Journal of the Statistical and Social Inquiry Society of Ireland
Vol. XLVII

Traffic Injuries in Ireland: a neglected problem

Jack Short*
Trinity College Dublin

(read before the Society, 12 October 2017)

Abstract: Road safety policy has focussed heavily on the reduction of the number of fatalities. However injuries are also a serious consequence of traffic collisions and they have received much less research or policy attention. This paper examines the extent of the traffic injury problem, the costs for society and appropriate policies to reduce these costs. The available data from a variety of sources are examined and assessed. The analysis then concentrates on three main data sources in Ireland: police, hospitals and the Injuries Board, from each of which anonymised datasets were available. These three sources provide different perspectives on aspects of the injury problem. All three have specific data problems and suggestions are made for improvements in the data collection and analysis processes. One conclusion is that no single data source can adequately reflect the extent and variety of the traffic injury problem. The data are analysed first as separate datasets and then together using probabilistic record linkage techniques. The results indicate that less than 20% of police-reported injuries are matched with hospital patients and less than 30% of hospital patients are matched with a police-reported injury. The linkage shows how police-recorded injuries understate the total number of injuries in the two sources by about 50% . Almost 36,000 people were hospitalised for transport accidents over the nine-year period but were not recorded by the police. Of these, over 6,400 were clinically assessed as being seriously injured, underlining how official data understate the true extent of the injury problem. The linkage also confirms that the definition of a serious injury used by police is neither an accurate nor a consistent measure. Linkage involving the three data sets shows lower matching rates than between hospitals and police, and indicates a large number of injuries in addition to those identified by police and hospital data. The total number of individual injuries in the three sources combined is more than three times the official police-reported number. Present social cost estimates understate by at least €500m annually the true social costs of injuries. These costs now exceed the social costs of fatalities. The analysis also shows how cost-benefit analysis parameters for injuries should be revised to better account for injury underestimation. These findings have significant policy implications. These include firstly, the need for improved data to better understand injuries, secondly the need for greater emphasis on injuries in the national safety strategy, and thirdly the need for specific policies to deal with groups with high injury rates like cyclists and other vulnerable users.

Keywords: Transport, Road Safety, Traffic Injuries, Social Costs, Record Linkage
JELs: R41, R42, L11, H23

1. INTRODUCTION

The benefits of mobility are taken for granted in modern societies. But mobility also has costs, with the toll of death and injury on the roads being the principal one. While there has been striking progress in reducing fatalities from the enormous levels of the early 1970s, the numbers killed and injured remain high, with over 1.2 million people killed and perhaps 50 million injured worldwide annually (WHO, 2013).

Ireland has replicated the fatality declines seen in other OECD countries and now has an annual death toll of fewer than 200, compared to over 600 in 1974. The present rate of about 42 deaths per million of population puts Ireland among the world's better performers in 2014 (IRTAD, 2016), though still some way behind Sweden, Norway, the Netherlands and the UK where these rates are below 30.

The success or otherwise of road safety policy has until now been evaluated almost entirely by the reductions in the number of fatalities. However, it is increasingly being realised that the consequences of road crashes cannot be fully or accurately summarised by the fatality total, and the numerous collisions involving injuries or material damage are a subject of increasing attention (ITF, 2012; EC, 2012; EC, 2016).

* The Road Safety Authority, the Health Service Executive and the Injuries Board are thanked for providing the data used in this paper.

There are at least five reasons why traffic injuries matter. Firstly, significant progress has been made in reducing fatalities, with many countries almost halving the number of fatalities in the last decade. Injuries, however, have not declined as rapidly (IRTAD, 2012). Secondly, the costs to society of injuries are very significant and in several countries, they are at least as large as the costs of fatalities (Department for Transport, UK, 2012b; Ministry of Transport, New Zealand 2014; SWOV, 2014). Thirdly, data on injuries are less reliable than on fatalities and are not comparable internationally. 2012). Fourthly, the larger number of injury collisions compared to those involving a fatality can provide statistically significant results for policy analysis of collision and risk factors. Finally, the policy focus on fatalities may mean that cost-effective policies to reduce injuries are not being given adequate attention.

The subject of traffic injuries is an appropriate topic for the Society, encompassing challenging statistical questions as well as having extremely important consequences for individuals and society. As the members of the Society know well, the way a problem is defined and measured influences the perception of its gravity, the policy attention it gets and the measures applied to deal with it.

The paper is structured as follows. First, there is a discussion on definitional and measurement issues. Then the main Irish sources are examined and assessed. In Section 4 the available data are analysed separately. Then in Section 5 record linkage and other techniques are used to obtain a better understanding of these data. Section 5 also includes a discussion on 2014 police data which became available only recently. The final section draws out some conclusions for policy and data development.

2. ROAD TRAFFIC INJURIES: DEFINITIONS, STATISTICAL AND MEASUREMENT ISSUES

Defining an injury is not straightforward; in the specific case of traffic injuries it can vary from a minor scratch to a life-threatening or permanently disabling condition. Some injuries are not visible, or are not immediately apparent, or may appear minor but have long lasting and debilitating consequences. Psychological trauma following accidents is common but may not be evident. Injuries have multiple and often serious consequences for the medical and insurance systems and of course for the individuals and their families. And similar injuries can have very different consequences for different individuals.

International road accident data confirm the measurement difficulty and show extremely wide variation between countries, even those with comparable fatality levels, from 150 injuries per fatality down to below20, with Ireland at about 40 (IRTAD, 2013).

A key reason for the disparities lies in the way injuries are defined and counted. For example, the definition of a minor injury used in the UK, unlike other countries, includes injuries such as cuts and bruises not requiring medical attention (Department for Transport, UK, 2012b). Persons with these kinds of lesser injuries are not usually counted as being injured in other countries.

In the transport sector, the international organisations working to harmonise statistical definitions have, until recently, defined a serious injury as one involving hospitalisation for more than 24 hours and a minor injury as one requiring medical attention (UNECE, 2009). Several countries applied this definition, including Belgium, Canada, France, Germany and Spain (ITF, 2012). Others, like the UK, New Zealand and Ireland define a serious injury as one that involves hospitalisation or is a specific kind of injury. In Ireland, 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" (RSA Factbook, 2011).

This definition is also used almost word for word by the UK (Department for Transport, 2012b) and New Zealand (Ministry of Transport, New Zealand, 2014).

Sweden uses a disability scale based on data from the insurance industry. Yet others, like Poland, define a serious injury as one involving more than 7 days in hospital. France defined a stay of more than six days in hospital as a serious injury (ONISR, 2011) but also identifies those hospitalised for over 24 hours. Increasingly countries are publishing data on hospitalisations (IRTAD, 2016). The European Commission shows 16 countries in November 2016 that can present a clinical classification of their serious injuries for 2105 (EC, 2016), though the wide variation (by a factor of 5) in injury rates shown underlines that harmonisation is still some way off.

In Ireland, therefore, the number of police-reported seriously injured people should, by definition, be larger than the number of traffic victims detained in hospital. But, as will be shown this is not the case and the police do not, in practice, apply their own definition.

In many countries the severity of road injury accidents is still determined by the police at the scene, but it is increasingly recognised that police assessments cannot be expected to give accurate or clinically comparable indications of severity (ITF, 2012). Indeed, it has been argued, for example by Cryer et al. (2001), that police should not determine injury severity and that this should be done by qualified medical staff. This view is now widely shared and is a central component of the strategies for injuries that are being developed (EC, 2012).

As far as severity measures from hospital data are concerned, several countries use variables like the number of hospital admissions and length of stay (LOS) (ITF, 2012). The LOS measures are used here but they are also problematic, being subject to administrative rules which can vary over time like availability of beds, hospitalisation policy or the age and circumstances of the person concerned (Cryer et al., 2010).

Defining thresholds for injuries is complex and has been the subject of academic and industry work for several decades; this work has involved medical practitioners, the car industry, safety agencies and academia (Association for the Advancement of Automotive Medicine, 1971; Fingerhut et al., 2004). The rapid growth in traffic injuries in the 1960s and early 1970s and the need to better understand the links between vehicle design and injuries led to a substantive body of new work on injuries. This involved developing comparable injury descriptions and classifications and resulted in the creation of the Abbreviated Injury Score and derivatives from it (Haddon, 1973; Baker et al., 1974). This work reflected the belief that severity assessment is a clinical matter requiring medical expertise and judgement.

The Abbreviated Injury Scale (AIS) is a specialised trauma classification of injuries. It was designed to distinguish between types of trauma of clinical importance as well as those important to vehicle designers and research engineers. It has been shown to provide a good basis for valid measurement of probability of death (Generelli and Wodzin, 2006). The AIS has two components, the injury descriptor and the severity score. The descriptor is a unique numerical identifier for each injury description, while the severity score ranges from 1 (minor) to 6 (untreatable) and is assigned to each injury descriptor. The AIS is based on anatomical injury, and not on any parameters of the person injured. This means that there is only a single AIS severity score for each injury, for any one person. The AIS is not a measure of impairment that results from the injury and it therefore does not measure injury consequences. But according to its proponents, it is not just a ranking of expected mortality from injury; it is based not only on probability of death but also on the diagnostic certainty, rapidity, duration, complexity and expected effectiveness of resolution (Generelli and Wodzin, 2006).

The AIS is a severity measure for a single injury. Several derivative severity scales have been developed to combine multiple injuries to create a single composite score for each patient. These include the Injury Severity Score (ISS) (Baker et al., 1974), the New Injury Severity Score (NISS) (Osler et al., 1996) and the Maximum Abbreviated Injury Score (MAIS) (Thomas et al., 2009), which is the maximum AIS of all injury diagnoses for an individual. It is one of the derivatives of the AIS. The AIS and its derivatives are now the recognised methods of measuring injury severity.

Research in the EU (Thomas et al., 2009) and OECD (ITF, 2012) resulted in a recommendation that a maximum abbreviated injury score of 3 or more (MAIS3+) be used as the definition of a serious injury. A key reason was that injuries of this kind were found to have a high probability of leading to hospitalisation, and MAIS3+ would therefore be comparable between countries. The EU adopted this recommendation (EC, 2012) and the intention was to begin to apply it from 2015. Implementing this recommendation requires that the AIS coding be applied to injuries; the AIS coding is usually not done directly– it takes time and clinical expertise. AIS coding is not undertaken in Ireland and instead, the AIS (and MAIS) scores need to be deduced from the tenth edition of the World Health Organisations International Classification of Diseases (WHO, ICD-10,2015) clinical coding that is applied to all hospital patients. There are maps that link from ICD-10 to AIS (European Centre for Injury Prevention, 2006). To the author's knowledge, these maps from the ICD-10 to the AIS have not been applied in Ireland and so far, there has been no publicly available data or research on the application of the AIS to Irish police or hospital data.

In summary, there are several dimensions to an injury. The key defining features are the causes and the consequences which the hospital clinical coding (ICD-10) tries to convey. Also important are the amounts of energy dissipated or absorbed, the threat to life and expected and actual mortality rates, the treatment time, cost and complexity if hospitalised, and the consequences for the person in terms of temporary or permanent disability or impairment and the effects on the quality of life (Generelli and Wodzin, 2006). These parameters illustrate that defining injuries and severity thresholds in a unique way risks ignoring or simplifying the many different circumstances involved.

In Ireland, as said above injury severity is assessed by the police at the accident scene or shortly afterwards and these assessments are used for official data. Whether or not medical attention was obtained would not necessarily be known to the police.

3. IRISH DATA SOURCES

In Ireland, as in many other countries, the police are the principal and official source for national and international collision data (ITF, 2012). But there are also data from hospitals and other medical sources, from insurance records and from other national, regional or local sources. For each source, the data content, quality and comparability are influenced by the purposes for which the data are collected, by the definitions and terms used as well as the manner in which they are gathered, coded, checked, and edited. These factors can have significant effects on the numbers themselves and on their interpretation, and therefore an understanding of them is important for correct use of the data.

There are several actual and potential sources for traffic injury data in Ireland and Table 1 below gives a brief summary of them. This paper concentrates on three of these sources- police, hospitals and the Injuries Board.

(a) Police Data

The police are required to make a statistical report on all road traffic accidents of which they are informed. Collisions which involve an injury and are reported to the police are the basis for the Irish official statistics, collected by the police and later published by the Road Safety Authority (RSA) in the annual road traffic collision reports (RSA Factbooks, various years). The RSA is an agency of the Irish Department of Transport, with responsibility for implementing the road safety strategy (RSA, 2013). In theory, police should be informed of all injury accidents as there is a legal requirement to report injury collisions involving a vehicle to the police (Irish Statute Book, 2015a). In practice, this does not appear to be the case and as this paper will show there are many injury accidents not reported in official data.

Table 1. Summary of Traffic Injury Data Sources in Ireland

Data Source	National Data (published by)	Availability of Datasets	Statistical Reporting Unit	Present Uses	Potential
Police	Annually (RSA)	Anonymised datasets on request	Police-reported collisions	Official Data	
Hospitals	Annually (HSE)	Anonymised datasets on request	Episode of Care	Limited	Major source of Injuries data
Injuries Board (IB)	Annually (Injuries Board)	Anonymised datasets available	Claim to Board	Limited	Could be developed significantly
Insurance	(1) Insurance Ireland (2) Central Bank	No	Insurance claims	No injury data	Complement to IB. Could show injuries
Accident and Emergency	Annual aggregates (HSE)	No	Presentation at A & E	No breakdowns	For less serious injuries
Medical Practitioners	None	No	Visit to doctor	None	For less serious injuries
Ambulance Service	None	For Dublin Area only	Ambulance trip	None	As check on police and hospital data
Health and Safety Authority	Annual (HSA)	No	Employer-reported Work-related injuries	Very limited	Specific accident types

Sources: see Section 3.

Ireland is not alone in this respect and International evidence shows that there is a substantial number of collisions that are not recorded by the police. The literature also shows that the likelihood of a collision being reported to the police is higher if there are vehicles involved and increases further if there are several vehicles, if there are serious injuries or if the incident occurs in an urban area (Amoros et al., 2006; Alsop and Langley, 2001; Jeffrey et al., 2009). As will be seen this work shows some similar results for Ireland.

