to the Irish Insurance Federation there were over 300,000 new motor insurance claims submitted in 2010[18]. While injury claims are not shown separately there are obviously a significant number. Finally, It should also be said that there are similar findings in other Countries notably UK [11] and Austria ,Netherlands and others as shown in the EU Safety Net project report [13].

On the other hand, there are several sources of potential error in the above including data errors, weaknesses in the matching data and the probability calculations and limitations to the application of the capture/recapture method. The conclusion from similar UK work is apposite "", … the estimates presented here should therefore be considered as broadly illustrative, rather than precise figures".[11]

Conclusions and Next steps

Injuries are difficult to define or measure and inevitably pose statistical problems. The main official source is police data but these cannot be considered as providing a full picture of road traffic casualties. Hospital data provides important complementary information both on

the numbers involved and also on the medical classification of injuries. Other sources like the Injuries Board , A and E ,ambulance and insurance data can provide insights into the extent of the problem. Obtaining a full picture requires using all sources and there remains a considerable amount of work to complete it.

At this stage the following conclusions can are drawn:

- Police injury data provide a continuous series and it would be wrong to abandon it. However, the way police data is collected, transmitted, coded edited and published could be significantly improved. A review of this would be a first step to meet Irelands aim to close the gaps with the best performing Countries;

- The definition of serious injury used by the Police is not applied in practice and understates the numbers who could be said to be seriously injured. The degree of understatement depends on the definition that is adopted and varies by road user category;

- A medical definition of serious injuries is to be adopted (MAIS 3+) but it can be applied only to hospital patients. This will bring more consistency to the figure used but will not necessarily provide a full total. Further work will be needed, not just to apply the medical definition, but to see whether adjustment factors will be needed;

- An examination of the hospital and police data sets shows they have similar demographic structures being comprised of almost two thirds men and a predominance of younger age groups. Yet there are differences and especially for cyclists and motorcyclists the numbers in hospital are higher than the police recorded numbers for even minor injuries;

- Data matching is a potentially valuable technique to identify overlap between data sets and also to bring together information on causes (police) and consequences (hospitals) of traffic collisions. There are limitations due to the insufficient specificity of the variables used. Nevertheless, it can provide indications for more detailed testing and analysis and it does allow the possibility to make new analyses on accident causes or types and consequences;

- The matching results indicate that, for the cases studied, fewer than half of those identified in one data set can be found in the other. The proportion varies by mode and severity. Police- reported serious injuries have higher matching rates then minor injuries but both are found in hospital data at all levels of severity as measured by length of stay .

- Applying the capture recapture method to the data would indicate that an estimate of the number of people injured could be at least twice and up to five times more than the number reported by the police. The conditions for use of this method are not met and these estimates should be seen as very tentative.

- On the assumption that the true number of minor and serious injuries are ,respectively twice and three times the police reported figures, and also accepting the existing assumptions and calculation methods, the estimated social costs of injury accidents in 2011 would be increased from their present figures of 162 million euros for minor injuries and 121 million for serious injuries to 324 and 363 millions respectively. This total of 687 million euros would balance the social costs of collisions more towards injuries, as in other Countries, and would also imply higher rates of return on measures that targeted injury reduction.

The work described can be developed in the following directions. First, the matching work can be strengthened in specific ways. Some matching variables can be made more precise to increase the likelihoods of finding matches; coding refinements can be made to make some variables more useful; the hospital injury coding can be verified and the medical coding can be used more; additional efforts can be made to understand the implications of

deviations from the models assumptions. Second, further matching exercises can be considered with other data sets like the Injuries Board and Ambulance data. Third, other sources, like insurance or A and E data can be explored in more detail to help provide a more complete view of the injury problem. One potential new source would be to use the national travel survey to obtain information on collisions. All of these can help improve understanding on a policy issue that will always have uncertainties because of the difficulty of fully satisfactory measurement.

Acknowledgement

The RSA and ESRI are thanked for providing the main data used here.

References

[1]Road Safety Authority (Ireland) 2013; Road Safety Strategy 2013-2020

[2] International Transport Forum IRTAD (2012) Road Safety Annual report, OECD

[3] International Transport Forum. IRTAD (2012). Reporting on Serious Road Traffic Casualties;OECD.

[4] World Health Organisation (WHO) International Classification of Diseases (ICD). available on line at http://apps.who.int/classifications/icd10/browse/2010/en

[5] Bedford Bedford, D., 2011, Sheridan, A., Howell, F., McKeown, N., Admission to Acute Hospitals for Injuries as a Result of Road Traffic Collisions in Ireland, 2005-2009.

[6] Winkler, William E. "Matching and Record Linkage". U.S.Bureau of the Census. Winkler WE (2010). *The state of record linkage and current research problems*. Available at http://eprints.kfupm.edu.sa/71379/1/71379.pdf

[7] H.B. Newcombe, J.M. Kennedy, S. Axford, and A. James, Automatic Linkage of Vital Records, Science, vol. 130, no. 3381, pp. 954-959, ...

[8] Fellegi, Ivan, Sunter, Alan (1969). "A Theory for Record Linkage". *Journal of the American Statistical Association* **64** (328): *pp.* 1183–1210.

[9] Clark DE. *Practical introduction to record linkage for injury research*. Injury Prevention 2004; 10(3):186-191.

[10] Estimating non-fatal road casualties in a large French county, using the capture-recapture method. Amoros E, Martin JL, Laumon B. Accident Analysis and Prevention, 2007 May; 39(3):483-90.

[11] Department for Transport UK,. Linking Police and Hospital data on Road Accidents in Engand; 1999 to 2009; Feb, 2012.

[12] Razzak, J. A. and Luby, S. P., 1998. Estimating deaths and injuries due to road traffic accidents in Karachi, Pakistan, through the capture-recapture method. International Journal of Epidemiology 27, 866-870.

[13] European Commission; Safety Net Report, Final Report on Task I.5;2008

[14] Hook, E. B. and Regal, R. R., 1999. *Recommendations for presentation and evaluation of capture-recapture estimates in epidemiology.* Journal of Clinical Epidemiology 52 (10), 917-926; discussion 929-933.

[15] Jarvis SN, Lowe PJ, Avery A, Levene S, Cormack RM. *Children are not goldfish--mark/recapture techniques and their application to injury data. Injury Prevention.* 2000 Mar; 6 (1):46-50.

[16] Morrison A and Stone D, Capture-recapture: a useful methodological tool for counting traffic related injuries? Injury Prevention, 2000. 6(4)299-304.
[17] Injuries Board Injuries Board,Ireland , Annual Review.
http://www.injuriesboard.ie/eng/News-Information/Articles/Injuries-Board-Annual-Review-2012.html

[18] Irish Insurance Federation. Factfile, 2010 (page 23)

http://www.insuranceireland.eu/media/documents/fact_file/Factfile%202010.pdf