The collection, compilation, editing and presentation of crash and injury data is a significant task. To start with, there are almost 200 data fields for each collision. Moreover, the circumstances of a crash are clearly not conducive to the calm collection of statistics. For the period covered by the data used in this work (2005 to 2013), there were several real and potential sources of error. These included the use of notebooks, the oral transmission of data and the delays at all stages of the process. There were no coding and editing manuals or checklists and limited editing checks for consistency once the data were entered into the police's PULSE system. For example, there are several data fields where residual responses are ambiguous (no answer, not known, other); in several cases these residual categories have the dominant share of entries. Logical contradictions between fields or incorrect codes are not edited out. As a result, the data are of lower quality than they could be. Changes were made in 2014, especially in the transmission of data from crash scenes to PULSE, in the data reception and checking procedures in Castlebar, and in the timeliness of the provision of data to the RSA. Section 5 concludes with some remarks on the data from 2104.

(b) Hospital Data

Hospital data on road traffic casualties in Ireland are a key source of injury data. The 57 acute hospitals involved in the Hospital in-Patient Inquiry (HIPE) system cover most hospital admissions and particularly the emergency admissions with which this work is concerned (HSE, 2013; ESRI, 2013). Some private hospitals and for example psychiatric hospitals are not included. Accident and emergency (A&E) cases are not included unless they lead to admission to hospital. The HIPE system, was managed by the Economic and Social Research Institute (ESRI) for the Health Service Executive (HSE) from 1990 until 2014. Since January 2014, the system is managed by the Healthcare Pricing Office (HPO), at present an agency of the HSE (HPO, 2015a).

Hospitals have specialised staff to undertake the clinical and other coding of the data. There is a set of coding manuals, instruction booklets and regularly updated information leaflets on coding (HPO, 2015b). Hospitals forward the data electronically to the HPO, where a unit undertakes the compilation and verification of the data. This unit also deals with research and other requests for extracts from the HIPE data.

The hospital data for this study were provided in anonymised form, that is there were no individual identifiers like names, addresses or social security numbers. The criterion for selection of the cases from the full HIPE dataset of all hospital discharges was that the person had been admitted as an emergency in-patient to a hospital in the HIPE system and that the person had been involved in a transport accident. Transport accidents are coded with the prefix letter V (indicating an external injury cause) in the World Health Organisations (WHO) International Classification of Diseases (ICD-10) (WHO, 1990). The data obtained contain administrative, demographic and clinical information for discrete episodes of care. The file contains about 30 variables of which about 20 are clinical codes describing the injuries. The information includes details on the type of transport accident as well as demographic and other details.

In addition to defining the transport cases of interest using the ICD V code, the analysis was restricted to emergency patients and therefore excluded day-patients (who generally make short repeated visits to hospital for treatment). Finally, HIPE data uses hospital discharges as the main statistical unit. The dataset used here is based on discharges but for consistency with the other datasets *admissions* from 1 January 2005 to the end of 2013 were used.

There are two significant statistical issues relating to the data. The first is that the statistical unit in the HIPE data is an episode of care. An episode begins on admission and ends at discharge or death. The same person may have more than one care episode for the same incident. Therefore, the number of episodes of care overestimates the number of people that were hospitalised. Estimates of return visits to hospital for the same injury can be made only in statistical terms, because of the lack of unique identifiers. This was done and matches were made on age, sex, residence, principal diagnosis and type of accident, taking account also of the time gap between admissions. The number of duplicate episodes varies depending on the criteria chosen. A range of criteria was tested. When matches were made on month and year of birth, gender, mode and principal diagnosis, about 2.5% of cases were found to be duplicates. These repeated episodes of care for the same incident were eliminated but the total length of stay was transferred to the retained record. This means that the total demand on the hospital system as measured by the number of hospital days remains correct. While unique identifiers would resolve this problem, better coding of the questions on whether the person transferred to or from a HIPE hospital could help make the matching more reliable. Even better would be a single code that indicated whether the episode was the first or a subsequent one for the incident.

The second significant statistical issue concerns the nature of the transport accident as described by the ICD external cause V codes. The V codes have ten main categories (V0 to V9) which distinguish different transport modes. Moreover further detailed coding indicates whether the accident is a traffic or non-traffic one (WHO, ICD; 2015). A traffic accident is one that occurs on the public highway and that therefore could appear in the police-reported data as it would correspond to international police-reported traffic statistic norms (UNECE, 2009). For example, accidents on mountain bike routes, or falls or collisions that occur in parks or driveways are not counted as road accidents by the police. The ICD-10 coding, at least in theory, reflects the statistical practice in the transport sector for counting road traffic accidents.

However, the distinctions between traffic and non-traffic accident occur at the fourth or even fifth digit of the V coding and the separation of the data into the two categories is not automatic. For this work, specific syntax (in SPSS) was written to categorise the cases as traffic or non-traffic. This syntax is available from the author.
Importantly, it is not certain that the coding in Ireland accurately respects the intention of the ICD-10 V coding to align with international statistical practice in defining traffic accidents. As an example in Ireland almost two thirds of cyclist hospitalisations are categorised as non-traffic accidents. This is far higher than the UK (Department for Transport, UK, 2012a). Moreover, an examination of the codes concerned leaves room for doubt on the accuracy of the coding (Short and Caulfield 2014). It is suspected that many injuries coded as non-traffic accidents should be coded as traffic accidents. A recent instruction to HSE coders has drawn attention to this issue and it will be interesting to see whether the relationship between traffic and non-traffic changes as a result. This issue is returned to in Section 5 on record linkage.

There are two further statistical issues that also need to be resolved. The first is how to determine a severity for those for whom there is no mapping from ICD-10 to AIS or for whom the mapping gives an indeterminate result. In this work only injuries with a defined injury severity from the mapping were used.
The other is how to deal with short-stay hospital patients, that is those in hospital for less than 24 hours. In many countries these are not counted as seriously injured. The data available did not allow this distinction as the shortest length of stay is 1 day and this could cover periods of hospitalisation of up to 48 hours.
For the future, and particularly for the introduction of new data series, further discussion on these issues will be required, including to take account of the efforts being made at European level to develop harmonised indicators.

The HIPE system requires substantial resources and has a rigorous approach to data collection, coding and editing. A continuous process of improvement is instituted through coding guidelines and online help (HPO, 2015b). Written records of specific coding problems are maintained and updated and new coding problems are documented and decisions shared. In these procedures, the system provides some lessons for the police data collection system where these features are less evident.

(c) Injuries Board Data
Data from the Injuries Board provide the third source of individual data for this paper. The Injuries Board was set up under the Personal Injuries Assessment Board Act of 2003, in response to concern about the high costs of injury claims and insurance in Ireland (Irish Statute Book, 2015b). While the concern was mainly about road collision claims, the Injuries Board was given a broader mandate covering work-related and public liability injuries. The concern about the high cost of insurance had also been reflected in the setting up of the Motor Insurance Advisory Board in 1998 (Motor Insurance Advisory Board, 2002).

Persons injured in an accident may make a claim for compensation. If they are unable to agree settlement with the person against whom the claim is made, they are obliged to submit their claim to the Injuries Board. There are three categories of claim: motor, public liability and workplace accidents. Motor claims are the largest share, accounting for around 60% of the claims and three quarters of the total payments made (Injuries Board, 2013).Only these motor claims are considered here. A claim is made by the “claimant” against a “respondent’’, usually represented by an insurance company. For motor claims, if the respondent is unknown or uninsured the claimant must apply to the Motor Insurance Bureau Ireland (MIBI).

Each claim has to be submitted with a payment (€45) and a medical report from their treating doctor confirming the injuries. The €45 and medical costs are reimbursed on acceptance of an assessment by the Injuries Board. While there have been references to ‘’claim farming’’ and suggestions of fraudulent types of claim (Injuries Board, 2012 Annual Report), there are no data on this topic.

Those at fault for collisions cannot make claims to the Injuries Board for their own injuries. And those who cannot identify a respondent would also be excluded. So the Injuries Board data are not a complete set of road traffic injuries. There could be cases where the respondent will be a local authority or road authority but this will be classed as a public liability claim. Similarly, falls by cyclists, common in the hospital data, would not generally be included. They could be included if, for example, the cyclist claimed that the fall was due to a pothole or some other negligence on the part of the local or road authority. However, as above, this would probably appear as a public liability rather than motor claim and would not show in the data used here.

The application to the Injuries Board leads to a number of possible outcomes as follows:

- An assessment is made and accepted by both parties and the agreed sum is paid.
- An assessment is made at 'nil' because the claim is not sufficiently founded. In this case, the claim is classed as "No Award''. According to the Injuries Board, there are virtually none in this category.
- An assessment is made but is not accepted by both parties. Such cases can then go back into litigation or can be settled separately between the parties. There is no information on the final outcomes of these cases. They are categorised in the dataset by the size of the award proposed.
- No assessment is made because the respondent does not consent to the claim being assessed by the Board. Such cases are released back into the litigation system. These can have different outcomes: settlement prior to the issue of proceedings, during the course of proceedings or on the steps of the court, or following a court hearing. These are classified as "No Award", though a payment could be made through settlement by an insurance company or following a court award.

In the data used here, therefore, the heading "No Award" covers a range of circumstances and does not imply that no compensation was made. For this paper, the size of the award is used as a proxy for injury severity. While the compensation paid is subject to many variables there should be a correlation with severity, given the clinical origins of the methodology used to assess costs (Book of Quantum, 2004).

For this study, the Injuries Board supplied anonymised individual data on claims for the years 2010 to 2013. The data fields are; the date of the incident, the date of closure of the file, the size of the award, the date of birth, gender and county or city of residence of the claimant.

There are a number of data issues that require clarification:

- A minor one is that the Injuries Board reports on claims received in a calendar year. The data used here are classified by year of incident to correspond with the other data sources. Claimants have up to two years after the incident to make a claim. Incidents occurring for example towards the end of 2010 would be included in the data used here for 2010, while it is more likely that they would be included in the 2011 official data from the Injuries Board as that is when the claim would probably be submitted to the Board.
- Claims take time to finalise; as with insurance data, the files on a particular year may not be closed for up to five years. Moreover, in general, the bigger the claim the longer it takes to be resolved. So looking at one year's data before the files are fully closed may lead to underestimates of the total costs and a bias against the largest claims. The Injuries Board is required to assess claims within 9 months of the date of consent and at present the average is 7 months (Injuries Board, 2014). Less than 3% of the claims used in this analysis had not been closed at the date of receipt of the data. And while award sizes do increase with the processing time the increase is not very steep. For these reasons, therefore, the risk of a bias may not be very significant.

The Injuries Board dataset brings further insights to the injury problem. The number of people making claims for injuries is significant, more than are seen in the hospital or police datasets. It is a system that depends on the rules for making claims and how these are interpreted, but the numbers are also affected by factors which influence whether someone will make a claim. Knowledge of the scheme, information and perceptions about whether claims can succeed and practices by insurance firms or solicitors can all influence the number of claims. Some of these factors can cause the number of claims to increase, even though other indicators may show that accidents are declining in number. It would therefore be incorrect to draw conclusions on trends in injuries or accidents from these data. As will be shown, many of the Injuries Board claims are not reported to the police and relatively few lead to hospital admissions. Many may be minor incidents but a significant number are serious enough to warrant large compensation payments and have long-term consequences for those involved. An example of an injury that would frequently not be reported to the police or lead to hospital admission is a soft tissue or whiplash injury.

These are common in the system and are believed to account for up to 80% of claims. The injuries reported to the Injuries Board have significant financial costs, as the compensation awarded for transport incidents is around €150 million per annum. This cost is reflected in the insurance premiums that are paid.

The data from the Injuries Board could be significantly more useful for safety policy and analysis if some improvements were made. A description of the incident and the modes involved would provide very useful additional information for injuries that may not appear in hospital or police sources. For the transport modes, a simple classification would be useful but the use of the ICD-10 V coding could be considered, though this is not straightforward. Moreover, since there is a medical assessment of each case using guidelines from the Book of Quantum, there is the potential to make a direct link to clinical severity measures. It would be valuable to test the possibilities by using a sample of cases to see how improved clinical information might be obtained. Recent communication with the Injuries Board indicates that these recommendations are being taken forward. This opens the possibility that Injuries Board data could become one of the cornerstones for future data on injuries.

(d) Other Sources

As Table 1 shows there are several other sources of actual or potential data on traffic injuries. Two of these sources have valuable data and the potential to assist in obtaining a full assessment of the real size of the problem. First, accident and emergency units at hospitals are obviously a crucial source of injury data but unfortunately no national data on reason for attendance at A&E units is collected. It would be useful and not demanding to add a code for reason of attendance (ideally using the WHO classification) to the data collected in A&E units. Second, injury data from motor insurance claims should be a valuable source. However, despite several requests and discussion, Insurance Ireland did not provide data on personal injury claims, declaring that they do not collect this information from their members. These data are important in themselves as they are the key to motor insurance costs; but they would also allow improved understanding of the links between the Injuries Board data and the Insurance system. It is to be hoped that the Insurance industry will begin to make efforts to contribute better to understanding the traffic injury problem.

4. WHAT THE IRISH DATA SHOW

The main data used in this paper consist of three anonymised datasets:

- Police data on collisions and injuries from 2005 to 2013 containing 75,290 cases.
- Hospital Data on transport injuries from 2005 to 2013 with 49,580 cases
- Injuries Board data on claims made from 2010 to 2013 with 70,087 cases.

Individual anonymised Injuries Board data were not available for the earlier years.

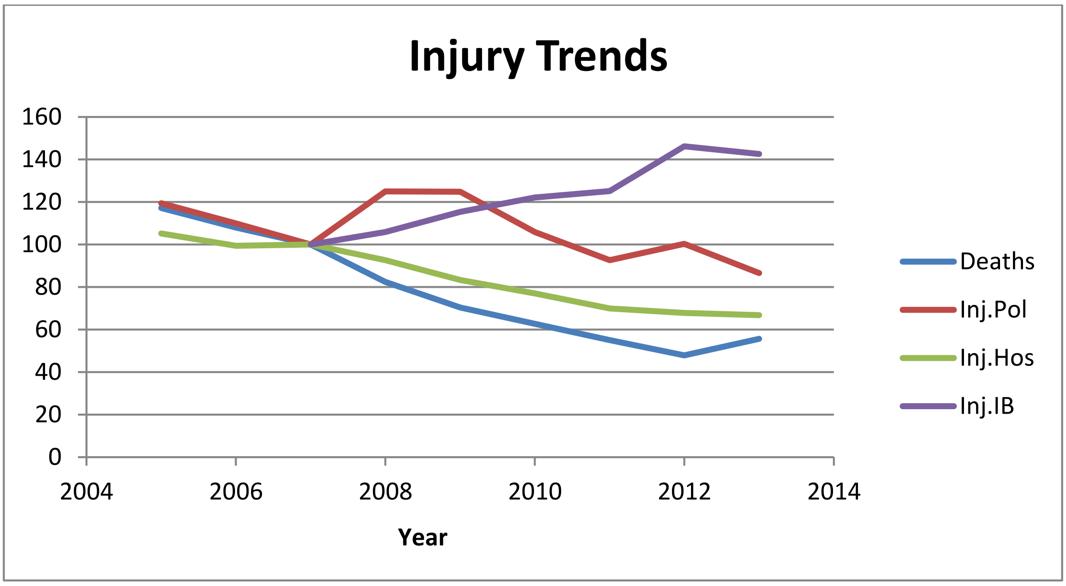
Table 2 summarises the broad aggregates from these datasets (with additional Injuries Board data also included) and Figure1 illustrates the trends in these three datasets.

Table 2. Aggregate Data from Police, Hospitals and Injuries Board, 2005-2013

Year	Police		Hospitals	Injuries Board
	Fatalities	Injuries	(Injuries)	
2005	396	9,318	6,843	
2006	365	8,575	6,464	
2007	338	7,806	6,505	13,073
2008	279	9,758	6,028	13,844
2009	238	9,742	5,423	15,079
2010	212	8,270	5,011	15,971
2011	186	7,235	4,548	16,351
2012	162	7,826	4,415	19,117
2013	188	6,760	4,343	18,648
Change 2005-2013 (%)		-27.4	-36.5	+ 42.6

Sources: RSA, Hospitals and Injuries Board.

Figure 1. Injury Trends from Different Sources (2007=100)

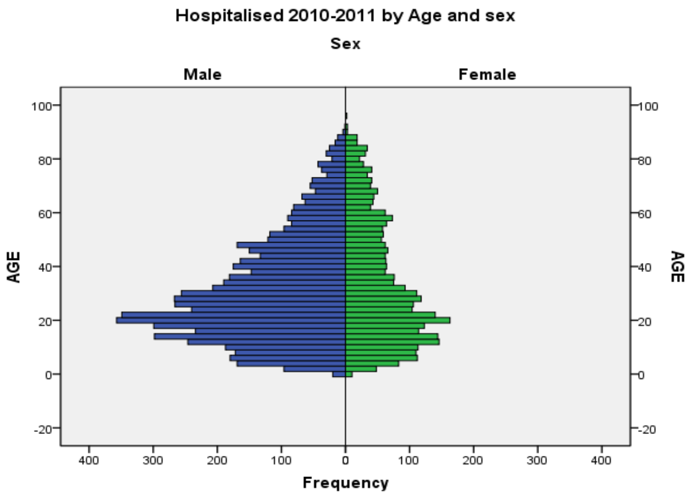


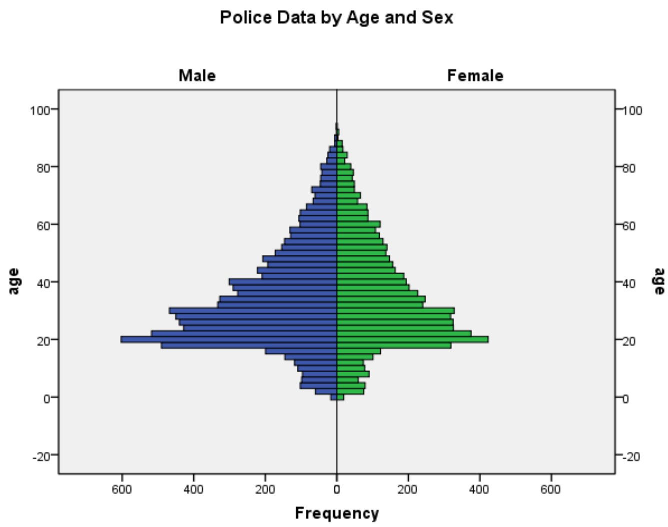
Sources: RSA, HSE, Injuries Board

It can be seen that the number of injuries board traffic claims is by far the largest with, by 2013, almost twice as many cases as in the other two combined. Over the period, police reported injuries declined by 27% and hospital admissions by 35% while in contrast Injuries Board claims increased by 42%. This increase most likely reflects factors liked the relative novelty of the Injuries Board system, a possible transfer from claims settled by insurance companies without passing through the injuries Board and public perceptions about claims and their likely success. Despite these factors and the possibility of false claims these injuries are formally recognised through legislation and need to be considered in an analysis of the injury problem.

Figure 2 illustrates the age and gender structure of the three datasets. Notice the strong concentration in the younger age groups (between 15 and 30) and the predominance of males especially in the hospital and police datasets. These are well known features internationally in traffic accident data; in fact traffic accident rates per head of population, for example using the hospital data, show risks four times higher for the younger age groups compared to middle-aged ones. The much stronger prevalence of males is mainly due to greater exposure to more dangerous modes like cycling and motorcycling and also greater distances travelled (CSO, 2011). There are also some interesting particularities as, for example, the large number of Injuries Board claims by girls under 15, possibly reflecting societal attitudes to the consequences of injuries to girls compared to boys.

Figure 2 Age and Gender structure for Three Datasets





Modal information is available only for hospital and police data and is summarised in Table 3.

Table 3. Modal Comparison, Police and Hospital Data, 2010-2013

Mode	**Police(P)**		**Hospitals(H)**		**Ratio**
	Number	**%**	**Number**	**%**	**(H/P)**
Pedestrian	3,742	*12.6*	1,867	*10.2*	.50
Cyclist	2,041	*6.9*	4,241	*23.3*	2.08
Motor Cyclist	1,282	*4.3*	1,562	*8.6*	1.22
Car Occupant	19,732	*66.4*	6,719	*36.9*	.34
Other	2,920	*9.8*	3,828	*21.0*	1.31
Total	29,717	*100*	18,217	*100*	.61

Sources: police and hospital datasets.

What is most striking in this table is the relatively much greater prevalence of cyclists and motorcyclists in hospital. Cyclists account for almost a quarter of hospitalisations, compared to 6% in police-recorded injuries and less than 1% of traffic (Short and Caulfield, 2014). More than twice as many cyclists have been hospitalised over the period 2010 to 2013 as have been reported as injured to the police. This finding that the likely understatement in police-reported injuries is more concentrated on specific modes is also found in international literature (Jeffrey et al., 2009; Amoros et al., 2006) and has significant policy implications, which will be drawn out later.

The earlier discussion has shown how thresholds for severity are difficult to define. Here, each of the three datasets has its own indicators of injury severity.

- For police data there is the police assessment on the spot, the official measure.
- For hospital data the clinical measure MAIS3+ as discussed earlier or the length of stay (LOS) can be used.
- For the Injuries Board the size of the award (for those who were assessed) is the only available indicator; different breakpoints can be taken like €100,000 or €40,000 or €20,000. The insurance industry shows a 'large claim rate' which is for claims over €100,000 and this is one option (Central Bank of Ireland, 2012). Until 2014, under the Courts Act the sum of €38,000 was the upper limit for cases that could be dealt with in the Circuit Court (Irish Statute Book, 2015c), and in a legal sense, this could be considered a plausible breakpoint. A further option is to use a breakpoint based on the relationship to household income or earnings. A figure of €20,000 can be used as it is close to average income. A case can be made for each of these.

Table 4 shows some comparisons on severity as defined for each dataset.

Table 4. Injuries and Serious Injuries for Police, Hospital and Injuries Board data, 2005-2013

Year	Police-reported Injuries		Hospitals		Injuries Board Awards	
	Total	Serious (%)	Total	MAIS3+ (%)	Total	Over €20K
2005	9,318	1,021 (11.0)	6,843	1,221 (17.8)		
2006	8,575	907 (10.6)	6,464	1,132 (17.5)		
2007	7,806	860 (11.0)	6,505	1,155 (17.8)		
2008	9,758	835 (8.6)	6,028	1,123 (18.6)		
2009	9,742	639 (6.6)	5,423	1,112 (20.5)		
2010	8,270	561 (6.8)	5,011	1,011 (20.1)	7,478	2,306 (30.8)
2011	7,235	472 (6.5)	4,548	913 (21.6)	7,162	2,315 (32.3)
2012	7,826	474 (6.1)	4,415	953 (21.2)	8,834	3,213 (36.4)
2013	6,760	508 (7.5)	4,343	921 (20.6)	8,025	2,852 (35.6)
Change (%)	-27.4	-49.8	-36.5	-24.6		

Sources: police, hospitals and Injuries Board datasets.

It can be seen that for Police data, both the number and the share of those seriously injured has more or less halved over the period shown. However, the total number of injuries has declined by only 27%. This suggests, plausibly enough on one level, that collisions have become less serious and the injuries recorded are less serious than before. On the other hand, the share of serious injuries among those admitted to hospital has remained fairly constant, even increasing a little. Assuming that the measurement of serious injuries is consistent in the hospital data this suggests something may not be quite right with the police data.

Moreover, the relatively large jumps in the police-recorded injury totals over the period (for example, in some years deaths are up and injuries down and in other years it is the other way round) are also cause for statistical suspicion. This will be discussed further in the next section.

In summary, police data show about 6% of cases as serious, hospital data about 20% with MAIS3+, and IB data about 12% of claims and about a third of awards exceeding €20,000.

Finally, in this overview of the different datasets, Table 5 looks at the injury incidence in the different counties for each of the three datasets.

The share of the national population by county is also shown. The proportions broadly follow the population shares across the datasets. Dublin is the only county with large differences, having a lower police-reported crash rate and a lower hospitalisation rate than its population share. Since the injury claim rate is above the population share this suggests fewer serious accidents involving the police or hospitalisation but more of the lesser injuries that would be common for the IB claims. The counties of Cork, Limerick and Galway also show an Injuries Board claim rate above the population share which suggests a stronger urban component to Injury Board claims with, on average, lower accident severity.

Table 5. Comparison of Injuries by County for Three Sources, 2010-2011

	Police		Hospitals		Injuries Board		Pop.
	Number	%	Number	%	Frequency	%(1)	%
Carlow	207	1.4	175	1.8	325	1.0	1.2
Cavan	362	2.4	223	2.3	494	1.7	1.6
Clare	391	2.6	227	2.4	767	2.5	2.6
Cork	1,619	10.7	1,080	11.3	3,605	11.7	11.3
Donegal	669	4.4	440	4.6	937	3.0	3.5
Dublin	3,293	21.8	1,671	17.5	8,993	29.2	27.7
Galway	939	6.2	500	5.2	1,985	6.4	5.5
Kerry	494	3.3	377	3.9	820	2.7	3.2
Kildare	614	4.1	479	5.0	1,146	3.7	4.6
Kilkenny	267	1.8	195	2.0	348	1.1	2.1
Laois	241	1.6	168	1.8	369	1.2	1.8
Leitrim	158	1	59	0.6	212	0.7	0.7
Limerick	791	5.2	377	3.9	2,286	7.4	4.2
Longford	168	1.1	103	1.1	520	1.7	0.8
Louth	551	3.6	265	2.8	1,314	4.3	2.7
Mayo	481	3.2	232	2.4	640	2.1	2.8
Meath	624	4.1	399	4.2	1,200	3.9	4.0
Monaghan	209	1.4	151	1.6	318	1.0	1.3
Offaly	255	1.7	146	1.5	441	1.4	1.7
Roscommon	321	2.1	143	1.5	327	1.0	1.4
Sligo	268	1.8	187	2.0	394	1.2	1.4
Tipperary	549	3.6	500	5.2	923	2.9	3.4
Waterford	391	2.6	251	2.6	609	1.9	2.5
Westmeath	282	1.9	226	2.4	495	1.6	1.9
Wexford	480	3.2	462	4.8	622	2.0	3.2
Wicklow	507	3.4	244	2.6	663	2.1	3.0
Foreign			273	2.9	1,032		-
Total	15,131	100	9,553	100	31,785	100	100

Sources: police, hospitals and Injuries Board datasets, CSO (2011).

In summarising the datasets we see that they show different aspects of the injuries problem and that they have different sizes and characteristics, underlining that the injury problem is not a simple one, though strongly suggesting that the official police dataset fails to capture correctly the nature and trends of the problem.

While these direct comparisons between the datasets are useful and illustrate some important issues they do not tell the full story. The next section goes further by looking at the overlaps between the datasets and by estimating the total number of people involved.

5. RECORD LINKAGE

The technique of bringing together corresponding records from two or more files is known as record linkage (Winkler, 1999). According to Felligi (1997), it began in the 1960s with the production of large files about individuals in different domains, as well as the increased role of government in data collection and analysis and the rapid development of computer technology. Where records do not have unique identifiers it allows matches to be made probabilistically on the basis of the different variables for which there is information.

The use of record linkage in road safety research and practice is relatively recent and has served different objectives. It has been used (usually together with the technique of capture–recapture) to make estimates of police underreporting of fatalities and injuries. The International Transport Forum (2012) cites 16 countries where the technique has been used in road safety. Papers from France (Amoros, Martin, & Laumon, 2007) and the Netherlands (Reurings & Stipdonk, 2011) are examples. The UK Government uses this method to calculate the social costs of crashes (Department for Transport UK, 2012a) and New Zealand uses it as a benchmark for the police as well as in the calculation of social costs (Ministry of Transport, New Zealand, 2014). The method can contribute to a better understanding of the crash problem; specifically, the combination of information from different sources can be a valuable research resource on crashes and their consequences.

Annex 1 sets out some of the mathematical and statistical background, including the idea developed in Short and Caulfield (2016) of creating a set of the “best “matches when clerical examination of the files does not allow decisions to be made on uncertain cases.

The starting point in record linkage is that there are no unique identifiers for the individuals in each of the datasets. The information available on the characteristics of the individuals like age, gender and location is used to make probabilistic calculations on whether two records, one from each dataset, relate to the same individual. The key concepts are the probabilities of matches on the specific characteristics by chance and of a true match given that certain characteristics match. For example, if two records (one each from hospital and police data) match on the county and gender this would not convincingly suggest a true match but a match on incident date, gender and age would be far more probabilistically credible.

The variables in the datasets are the dates of the crash and hospital admission, age, gender, transport mode and county of crash or residence. No other variables are common to the datasets and linkages on the basis of these five variables are the best that can be achieved with the data available.

Date of Incident: In general, hospital admission following a serious accident would occur on the same day as the accident. However, hospital admission can legitimately occur one, two or even more days after the crash date. Examination of possible matches showed few admissions more than 2 days after a crash and in this work, hospital admission one and two days after the crash is used in the linkage procedure.

Age: Police and Hospital datasets show the age of persons involved- this is a key variable in the linking process. However, data available does not show exact age (based on day, month and year of birth) and age in years only is available. For data protection reasons, birthdate in the HIPE system is assigned to the 15th of the month for all patients. Moreover, age was not stated for 2.5% of the cases in the police file.

County: The county of collision is available from the police file, and the county of residence from the hospital and Injuries Board data. These are not necessarily identical. While most crashes occur near home (from the police data base, in 85% of collisions those involved are familiar with the location) they can occur in neighbouring or other counties.

Mode: No modal information is available from the Injuries Board data. The categories in Police and Hospital datasets are not identical (for example, the motorcycle category in the hospital data includes three-wheeled vehicles, a category not mentioned in the police classification; similarly, vans, minivans, taxis and buses may not necessarily be included in comparable categories in the two datasets). Therefore, there can be genuinely matched pairs where the indicated mode is not the same. In addition, there are significant numbers in residual categories, especially in the hospital data for this variable.

Gender: This is an obvious variable for linkage. However, gender was not reported in as many as 5% of the cases on the police file.

Twenty studies cited by the ITF (2012) showed that the variables used in the linkage described here for Ireland are similar to those used in record linkage exercises for other countries. Some of these studies have additional personal information like names or initials or precise dates of birth, as for example in the United States (Johnson and Walker, 1996) or in Australia (Ferrante et al., 1993). Other countries like the Netherlands have additional items of information, for example on the location of the hospital, on the time of admittance to hospital and time of collision (Reurings and Stipdonk, 2011). By international standards for linkage exercises, while variables available in Ireland are similar, the number of variables used is often fewer and, for the reasons just set out, their precision is not any better and is likely to be worse particularly for the location variable. This implies that results of data linkage in Ireland risk to be less accurate than those cited above.

The methodology has several steps to prepare and edit the data, to obtain comparable variables and to test different scoring schemes. It followed, with some refinements, the recommendations set out in the International Working Group for Disease Monitoring and Forecasting (lWGDMF 1995a, 1995b). Further details are in Short (2016).

Box 1. Method to Match Police Hospital and Injuries Board Datasets

1. The data are the police, hospital and Injuries Board anonymised datasets for 2010 to 2013.

2. The datasets are matched pairwise probabilistically using the specialised software LinkageWiz (LW). For each pair of datasets, slightly different variables and weights are used for the matching. The preparation of the datasets requires consistent variables and the assignment of unique reference numbers for each record.

3. The Linkage-Wiz output files are exported to SPSS and are refined using specific syntax, resulting in three pairwise datasets of ''best matches'' using the theory developed in Annex1. These three sets of "best matches" are between police and hospitals, police and Injuries Board and hospitals and Injuries Board.

4. These three sets of pairs or ''doubles'' are then combined and matched deterministically using SPSS matching software on the basis of the unique reference numbers. This allows "triples"- individuals in all three datasets- to be found. When a triple is found from two different pairs, one of the pairs is dropped and needed information is transferred to the retained record. This file of "doubles" and "triples" is combined successively with the three individual datasets and duplicate records are deleted. This gives the file of unique individuals involved in an accident and recorded in at least one of the datasets.

6. RESULTS

Two main sets of results are presented - the matching of police and hospital data over the period 2005-2013 and of police hospital and Injury Board data over the years 2010 to 2013.

(a) Police Hospital matching for the period 2005 to 2013

Table 6 shows the number of matches, or overlaps between the two datasets for each of the categories shown. The total expected number of matches is almost 14,000, equivalent to 19% of the police file and 28% of the hospital file. For most characteristics, the overlaps between the datasets account for no more than a third of the hospital cases and 20% of the police-recorded cases. From international data these orders of magnitude are not unusual and have been seen in several countries including the UK (Department for Transport, UK, 2012a), New Zealand (personal communication) and other countries (EC, 2008; ITF,2012). This general finding that only a third of patients hospitalised as a result of a transport accident have been recorded by the police is a clear indication of how official data understate the total number of injuries.

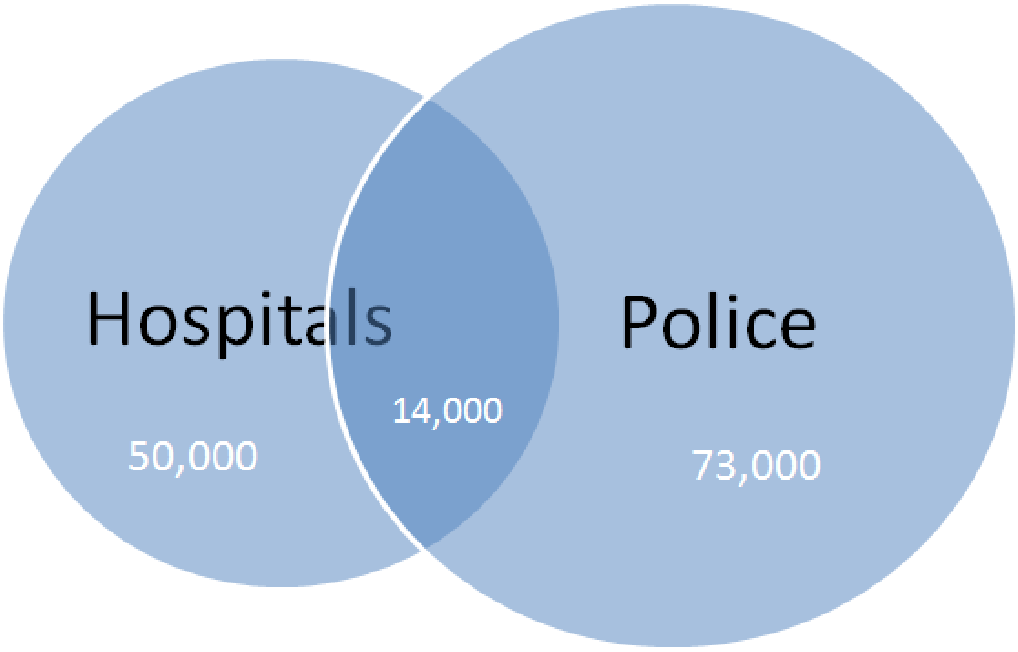
Table 6. Matching Results for Police and Hospital Data, 2005-2013

Category	No. of Cases				Match Rates[1] (%)	
	Police	Hosp.	Overlap	Total	Police	Hosp.
Male	39,207	32,057	8,801	62,463	22.4	27.4
Female	29,097	17,522	4,970	41,649	17.1	28.4
Pedestrian	8,690	5,327	2,437	11,580	28.0	45.7
Cyclist	3,392	8,707	784	11,315	23.1	9.0
Motor-C	3,494	4,562	1,239	6,817	35.5	27.2
Car	50,134	21,058	8,333	62,859	16.6	39.6
Other	7,670	9,926	1,114	16,482	14.5	11.2
2005	8,600	6,842	1,704	13,738	19.8	24.9
2006	8,342	6,464	1,795	13,011	21.5	27.8
2007	7,677	6,505	1,646	12,536	21.4	25.3
2008	9,521	6,028	1,746	13,803	18.3	29.0
2009	9,563	5,423	1,766	13,225	18.5	32.6
2010	8,053	5,011	1,514	11,550	18.8	30.2
2011	7,078	4,548	1,338	10,288	18.9	29.4
2012	7,826	4,415	1257	10,984	16.1	28.5
2013	6,760	4,343	1,143	9,960	16.9	26.3
Traffic		27,250	10,063			36.9
Non-Traffic		22,330	3,365			15.1
Serious (P)	6,277		3,216		51.2	
Minor (P)	67,103		10,693		15.9	
MAIS 3+		9,541	3,100			32.5
MAIS<3		28,825	7,941			27.5
0-14	5,890	9,954	1,748	14,096	29.7	17.7
15-24	18,721	12,504	3,908	27,317	20.9	31.2
25-34	15,781	8,961	2,399	22,343	15.2	26.8
35-44	10,613	5,984	2,067	14,530	19.5	34.5
45-54	7,403	4,510	1,427	10,486	19.3	31.6
55-64	4,961	3,299	990	7,270	20.0	30.0
65+	4,970	4,367	1,367	7,970	27.5	31.3
Dublin	14,978	8,754	2,215	21,517	14.8	25.3
Cork	7,577	5,513	1,834	11,256	24.2	33.3
Galway	4,107	2,768	708	6,167	17.2	25.6
Donegal	3,618	2,511	882	5,247	24.4	35.1
Kildare	2,783	2,096	468	4,411	16.8	22.3
Limerick	3,922	1,741	551	5,112	14.0	31.6
Wicklow	2,261	1,279	310	3,230	13.7	24.2
Wexford	2,625	2,234	685	4,174	26.1	30.7
Others	31,499	22,684	6,256	47,927	19.8	27.6
Total	***73,370***	***49,580***	***13,909***	***109,041***	***19.0***	***28.1***

Sources: police, hospitals and Injuries Board datasets; matching software.

The total number of distinct injury cases in the two datasets together is also shown (this is the sum of the police and hospital totals less the overlap). This total number of distinct injuries between the two sets is over 109,000, comprising almost 73,000 from the police, 50,000 from hospitals less the 14,000 in both datasets. This is illustrated in Figure 3.

Figure 3 Injury Data for Police and Hospitals, 2005-2013



When comparable data are available from both datasets, the police-hospital match rate is the overlap divided by the total number in that category. When the data are available in one dataset only (like MAIS3+ for hospital data) the match rate is the overlap as a percentage of the known total. So for MAIS3+, the figure of 32.5% in the Hospital Match rate column indicates that this percentage of MAIS3+ hospital patients were matched with a police-recorded injury, showing how it was slightly more likely that a clinically seriously injured person would be recorded by the police.

There is significant variation in the proportions that are matched and many of the differences in proportions are statistically significant (with the large sample sizes, differences of over 2 percentage points from the average are statistically significant at the 95% level). The highest match rate is for police-recorded serious injuries where over 50% are linked with a hospital patient.

The linkage shows how a significant number of those hospitalised are not recorded by the police. The evidence here indicates that over the period 2005 to 2013, there were over 6,400 people who were clinically seriously injured (MAIS3+) who were not reported to police. This number is roughly the same as the total police-reported number of serious injuries over the period. In addition there were almost 21,000 hospitalised people (with MAIS<3) and a further 6,800 hospitalised people without a clinical severity rating who were not reported to the police. Even though Irish law requires that road accident injuries are reported to the police (Irish Statute Book, 2015a, Road Transport Act 1961) the evidence here is that a large number of road traffic injuries are not reported. The total of almost 36,000 people hospitalised but not recorded by the police over the nine-year period demonstrates the additional cost of road traffic accidents for society beyond those indicated by official data.
Cyclists illustrate the differential in matching rates most strikingly, with only 9% of the hospitalised cyclists being matched with a police-recorded injury (Short and Caulfield, 2014). Here the coding problem is significant as the hospitalised figures include mountain bike and other off-road crashes that should not be included as road traffic accidents. This issue needs further work and the coding needs to be improved as many of those classed as non-traffic accidents in the HIPE system were matched with police cases. But even if all HIPE coded non-traffic accidents were correct, there remains a significant injury understatement in police data. This problem has been noted by HIPE coders and an instruction has recently been issued to make the coding correspond with the intentions of the ICD system which is that the default code is a traffic accident.

Relatively low match rates in one source can be accompanied by relatively high ones in the other. For young people, for example, the hospital match rate is low (17.7%) and the police one is high (29.7%). The low hospital match rate suggests a relatively higher understatement in official data; a high police match rate indicates that this category has a relatively higher probability of being hospitalised. Car occupants provide an example. Car drivers and passengers in hospital are more likely to be matched with police-recorded injuries than other modes except for pedestrians. In contrast, police-recorded car occupants with injuries are less likely to be matched to hospital patients than other categories of road user. This shows that there are many car occupant injuries which do not result in hospitalisation. But it also demonstrates that a higher proportion of hospitalised car occupants have been reported to the police than other hospitalised traffic casualties.

In summary, in police data, if a person is seriously injured or is from Cork or is over 65 years or under 14 or a motorcyclist or pedestrian then they have a statistically significant higher likelihood to be matched. In the hospital data, the following categories are associated with a statistically significant higher probability of being matched: pedestrians, patients from Cork, those with serious clinical injuries (MAIS3+), traffic injuries and car occupants. Cyclists and other road users have lower matching rates. The extremely low cyclist-matching rate for hospital patients has been pointed out already (Short and Caulfield, 2014) and cycling is the mode where police understatement of the number of injuries is greatest.

Table 6 also shows that there are higher matching rates in Cork and lower ones in Dublin. This could be due to a lower likelihood of data complications in Cork where there are fewer hospitals and the location variable problem mentioned above of a difference between county of crash and county of residence might be less than around Dublin.

One useful aspect of the linking is that it allows additional information in the matched set on variables contained in only one of the datasets. Specifically, it allows comparisons on the assessed injury severity from both datasets as in Table 7 below.

Table 7. Severity Comparisons for Police and Hospital Matched Data, 2005-2013

		Police Reported Injuries		
		Serious (%)	**Minor (%)**	**Total(%)**
Hospital Injuries	MAIS3+	1,360 (12.6)	1,740 (16.1)	3,100 (28.7)
	MAIS<3	1,502 (13.9)	6,193 (57.4)	7,695 (71.3)
	Total	2,862 (26.5)	7,933 (73.5)	10,795 (100)

Sources: police and hospital datasets; matching software and own calculations.

Table 7 shows that for the matched sets over the period, the severity indicators correlate but by no means perfectly. For those assessed as seriously injured by the police, it is equally probable that they will be assessed as clinically seriously injured or not seriously injured. Over 20% of those classed with a minor injury by the Police have clinically serious conditions. Equivalently, more than half of those with clinically serious injuries have been reported by the police as having minor injuries.

Taking the comparison of severity assessments further, Table 8 examines, for the matched sets over the period 2005 to 2013, how police-reported injuries are assessed clinically.

Table 8. Police and Hospital Assessments of Serious Injuries for Matched Data, 2005-2013

Year	Police-reported Serious Injuries			Police-reported Minor Injuries		
	Total	Matched	MAIS3+ (%)	Total	Matched	MAIS3+
2005	1021	400	154 (38.5)	7,579	918	164 (17.9)
2006	907	406	183 (45.1)	7,435	1,009	199 (19.7)
2007	860	362	152 (45.3)	6,817	935	168 (18.0)
2008	835	349	164 (46.4)	8,686	949	212 (22.3)
2009	639	329	169 (51.1)	8,884	1031	243 (23.6)
2010	561	298	152 (51.0)	7,492	867	194 (22.4)
2011	472	254	127 (50.0)	6,606	792	210 (26.5)
2012	474	226	117 (51.8)	7,352	748	175 (23.4)
2013	508	238	133 (55.9)	6,252	684	175 (25.6)

Sources: police and hospital datasets; matching software and own calculations.

It can be seen that for those reported as seriously injured by the police, the proportion that is clinically serious (MAIS3+) has increased from 38.5% in 2005 to 55.9% in 2013. At the same time the proportion of those with minor injuries that is clinically serious has increased from 17.9% to 25.6%. These facts together show how the police assessment of severity is not consistent with the clinical assessment and how the assessment seems to have changed over time in the sense that fewer people are categorised as seriously injured. This helps explains the more rapid decline in police-reported serious injuries compared to clinically serious ones. These matching data provide strong evidence that police-reported assessments of severity are neither an accurate nor a consistent indicator of the problem.

(b) Police Hospital and Injuries Board matching from 2010 to 2013

Matching of three datasets is not usual and there are few examples in the road safety field. The process of matching the three sets is described in Box 1. Refinements to this method are undoubtedly possible in future, obviously if there is improved data, but also using entirely probabilistic methods. The method used here is consistent and provides estimates for the first time of the size of the overlaps between the datasets. Statistical errors (meant as false positives or negatives), which could each be of the order of 10% of the matched file, as well as possible biases due to data problems require that care be taken with the interpretation of the results, and that further iterations and experience are needed.

Table 9 summarises the results and Figures 4 and 5 illustrate them.

Table 9. Matching Results Police Hospitals and Injuries Board 2010-2013

Year	Police	Hospitals	IB	PHIB	PH~IB	P~HIB	~PHIB	Unique
2010	8,053	5,011	16,160	708	811	1,839	410	24,748
2011	7,078	4,548	16,332	609	733	1,697	327	23,983
2012	7,826	4,415	19,117	551	706	2,013	384	27,153
2013	6,760	4,343	18,648	591	552	1,788	249	25,980
Total	29,717	18,317	70,257	2,459	2,802	7,337	1,370	101,864

Sources: police, hospitals and Injuries Board datasets; matching software; own calculations.

Figure 4. Injuries 2010-2013, by Dataset Category

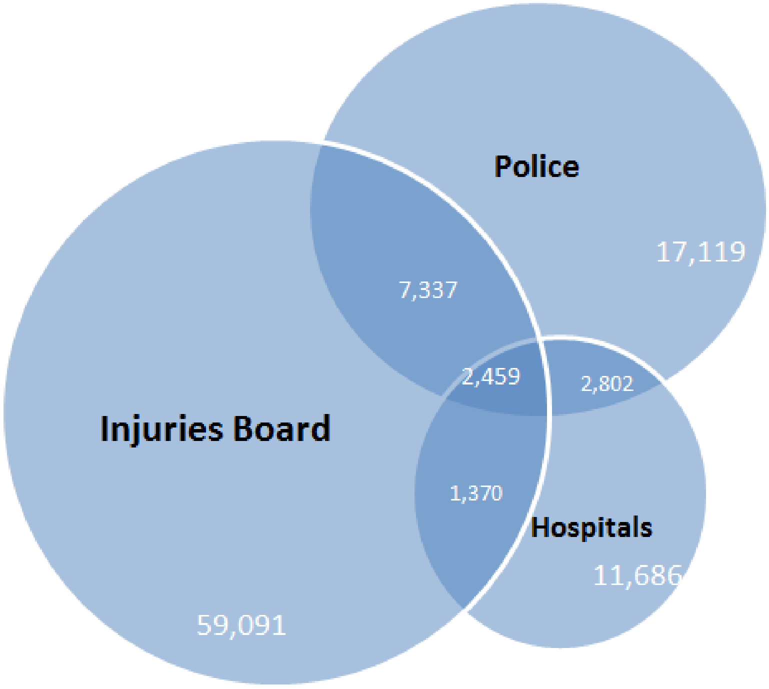
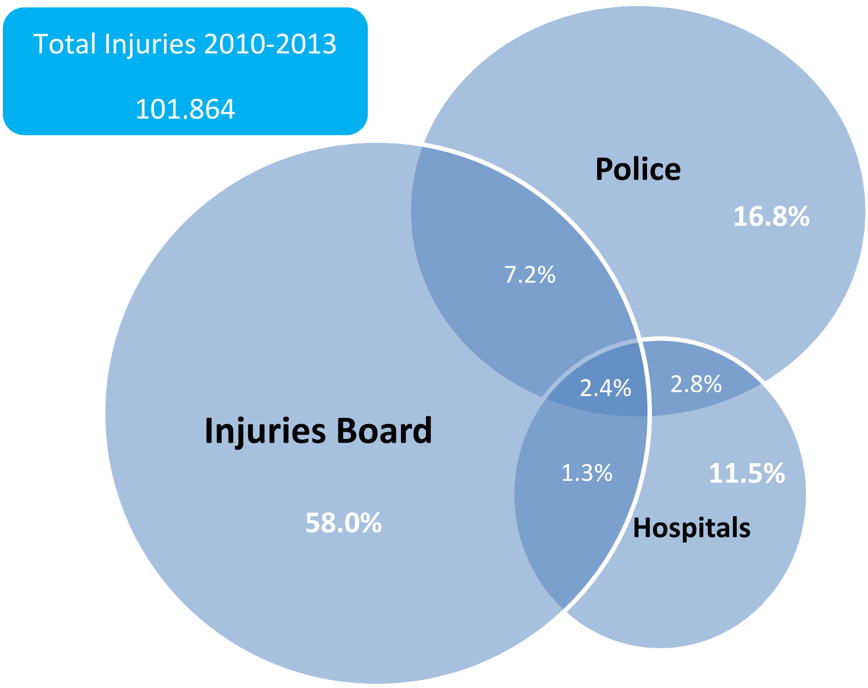


Figure 5. Distribution of Injuries by Dataset, 2010-2013



It is noteworthy how the overlaps are relatively small as shares of the datasets, as can be seen from the figures. As a result the total number of unique injury cases identified in the three datasets is around 25,000 per annum. This is over three times the official injury figure.

A dataset of these individuals can be constructed (called the Combined Injuries Dataset) and allows explorations of the injuries especially when individuals are in more than one dataset as set out in Table 9. The notation in the table refers to the datasets in which an individual is present. So for example ~PH~IB concerns individuals in the hospital dataset but not in the other two. Persons appearing in all datasets compared to those in just one are more inclined to be male (62% against 54%), to be slightly older (35 years old compared to 33, with 9% over 65 compared to 5%) and are much more likely to be pedestrians (24% compared to 5%). Restricting analysis to severity illustrates how the combined dataset can be a useful analytical tool.

As mentioned, no single measure of severity applies across the datasets and different measures can be used. These are: a serious injury as assessed by police at the scene, Maximum Abbreviated Injury Score (MAIS) and length of stay (LOS) for hospital patients and size of award for Injury Board claimants. While these are not directly comparable each can give a threshold for a serious injury and when the same individual is in more than one dataset, severity assessments from more than one source are available and can therefore be compared. The severity information in the Combined Injuries Dataset shows that the different measures are correlated but not always strongly so. Severity increases with the number of dataset appearances but where comparisons can be made, all combinations occur.

To extend the analysis, let a "bad injury" be defined as one that is "serious" by any of the measures for the individual datasets. So a person with a bad injury is either classified as seriously injured by the police or is clinically seriously injured (MAIS3+) or is long stay (6 days or more) for those in the hospital dataset or offered an award of over €20,000 by the Injuries Board.

Table 10 below shows the results for 2013.

Table 10. Bad Injuries by Dataset

Dateset Presence		Bad Injury		Total
		No	Yes	
PHIB	Number	207	385	592
	%	*35.0*	*65.0*	*100.0*
PH~IB	Number	336	262	598
	%	*56.2*	*43.8*	*100.0*
P~HIB	Number	1,368	489	1,857
	%	*73.7*	*26.3*	*100.0*
P~H~IB	Number	3,540	173	3713
	%	*95.3*	*4.7*	*100.0*
~PHIB	Number	168	164	332
	%	*50.6*	*49.4*	*100.0*
~P~HIB	Number	13,671	2,196	15,867
	%	*86.2*	*13.8*	*100.0*
~PH~IB	Number	1,986	835	2,821
	%	*70.4*	*29.6*	*100.0*
Total	Number	21,276	4,504	25,780
	%	*82.5*	*17.5*	*100.0*

Sources: police, hospitals and Injuries Board datasets; matching software and calculations.

It can be seen that 17.5% of the total, that is over 4,500 person are ''badly" injured using this composite measure. This total is almost nine times the official police-reported serious injury total. Such a composite measure can obviously be criticised but it provides an alternative indicator of a difficult-to-measure concept compared to the present one that is obtained from the judgement of a medically untrained police officer at the scene of an incident. So while there is an arbitrary element to this broader definition, it provides a wider framework for analysis and policy than one based only on a single measure. If, for example, a weaker definition of bad injury was taken with only those obtaining IB awards over €40,000 counted as a serious injury, the number of bad injuries would be around 2,200. Thus, even on this more conservative definition the number of bad injuries would be over four times the official total. Table 10 also shows, not surprisingly in view of its definition, that the proportion of people with bad injuries is significantly higher if the persons are in two or three datasets.

Capture-recapture methods
The analysis above shows a more diverse and larger injury problem than official data indicate. But the three sources are not exhaustive and there are also other injuries that may not be included. Potential sources for these include insurance, A&E or doctors' surgeries but at present there are no useable data from these sources. It would obviously be helpful if there were some way to estimate the full total of injuries based on the information we have. A technique to do this, developed mainly in animal populations, is called capture-recapture and has led to sophisticated methods to estimate a true population size from a number of samples (Amstrup et al; 2005). In road safety research a simple estimator called the Lincoln-Petersen (L-P) is often used to estimate the total population of fatalities or injuries based on the records on two lists, usually police and a medical source. Under specific conditions, the population size is estimated by taking the product of the two file sizes and dividing by the overlap (IWGDMF, 1995; Hook and Regal 1999; ITF, 2012). Because of its simplicity there is a temptation to use the L-P estimator even though the conditions for its correct use may not be met. The possibility to use this estimator is explored in Short and Caulfield (2016). It is concluded that its use in road safety studies may not be appropriate. More sophisticated methods are required, though the initial attempts made in Short (2016) are not encouraging. The main problems lie in the heterogeneity of the human population and the lack of sufficiently gradated measures for the type and nature of the injuries. Models to estimate the total injured population are likely still some way off and there will be no substitute for better data from the missing sources.

Data from 2014
Detailed Police data for 2014 were made available by RSA in July 2017. Where possible the tables and charts in this paper include these data. But because there are some significant developments in 2014, this section has been added to the paper.

In 2014, as stated in Section 3, the police data collection changed from a paper based reporting system to one which included on- line data transfer to the PULSE call centre in Castlebar. Moreover, additional efforts were made in the call centre to verify and improve the consistency of data.

Table 11 below compares the broad aggregates from the 2013 and 2014 police data.

Table 11. Comparison of 2013 and 2014 Aggregates

Measure	2013	2014	% Change
Fatalities	188	193	2.7
Injury Collisions	4,797	5,618	17.1
Material Damage	21,734	33,510	54.2
All Injuries	6,880	8,079	17.4
Serious Injuries	508	755	48.4
Pedestrian	97	180	85.6
Cyclist	50	106	112.2
Motorcyclist	47	87	85.0
Vehicle occupant	314	382	27.4

Source: RSA, 2017

The year 2014 saw a small increase in fatalities, a larger one in injuries (after a sharp decline the previous year). However, more noteworthy is the increase of almost 50% in police-reported serious injuries and the far larger increases in serious injuries for vulnerable users. Note also the extremely large increase in the number of material damage collisions reported.

How should these data be understood? Do they in some sense measure real changes or do they simply result from changes in procedures? Certainly, there were not dramatic changes in other injury data or in traffic. Hospital admissions were up by 2.6% and the number of clinically serious injuries was stable at 920. Traffic increased by 1.7% (CSO, 2017) and cycling has increased significantly in recent years. Nevertheless, it is highly unlikely that these new injury data could reflect real changes of these levels.

The text accompanying these 2014 data (RSA, 2017) states there have been "…enhancements in the validation process" and that 2014 represents... "a break in the series" and " it will take 5 years before any appreciable trends can be confirmed''. On the first of these points, the enhancements are not described and it is not clear what is meant. Is it the police or RSA validation process that is meant? Are some injuries "upgraded" by the police or RSA? The proportion of collisions that are deemed serious has been declining for several years until it increased slightly in 2013 and then jumped significantly in 2014 from 8 to 11 % of collisions. The result seems to be a kind of correction to what this paper has shown as the disproportionately large decline in the number of serious injuries. But the mechanisms are not clear. Is it possible that reports of injury undercount may have somehow influenced the way the police classify severity?

The argument that 2014 represents a break in the series that will take five years to remedy seems to imply that severity data will be more or less meaningless for at least a further five years. This is clearly unsatisfactory and the RSA statements do not make for good statistical practice. If there is a break in a series the reasons for it, and any definitional or procedural changes should be described. A new series could then begin on the new basis; it is not correct that the series should be unclear or undefined for an indefinite period.

More importantly, these changes can be seen as further confirmation that assessing injury severity is not a job that police can or should do. The next section explores the policy consequences of this point and the other findings in the paper.

7. POLICY IMPLICATIONS AND RECOMMENDATIONS

Introduction

This paper has shown that the injury problem due to road traffic in Ireland is significantly larger and structurally different than indicated by official data. Specifically, official police-reported injury data understate the number of traffic injuries by a factor of at least three, with the figure being greater for vulnerable users, particularly cyclists. Depending on how they are defined, serious injuries are understated by a factor of from two to eight. The understatement in police-reported serious injuries has increased over the period 2005-2013, as police assessment of a serious injury seems to have changed.

These facts imply that traffic injuries deserve more policy and research attention. Up to now the number of fatalities has been the main focus of attention and has been a powerful political driver for improvements in policy. Difficulties in defining injuries and monitoring them consistently are among the main reasons for their relative neglect in policy and research. However, injuries matter, not only because they are very costly to society but also because the objective to make the transport system safe requires attempting to avoid collisions and injuries as well as fatalities.

The following sections examine some implications of these conclusions for road safety policy. First, injuries are discussed in the context of the national road safety strategy, with a particular focus on serious injuries and the question of a target for reducing the number. Data issues are then examined and suggestions made for improvements to existing data and for new indicators. Some examples are given of areas where specific policies for injuries might be indicated. Finally research needs arising from this work are set out.

Policy Orientation and Road Safety Strategy

There is a growing understanding internationally of the significance of injuries in road traffic and the need to make them a more central part of road safety strategies. In Ireland, the road safety strategy for the period 2013-2020 contains 144 actions (RSA, 2013). Of these, there are only two specific references to injuries, in actions 116 and 117. Action 116 refers to the need to define serious injuries and then set a reduction target. Action 117 suggests the collection of data on injuries through a National Health and Welfare Survey. These ideas are examined in the following paragraphs.

Defining and Measuring Serious Injuries
This work has shown how a serious injury is a complicated concept that is difficult to define precisely. All the definitions used here (police-reported serious injury, clinically serious injuries, long stays in hospital and large compensation awards), show how individuals can be categorised as seriously injured by one of the measures and not by any of the others.

Clearly, the use of a clinical indicator (MAIS3+) will provide improved consistency to the measurement of serious injury. From the evidence presented here, the number of seriously injured people as measured by MAIS3+ is, at least until 2013, around twice the official police-reported number, with a much larger multiple especially for cyclists but also for motorcyclists. The MAIS3+ totals can be extracted from the hospital data and published as a new serious injuries series. This can be done without record linkage and all that is needed is the mapping from the ICD-10 codes to the MAIS indicator. While this is relatively straightforward, there are a number of definitional decisions that will need to be taken before this mapping is applied. These are described in Section 3. Such a data series should initially be provisional with the possibility to correct and backdate it.

However, hospital data alone is a limited way to deal with a complex topic of growing importance. A broader approach, using other available data sources will also be needed to provide improved indicators on injuries and their severity.

Police data are of particular importance in this regard as they have always been and will continue to be the main source of information on traffic collisions. The evidence here shows that police cannot and should not be relied on to assess injury severity. The recently available data from 2014 reinforces this view. Police obviously need to continue to record injuries in collisions. And they should collect information on the nature of the collision. A definition of serious **collision** (which would not necessarily involve a serious injury but which resulted in major traffic disruption or road closures) might be useful. One possibility is that, as in parts of the UK, a list of injuries be provided as a guideline to determine severity. Police should also check if the collision led to a hospital stay, as they theoretically should do at present. The existing series is counterproductive for policy, giving up to 2013 a false impression that serious injuries were declining faster than fatalities and then, after a slight increase in 2013, showing a 50% increase in 2014. Given these problems, a review is needed of the data collection, the definitions used and procedures followed, as well as on the allocation of responsibilities between police and RSA. This review should be led by the Department of Transport, with the involvement of the agencies and the Department of Justice as well as outside experts. Police would also be helped by a clearer requirement to report traffic injuries than is set out in Section 106 of the Road Traffic Act of 1961.

In the end, a solution like that in New Zealand where the police are specifically financed through the Department of Transport to provide safety services including data collection and enforcement might be one way forward.

For a deeper understanding of injury severity, data from other sources and record linkage techniques are needed. These are discussed below; but first, a brief discussion on the issue of setting targets for injuries.

Setting a Target for Serious Injuries
While in principle injury targets might appear to be an evident addition to the measures available, there are two main problems with such targets at present. The first is that, as shown here, police-reported figures for serious injuries are neither an accurate nor a consistent measure of the severity of the injuries incurred. Thus existing police-reported data do not measure the problem accurately and it would be inappropriate to use them for setting targets. Secondly, while targets can be a very useful stimulus to political action, they need to be evidence-based and be susceptible to policy influence. However, it is not clear how a separate serious injury target could meet these criteria at present. This is largely because there are not yet specific policies that might influence injuries in a distinguishable way from fatalities. As an example, Ireland's road safety strategy adopts an indicative target for a reduction in serious injuries for 2020. It is that police-reported serious injuries be reduced from 474 in 2012 to 330 by 2020 or 61 per million of population[2] . Such a target would imply that between 2013 and 2020 serious injuries would need to decline by 3.9% per annum while fatalities are targeted to decline from 162 in 2012 to 124 in 2020, a rate of 3.3% per annum. The data presented here shows that fatalities have declined more rapidly than injuries and a target that requires injuries to decline faster than fatalities would imply additional specific measures aimed to reduce injuries. But at present such measures or their costs or potential benefits are not known. The conclusion is therefore that before injury targets are set, there needs to be a better understanding of the size and

[2] The injuries target is arithmetically inconsistent in the RSA strategy document (RSA, 2013). If the target of 61 serious injuries per million is correct the numerical target should be 303 injuries. This would require larger annual reductions than the figure of 3.9% shown.

nature of the problem and of the specific policies and measures that can make a difference. Once this knowledge is in place, injury targets can be a helpful addition to road safety's policy tools. This argument applies also to the decision in the EU to set a 50% reduction target for serious injuries. Few Countries can show serious injuries data in a comparable way3 and until this is improved, such targets are virtually meaningless.

Improving the Data
This work has shown that existing published data give an inaccurate assessment of the injury problem. This has led to an underestimation of the relevance of injuries in safety policy and may also have led to errors in policy prioritisation. Better data will reduce the risks of policy bias and are an essential component of a strategy to reduce traffic injuries.

Existing Data
The review of the existing data from police, hospitals and the Injuries Board showed how each source identifies specific features of the injuries problem but also pointed to weaknesses in each of these sources. Continuous improvement of these datasets, along the lines set out in Section 3, are essential to an improved understanding of the traffic injury problem. But this will not be enough and new series are needed to get a better understanding of the injury problem.

New Data Series
This paper underlines how no single source can provide an adequate measure of the extent of the traffic injury problem. Different sources need to be used and combined. At the simplest level, aggregates and analyses from the separate sources can be published and used. Hospital data should be used for serious injuries. More sophisticated methods involving especially record linkage and possibly capture-recapture as described above can also be used. Record linkage is required in order to obtain estimates of the true totals involving several sources. Importantly, the technique can be used to generate new data series that give a fuller picture than any of the sources individually. Such series would be valuable as they can give indicators on the main trends and also on specific types of injuries (for example to cyclists) that are seriously understated in the present official data. The linkages between hospital and police data are the starting point and can, in addition to giving a new series for injuries, provide a wealth of useful analytic material. Furthermore, using also the Injuries Board data, especially if it can be improved by the addition of modal and clinical variables, would give a data series that includes a range of injuries not shown in police or hospital data and would provide a valuable and much broader indicator on the size of the traffic injury problem.

A data series based on the results of record linkage of the three sources will take some time to develop into a reliable or more formal indicator, both because the method is new and untested and because the data need to be better understood. However, such a series would give a more representative view of the extent of the injuries problem and could be started as part of a set of more informal indicators.

A valuable addition to this accumulation of data from diverse sources would be to try to obtain injury information in another way. One possibility is to add some questions to the National Travel Survey on whether the respondents were involved in a road accident in the previous 3 or 6 months. This has been done in other countries and can provide consistent data on injuries.

Policies for Injuries
Road safety policy has focussed mainly on reducing fatalities. The assumption behind this is that by reducing fatalities, injury and material damage consequences will also be reduced. However, this assumption is not entirely accurate. In the period from 2005 to 2013, fatalities reduced by 50% while injuries as measured by the total in the police and hospital dataset fell by only 27%. Furthermore, the trends for the modes are quite different, with cyclist injuries actually showing an increase over the period. It is therefore not automatic that actions to reduce fatalities will reduce injuries to the same extent. This does not imply that the focus on fatalities is inappropriate. But it does indicate that injuries and potential injury consequences of policy measures need to be taken into account to a greater extent than they have until now.

Specifically, the injury implications of different policy measures need to be evaluated, particularly the impacts on the vulnerable modes. A key to a Safe System approach (ITF, 2008) is that measures, their costs and their likely

[3] Published figures from the Commission show widely varying (by a factor from 1 to 6) in the 17 Countries with clinical severity data. Ireland is included in these data as the best performing country on serious injuries but the figure used does not correspond to any published data (EC, 2016).

effects are considered together. The following sections illustrate some examples where specific policies focussed on injuries may be appropriate.

Cyclists
The data show the very large difference between police and hospital data. Over the period 2005-2013, police reported 211 seriously injured cyclists (50 of them in 2013!), while hospitals showed 8,707 cyclist admissions. Of these, 1,650 were categorised as clinically seriously injured (MAIS3+), almost eight times the police-reported number. About 16% of the hospital bed nights in 2013 accounted for by transport injuries were for cyclists compared to their 0.6% share of traffic (Short and Caulfield, 2014). Clearly, using police data alone to assess cycling safety greatly underestimates the risks with this mode of transport. Putting this in a more positive way, the benefits and the benefit/cost ratio of measures to make cycling safer are likely to be far greater than previously believed. These measures include helmet wearing, reducing traffic speeds and investments in cycle infrastructure.

- *Helmet wearing*

Cycle helmet wearing is not compulsory, although over 50% of cyclists do wear one (RSA,2013). Unfortunately, data on helmet wearing is not collected by the police at accidents but the hospital evidence shows that 37% of cyclists have head injuries, the highest for all modes, and much higher than the 15% for motorcyclists. Though there is an intensive debate among cyclists on the benefits of helmet wearing, the arguments in favour are strengthened by the data and analysis in this work. In particular, the case to introduce compulsory helmet wearing by children appears compelling. As shown in Short and Caulfield (2014), the odds against a cyclist in a collision with a vehicle are 40 to 1[4] and measures to reduce this disparity should be welcomed.

- *Urban speeds*

Severity of cyclist injury goes up rapidly with speed. Only 5% of cyclists in accidents are seriously injured in 30 km/h speed zones compared to 9% in 50 km/h zones and 36% in 100 km/h zones. Policies to encourage cycling need to be accompanied by measures to reduce vehicle speeds where there is mixed traffic. The generalisation of low speed zones (30km/h or even 20km/h) in towns with mixed traffic is an essential complement to the pro-cycling measures. Moreover, urban speeding is endemic as only 22% of traffic on the urban network obeys the 50 km/h speed limit (RSA, 2013). In short, a combination of improved enforcement and lower speeds is required to facilitate the further increase in cycling that is anticipated.

- *Cycleways and separate facilities*

Separation of cyclists from other traffic obviously reduces conflicts and is preferred by cyclists themselves. Investment in such facilities becomes more cost-effective when the social cost data are taken into account, as these show that the potential benefits are understated using existing traffic accident counts. Appraisal methodologies applied to infrastructure projects include estimates for injury undercount, but these are not sufficiently adapted for specific modes.

Whiplash Injuries
A second illustrative area where specific injury policies may be appropriate is for whiplash or soft tissue injuries. The Injuries Board indicates that a large proportion of the claims which are settled for under €20,000 are for whiplash injuries. Many of these collisions will not involve the police or hospitalisation and in general are relatively low speed rear-end car collisions. An examination of these collisions and how their number and consequences could be reduced (including improved medical categorisation as well as measures such as head-rest positioning, insurance incentives to introduce automatic emergency braking, information campaigns and efforts to eliminate bogus claims) is an example of an important injury area where there are probably highly cost-effective solutions.

Urban Planning and Projects
Urban improvements or traffic management projects such as pedestrianisation, pavement widening, accessibility improvements or cycleways are among the kinds of projects where there may be significant safety benefits. In general, these projects may not correspond to the kinds of schemes where traditional cost-benefit analysis is used. However, these are likely to have extra safety benefits that cannot be identified from the police-reported data alone. The contribution these projects can make to reducing injury and injury risks should be examined.

Economic Considerations and Project Appraisal
The costs to society of traffic fatalities and injuries are published annually by the RSA (RSA, 2010-2013). Table 12 shows the social costs of injuries for the years 2010 to 2013. Also shown are revised estimates of these costs,

[4] In the sense that in collisions involving cyclists and vehicles, for every injury to a vehicle occupant there are 40 cyclist injuries.

using the evidence in this paper from the three injury sources. In these calculations a serious injury is defined as a clinically serious one (MAIS3+) and all other injuries are considered minor. The social costs per injury are those used by the RSA.

Table 12. Official and Estimated Injury numbers and their Social Costs, 2010-2013

Injuries	Year	Official (using police data)			Revised (Combined Dataset)		
		Serious	Minor	Total	MAIS3+	Other	Total
Numbers	2010	561	7,492	8,053	1011	23,766	24,748
	2011	472	6,606	7,078	913	23,091	23,983
	2012	474	7,468	7,942	953	26,200	27,153
	2013	508	6,372	6,880	921	25,059	25,980
Costs (M €)	2010	141	176	317	254	558	812
	2011	121	162	283	234	566	800
	2012	120	182	303	242	650	892
	2013	148	161	309	268	644	912

Sources: RSA, own calculations

On this basis it can be seen that the annual social cost of traffic injuries is at least three times, and for 2013, almost €600m greater than the official estimate. It should also be noted that the social costs of injuries now significantly exceed those of fatalities, which were estimated , for example in 2013, to cost €497m. This confirms why injuries need to be given more policy attention and also implies that measures to reduce injuries will in general have a higher benefit to cost ratio than previously considered. One consequence is that transport appraisal should revise upwards the values used for the estimates for injury undercount. A factor of 2 to 3 should be used for serious injuries (instead of the current 1.5) and a factor of at least 3 for minor injuries (which is the same as that used at present). But most importantly, these values should be varied by mode with cyclists especially having a much higher figure.

Research Needs

This is an area of policy that should become more important and on which there has been little research. The recommendations above on improved and more readily available data are the starting point for a better understanding of the injury problem. In addition, research on the origins and causes of injury accidents, their locations and the costs and consequences of injury accidents would all be valuable.

Causes of Injury Accidents

This work has not examined the causes of the injury accidents in any detail nor examined how they might differ by mode or other factors. This information could help identify specific combinations of circumstances that might be likely to result in injury. In general, the causes will be the same as for fatal accidents, but it would be beneficial to obtain a more detailed understanding of them. An analysis of injury risk factors for the vulnerable modes especially would be helpful to improve the appraisal of projects suggested above.

Injury Locations

Fatalities in road crashes have become rarer; in the past, information on the location of these crashes has provided specific indications on infrastructure or other problems. This has, however, become of somewhat lesser importance as remedial measures have been carried out on many of the accident black spots. But since they are more numerous, an examination of the locations of injury accidents can identify concentrations at particular locations or roads. Unfortunately, the precise location is available only in the police data. But such work, can provide vital indications for remedial measures. In particular, such exercises have not been carried out for the vulnerable modes and would make a valuable contribution to dealing with them.

Consequences of Injuries

Injuries encompass a wide range of circumstances and consequences that are not adequately summarised by any of the measures used. There have been recent cohort studies in other countries on the consequences of injuries including impairments, disability and psychological impacts (Hours et al., 2014; Tingvall et al., 2014). The study in France shows for example that almost half of those with minor injuries (MAIS2 or MAIS1[5]) have regular pain one year after the accident. The cited work in Sweden confirms this finding showing how MAIS2 injuries have lasting and serious consequences for many of those involved. Cohort studies on accident victims would provide a far better understanding of the range of consequences and their duration and would help gain insights into the categorisations that could be used for injuries. The clinical injury category MAIS2 has the largest number of people and a deeper understanding of the nature and consequences of these injuries would also be valuable.

Social Costs of Injuries
A more solidly-based and broader set of estimates for the social costs of injuries would aid policy in finding the most cost-effective measures. A model is New Zealand where social costs are published for many categories of accidents, providing an analytic basis for policy interventions (Ministry of Transport, New Zealand, 2014). There is also a need for an international study to revisit the human cost valuations used for injuries as these are based on very small samples and date from at least twenty years ago. Recent recommendations in the UK and New Zealand (NERA, 2011; NZIER, 2015) confirm the need to update and refine this work.

8. SUMMARY

The research for this paper shows that road traffic injuries are both more serious and structurally different than previously considered. From this, four principal conclusions are drawn. Firstly, injuries should become a more integrated part of Ireland's safety strategy in future. Specifically, priorities in the strategy need to be re-examined in light of the findings here that injuries fall disproportionately on vulnerable traffic users. Research to fill some of the gaps identified would also contribute. Secondly, project appraisal needs to take more account of the injury consequences of proposed measures, not only by revising the undercount estimates used in cost-benefit analysis, but also by specifically examining the potential injury impacts for different groups of road users. Thirdly, improvements to existing data, along the lines suggested here work will contribute to obtaining a clearer understanding of the extent of the problem and will greatly facilitate exercises like record linkage. Finally, new data series and indicators are needed to monitor and better understand the variety of circumstances involved in injuries.

In the end, a deeper understanding of the injury problem will be vital to safety policy in the future and improved data, a higher policy profile and additional research can all contribute to this aim.

[5] Recall that MAIS2 and MAIS1 are clinical classifications of injury with maximum scores of 2 and 1 on the abbreviated injury scale. They have a lower risk of death than MAIS3 but can have long lasting and serious consequences.

Bibliography

Alsop, J. and Langley, J. (2001). Under-reporting of motor vehicle traffic crash victims in New Zealand. *Accident Analysis and Prevention*, *33*(3), 353-359.

Amoros, E., Martin, J. L. and Laumon, B. (2006). Under-reporting of road crash casualties in France. *Accident Analysis and Prevention*, *38*(4), 627-635.

Amoros, E., Martin, J. L. and Laumon, B. (2007). Estimating non-fatal road casualties in a large French county, using the capture-recapture method. *Accident Analysis and Prevention*, *39*(3), 483-490.

Amstrup, S. C., McDonald, T. L., Manly, B. F. J. (editors) (2005). *Handbook of Capture-Recapture Analysis*: Princeton Press.

Association for the Advancement of Automotive Medicine (AAAM) (1971). Rating the severity of tissue damage, I. The abbreviated scale. *JAMA*, *215*(2), 277-280.

Baker, S. P., O'Neill, B., Haddon, W. Jr. and Long, W. B. (1974). The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *Jounal of Trauma*, *14*(3), 187-196.

Book of Quantum (2004). Injuries Board, Ireland http://www.injuriesboard.ie/eng/Forms-Guidelines/Book-of-Quantum.pdf.

Central Bank of Ireland (2012) Private motor insurance statistics, 2009.

Cryer, P. C., Westrup, S., Cook, A. C., Ashwell, V., Bridger, P. and Clarke, C. (2001). Investigation of bias after data linkage of hospital admissions data to police road traffic crash reports. *Injury Prevention*, *7*(3), 234-241.

Cryer, C., Gulliver, P., Langley, J.D. and Davie, G. (2010). Is length of stay in hospital a stable proxy for injury severity? *Injury Prevention*, *16*(4), 254-260.

CSO (2011). Central Statistics Office, Ireland National Travel Survey, 2009.

CSO(2017). Central Statistics Office, Ireland. Kilometres Travelled by Road Traffic.

http://www.cso.ie/multiquicktables/quickTables.aspx?id=tha10_2

Department for Transport UK (2012a). Linking Police and Hospital data on Road Accidents in England, 1999 to 2009.

Department for Transport UK (2012b). Reported Road Casualties in Great Britain; a valuation of road accidents and casualties in Great Britain in 2011. https://www.gov.uk/government/statistical-data-sets/ras60-average-value-of-preventing-road-accidents#history

EC (2008). European Commission. Estimation of real number of road accident casualties: Safety Net project, final report.

http://erso.swov.nl/safetynet/fixed/SafetyNet%20Final%20Activity%20Report_V3_Final.pdf

EC (2012). European Commission. Next Steps in the Development of the Injuries Strategy.

https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/injury_next_steps.pdf

EC (2016). European Commission. Road Safety in the European Union; Trends Statistics and Main Challenges. https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/vademecum_2016.pdf

ESRI (2013).Economic and Social Research Institute Activity in Acute Public Hospitals in Ireland 2012. https://www.esri.ie/__uuid/462574ea-80a8-4bb4-96ea-31bf0fdcfc2d/2011-HIPE-Annual-Report-Final.pdf

European Centre for Injury Prevention (2006). University of Navarra, Algorithm to transform ICD-10 codes into AIS and ISS, version 1 for SPSS. Pamplona, Spain 2006.

Fellegi, I. P. and Sunter, A. B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association*, *64*(328), 1183-1210.

Fellegi, I. P. (1997). Record Linkage and Public Policy – A Dynamic Evolution in Record Linkage Techniques. In *Proceedings of an international workshop and exposition*. Arlington, VA, March 20-21, 1997: Federal Committee on Statistical Methodology.

Ferrante, A., Rosman, D. and Knuiman, M. (1993). The construction of a road injury database. *Accident Analysis and Prevention*, *25*(6), 659-665.

Fingerhut, L. A. (2004). International collaborative effort on injury statistics: 10 year review. *Injury Prevention*, *10*, 264–267.

Gennarelli, T. A., Wodzin, E. (2006). AIS 2005: a contemporary injury scale. *Injury*, *37*(12), 1083-1091.

Haddon, W., Jr. (1973). Energy damage and the ten countermeasure strategies. *Journal of Trauma*, *13*(4), 321-331.

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.

Hours, M., Chossegros, L., Charnay, P., Tardy, H., Nhac-Vu, H. T., Boisson, D. and Laumon, B. (2013). Outcomes one year after a road accident: Results from the ESPARR cohort. Accident Analysis and Prevention, 50, 92-102.

HPO (2015a) Healthcare Pricing Office. Activity in Acute Public Hospitals in Ireland http://www.hpo.ie/ .

HPO (2015b). Healthcare Pricing Office. Hospital In-Patient Enquiry Scheme (HIPE); DataDictionary; Version 5.0. http://www.hpo.ie/hipe/hipe_data_dictionary/HIPE_Data_Dictionary_2013_V5.0.pdf

HSE (2013). Health Service Executive, Ireland, Annual Report, 2012. http://www.hse.ie/eng/services/Publications/corporate/annualreport2012.pdf

Injuries Board (2007-2016). Annual Reports and Accounts (2006-2015). http://www.injuriesboard.ie/eng/News-Information/Annual-Reports-Archive/ .

Injuries Board; Statistics (2007-2014). http://www.injuriesboard.ie/eng/Statistics

Irish Statute Book (2015a). Road Traffic Act, 1961 http://www.irishstatutebook.ie/ResultsTitle.html?q=Road+Traffic+Act+1961&search_type=all&button=Search .

Irish Statute Book (2015b). Personal Injuries Assessment Board Act, 2003 and 2007 amendments. http://www.irishstatutebook.ie/2007/en/act/pub/0035/index.html

Irish Statute Book (2015c). Courts Act 1991. http://www.irishstatutebook.ie/1991/en/act/pub/0020/index.html.

IWGDMF (1995a). International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation I: history and theoretical development. *American Journal ofEpidemiology*, *142*(10), 1047-1058.

IWGDMF (1995b). International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation II: Applications in human diseases. *American Journal of Epidemiology*, *142*(10), 1059-1068.

IRTAD (2013-2016). International Road Transport Accident Data, Annual Reports, 2012-2015, OECD, Paris.

ITF (2008). International Transport Forum, OECD. Towards Zero; Ambitious Road Safety Targets and the Safe System Approach. OECD, Paris.

ITF (2012). International Transport Forum; IRTAD (2012). Reporting on Serious Road Traffic Casualties: OECD, Paris.

Jeffrey, S., Stone, D. H., Blamey, A., Clark, D., Cooper, C., Dickson, K. and Major, K. (2009). An evaluation of police reporting of road casualties. *Injury Prevention*, *15*(1), 13-18.

Johnson, S. and Walker, J. (1996). The crash outcome data evaluation system (CODES): Washington. National Highway Traffic Safety Administration.

Ministry of Transport, New Zealand (2014). Social Cost of Road Crashes and Injuries, June 2013 update.

Motor Insurance Advisory Board, Ireland (2002). Final Report; Summary of recommendations.

NERA consultants (2011). Updating the VPFs and VPIs; report for the Department for Transport, UK.

NZIER (2015). Approaches to valuing injury and mortality risk in transport assessments. Wellington: for Ministry of Transport, New Zealand.

ONISR (2011). L'Observatoire National Interministériel de la Sécurité Routière. Bilans annuels. http://www.securite-routiere.gouv.fr/la-securite-routiere/l-observatoire-national-interministeriel-de-la-securite-routiere/accidentalite-routiere/bilans-annuels

Osler, T., Rutledge, R., Deis, J. R. and Bedrick, E. (1996). ICISS: An International Classification of Disease-9 Based Injury Severity Score. *The Journal of Trauma: Injury, Infection, and Critical Care*, *41*(3).

Reurings, M. C. B and Stipdonk, H. L. (2011). Estimating the number of serious road injuries in the Netherlands. *Annals of Epidemiology*, *21*(9), 648-653.

RSA Factbooks; for years 2006-2013. Road Safety Authority, Ireland.

RSA (2013). Road Safety Authority, Ireland. RSA Strategy, 2013-2020.

RSA (2017) Road Safety Authority, Ireland. Road Casualties and Collisions in Ireland 2014;Tables

Short, J. and Caulfield, B. (2014). The safety challenge of increased cycling. Transport Policy, 33, 154–165.

Short, J. and Caulfield, B. (2016). Record Linkage for two and three datasets for road traffic injuries in Ireland using police hospital and injury claims data. Journal of Safety Research 58, 1-14.

Short, J. (2016). The Road Traffic Injury Problem ; Size, Costs and Policies. PhD Thesis. Trinity College, Dublin

SWOV (2014). The valuation of human losses of road deaths. Fact Sheet. https://www.swov.nl/rapport/Factsheets/UK/FS_Immaterial_costs.pdf

Thomas, P., Chambers-Smith, D., Morris, A., Yannis, G. et al. (2009). Building the European Road Safety Observatory; Final Report on Task 1.5 (SafetyNet), European Commission.

Tingvall, C., Ifver, J., Krafft, M. et al. (2013). The Consequences of Adopting a MAIS 3 Injury Target for Road Safety in the EU: a Comparison with Targets Based on Fatalities and Long-term Consequences; IRCOBI Conference Paper.

UNECE (2009). United Nations Economic Commission for Europe. Glossary of Transport Statistics, Geneva.

WHO (2013).World Health Organisation. Global Status Report on Road Safety. Geneva.

WHO (2015). World Health Organisation. International Classification of Diseases, 10th Revision. Geneva. http://www.who.int/classifications/icd/en/.

Winkler, W. E. (1999). The State of Record Linkage and Current Research Problems. Governmental Statistical apers, Canada https://www.census.gov/srd/papers/pdf/rr99-04.pdf

ANNEX 1: THEORY OF RECORD LINKAGE

This Annex summarises record linkage theory as set out originally by Fellegi and Sunter (1969). It also presents a Bayesian method of explaining the theory.

The following notation is that in the paper by Felligi and Sunter.

Let A and B be two sets of records, with elements *a* and *b* respectively. For each record a in A and b in B, let α (a) and β(b) be vectors of the observed characteristics of a or b. So $\alpha(a) = (a_1, a_2 \ldots a_k)$ is a vector, for example, consisting of values for age, sex, date…. for record a.

Let AXB={(a,b): a∊A, b∊ B} the set of pairs with one from each set, the set of comparison pairs. For each pair (a,b) a comparison vector for the elements a and b is obtained.

$\Upsilon(\Upsilon) = (\Upsilon^1(a_1,b_1), \Upsilon^2(a_2,b_2), \ldots . \Upsilon^k(a_k,b_k))$ a vector of comparisons Υ^j on each of the characteristics between a and b. The Υ^i values in the comparison vector are 0s and 1s, with a 1 indicating a link on variable i and 0 a non-link. In this work, the ϒ (ϒ) are either 4 or 5 dimensional vectors, consisting of 0s and 1s. In the case of 5 variables, there are 32 possible combinations for ϒ.

Let Γ be the set of all comparisons, called the comparison space. Now, Γ is a function on AXB with a probability distribution.

Let M= {(a, b): a = b; a ∊ A, b ∊ B} the set of matched pairs.

Let U= {(a, b): a ≠ b; a ∊ A, b ∊ B} the set of unmatched pairs.

At the core of record linkage is a decision rule D that allocates, on the basis of the observations of ϒ, pairs (a,b) as matched i.e. (a,b) ∊ M, not matched i.e. ∊ U, or undecided.

A linkage rule is a mapping from Γ to a set of decision functions D = {d(ϒ)} where

$d(\Upsilon) = (P(A_1|\Upsilon), P(A_2|\Upsilon), P(A_3|\Upsilon))$ with the sum of the probability components being 1 and where A_1 A_2 and A_3 are the sets deemed to be matches, possible matches and non-matches respectively. For each value of ϒ the decision rule assigns the probabilities for each of the three possible actions.

The decision that pairs in A_1 are matched is known as a positive link. The decision that pairs in A_3 are unmatched is known as a positive non-link. The decision that pairs in A_2 are neither matched nor unmatched, is known as a possible link.

The conditional probability of ϒ given that (a,b) ∊ M is

$$m(\Upsilon) = \sum P\{\Upsilon(\alpha(a), \beta(b)) | (a,b) \in M\}$$

$$= P\{\Upsilon(\alpha(a), \beta(b))\}. P(a,b)| M\}$$ with the sum taken over all (a, b) in M.

The conditional probability of ϒ given that (a, b) ∊ U is

$$u(\Upsilon) = P\{\Upsilon(\alpha(a), \beta(b)) | (a,b) \in U\}$$

$$= \sum P\{\Upsilon(\alpha(a), \beta(b))\}. P(a,b)|U\}.$$ summed over (a, b) in U.

These are the *m* and *u* values used throughout the text. Both are conditional probabilities, with *m* being the probability of a match given a link and *u* the probability of a match given a non-link.

Levels of Statistical Error

There are two kinds of statistical error in record linkage. First, when linked pairs are not truly matched; and second, when unlinked pairs are matched.

Using the notation above the two kinds of error are:

$P(A_1|U) = \mu = \sum u(\Upsilon) P(A_1|\Upsilon)$, with the sum taken over the values of ϒ. These errors are also known as false positives or homonyms (Clark,2004).

$P(A_3|M) = \lambda = \sum m(\Upsilon)$. $P(A_3|\Upsilon)$ the sum over Γ; known as false negatives or synonyms.

These probabilities can be understood to correspond respectively to Type 1 and Type 2 errors in statistical hypothesis testing, where the set of linked pairs is considered as the body of the distribution and the set of unlinked pairs the tail. A Type 1 error is the probability of rejecting the null hypothesis when it is true, that is, by chance getting an outlying value in the tail of the distribution. A Type 2 error is the probability of accepting the null hypothesis when it is false, that is wrongly getting a test value in the tail of the distribution when it should be in the body.

The aim in record linkage is to minimise A_2 for fixed levels of μ and λ that is to minimise the number of cases that are classed as possible links for the fixed levels of error. Fellegi and Sunter's main theorem is that for linkage rules L on Γ which have these levels of error, the optimal one is defined as that which minimises $P(A_2|L)$. This maximises the number of positive dispositions that is the number assigned to either M or U and makes intuitive sense as costly examination of the undecided cases should be avoided as far as possible.

The proof begins with a unique ordering of realisations of ϒ so that $m(\Upsilon)/u(\Upsilon)$ is a monotonically decreasing sequence (taking care when values are zero or equal to define the ratio). The ordered set {ϒ} is indexed as Υ_i with i=1,2….N, with $u_i = u(\Upsilon_i)$ and $m_i = m(\Upsilon_i)$

With μ and λ as admissible levels of error, select n and n' so that the sum of the u_i to n-1 is just less than u and the sum of the m from n' to N is just less than λ.

For practical purposes, the Fellegi-Sunter paper makes a key simplifying assumptions that the components of ϒ can be listed and that they are statistically independent with respect to each of the conditional distributions. This means that *m* and *u* can be shown as products of the individual marginal probabilities. This assumption is made systematically in record linkage projects. It implies that errors on one variable like date or age are independent of errors on others. This may not be true, (since poorly completed records may have more than one error) but is difficult to test without true records.

Bayes' Theorem and Prior and Posterior Odds

Neither Newcombe or Fellegi-Sunter mention Bayseian probabilities in their expositions. However, the odds ratio form of Bayes's theorem can be shown to give the same results as well as additional insights. The following exposition is based on Clark (2004) and also uses his notation. The Irish data and linking possibilities are used to illustrate the theory.

Given two files or lists of records, let $A_1, A_2, \ldots A_n$ mean the events that records selected at random from the two lists are identical on characteristic 1,2,…. n. So A_1 means that characteristic 1 (for example, incident date) is the same on both lists. In the Irish case, the two lists are the police and hospital injuries data, n=5 and the variables are crash and admission date, age, sex, mode and county.

Let B be the event that records from the two lists are a true match. P(B) is the probability that a record chosen at random from each list is a true match, i.e. that two records on different lists belong to the same individual.

The ratio $P(B)/P(\overline{B})$ is the odds of B, also referred to as the prior odds, as it is the odds of B without additional information. Here $\overline{B}$ means not B and note that $P(\overline{B}) = 1 - P(B)$.

Working with odds is conventional in this domain and has an intuitive sense. It is also convenient for numerical manipulation. Odds can be converted easily into probabilities since,

$$\text{Odds}(B)/1+\text{Odds}(B) = \frac{P(B)/P(\overline{B})}{(P(B)/P(\overline{B}))+1} = P(B).$$

Now P(A.B) is the probability that both A and B occur. By Bayes's theorem P(A.B)=P(A).P(B|A) where P(B|A) is the probability that B occurs given A has occurred. Similarly, P(A.B)=P(B).P(A|B). And so P(B).P(A|B)=P(A).P(B|A) and therefore;

$$\mathrm{P(B|A)} = \frac{\mathrm{P(B).\,P(A|B)}}{\mathrm{P(A)}} \qquad (1)$$

Similarly

$$\mathrm{P(\overline{B}|A)} = \frac{\mathrm{P(\overline{B}).\,P(A|\overline{B})}}{\mathrm{P(A)}} \qquad (2)$$

Now, $\frac{\mathrm{P(B|A)}}{\mathrm{P(\overline{B}|A)}}$ is the odds of B given A, also known as the posterior odds of B. It is the odds of a match given A. Therefore, from (1) and (2),

$$\frac{\mathrm{P(B|A)}}{\mathrm{P(\overline{B}|A)}} = \frac{\mathrm{P(B).\,P(A|B)}}{\mathrm{P(\overline{B}).\,P(A|\overline{B})}} \qquad (3)$$

That is, the odds of B given A is the product of the odds of B by the odds of A given B.

This is known as the odds ratio form of Bayes' theorem and applies for any A.

Then, on the assumption that the As are independent,

$$\frac{\mathrm{P(B|A_1A_2 \ldots A_n)}}{\mathrm{P(\overline{B})|A_1A_2 \ldots . A_{\mathit{n}})}} = \frac{\mathrm{P(B).\,P(A_1|B).P(A_2|B) \ldots P(A_{\mathit{n}}|B)}}{\mathrm{P(\overline{B}).\,(P(A_1|\overline{B}).\,P(A_2|\overline{B}) \ldots . P(A_{\mathit{n}}|\overline{B})}} \qquad (4)$$

Recall that P(B) means the probability that two records on different lists refer to the same person and A_i means that characteristic i (age, sex, mode…) is the same on both lists.

P(A|B) is known as the *m* probability. It is the probability of a match on characteristic A, given the records are a true match. $\mathrm{P(A|\overline{B})}$ is known as the *u* probability and is the probability of a match on characteristic A, given the records do not match. These correspond to the *m* and *u* functions of Fellegi-Sunter, mentioned above where the relationship is:

$$\frac{\mathrm{P(A|B)}}{\mathrm{P(A|\overline{B})}} = \frac{m}{u}$$

Therefore, from Equation 4, the posterior odds of a true match is the product of the prior odds by the likelihood ratios. It is the prior odds multiplied by the *m*/*u* factors, where specific characteristics match and by (1-*m)*/(1-*u*)where they do not match.

This link between prior and posterior probability can be used to derive a useful result as follows. First, a monotonically decreasing sequence of the ϒ values is defined. Now, the function

P= c(ϒ/1+ ϒ) where c is any positive constant, is also a monotonic decreasing function which respects the ordering of ϒ. If c is the prior probability, then P is an ordered set of posterior probabilities, p_i. Moreover, the matches, non-matches and possible matches are ordered in the same way using this function.

Now, consider that each selected pairing is made without replacement so that there are only M pairs of best matches, where M is the size of the smaller of the two datasets being matched. Then, consider $\sum_1^M p_i$ the sum of the posterior probabilities with the sum taken over all M pairs.

An event with probability p_i, repeated for N trials has an expected number of occurrences of Np_i. Here for each probability the number of trials is 1. Therefore, for each p_i, the expected number of matches is p_i. Since the p_i are independent, the total expected number of matches is the sum of the probabilities $\sum_1^M p_i$.

This is the main result used in the linkages undertaken in this work. The sum of the posterior probabilities provides an estimate for the expected number of matches. Using this number of matches, sets of 'best' matches can be obtained and this facilitates analysis and understanding.

Weights and Probabilities

Newcombe et al. (1959) introduced logarithms into the exposition of linking methods. Taking logs of both sides of Equation 3 above gives the result that the posterior log odds that the records refer to the same person is a constant (the prior log odds) plus the sum of the log-likelihood ratios for each element.

If the weights are applied according to the theory, and each weight is the likelihood ratio *m*/*u* then the relationship between the total score (S) and the likelihood ratios is as follows;

$$S = \sum_1^k log_2 \frac{m_i}{u_i}$$

$$\text{Therefore, } 2^S = \prod_1^k \frac{m_i}{u_i} \text{ as } \sum_1^k log_2 \frac{m_i}{u_i} = \log \prod_1^k \frac{m_i}{u_i} \qquad (5)$$

Conventionally, as explained above, the weights for a record linkage exercise are the logs of the likelihood ratios and the total score for each combination is the sum (S) of these log likelihood ratios.

From the exposition above:

$P(B|A_1.A_2...A_k)/P(\overline{B}|A_1.A_2....A_k) = P(B)/P(\overline{B})*2^S$.

That is, the Posterior odds= Prior odds * 2^S.

And therefore the posterior probability, P_{post} is given by:

$$P_{post} = \frac{\text{Posterior odds}}{\text{Posterior odds+1}} = P(B) * 2^S \qquad (6)$$

That is the posterior probability is the product of the prior probability and 2 to the power of the score. This also demonstrates how the odds double when the score increases by 1 and was the reason Newcombe used logs. The values of the logs of the *m* and *u* factors were called “binits”. Mathematically, it is not necessary to use logs but they have this intuitive sense and in this work the practice is followed, not least because the computer programme is set up to use them (Linkage Wiz, 2014).

The Prior Odds

The prior odds are the original odds of a match by chance. So it is the probability that two records selected at random, one from each list, match. This is the ratio of the number of possible true pairs to the total number of possible pairs which is the product of the two file sizes. In practice, the true number of matches is not known and needs to be estimated. If there is a reason to believe that the number of matching records is N_x, with N_A and N_B being the numbers in the two files, the probability P of choosing a matching pair by chance is given by:

$$P = \frac{N_X}{N_A N_B}$$

If N_A and N_B are large this will be a very small number and its log_2 a large negative one. The odds will also be a small number very close to the probability. A benefit of the method used in this work is that the prior odds can be estimated initially and revised when the true number or an improved estimate is known.

DISCUSSION

Seán Barrett: Chairman, I wish to join in the welcomes already expressed for Dr Short's paper this evening. Extending and improving official statistics and applying them in policy fields has been a major concern of this Society since its foundation in 1847. John Kells Ingram, elected an honorary member of the American Economics Association in 1891, and a founder of this Society, wrote that "it has not occupied itself with dilettante statistics, collected with no special purpose and tending to no definite conclusion. It has from the first, applied itself, in the spirit of earnest inquiry to the important questions affecting the condition of the country." As Dr Short says in his opening paragraph road transport kills 1.2m people worldwide annually with some 50m people injured. The paper brings the experience of the author from the ESRI, the Department of Transport and the OECD where, as Director General, he skilfully transformed the old European Conference of Ministers of Transport into the International Transport Federation which now has 57 members worldwide.

The societal cost of road transport in terms of lives lost and blighted by injury has been overlooked in policy debates. Had these fatalities and injuries been concentrated in location and time societies the social costs of roads might be more widely recognised. The explosion that downed Air India Flight 182 in 1985 killed 329 people and is remembered today at Ahakista. In 1985 road deaths in Ireland, at 410, were one quarter more than on the Air India flight in the same year. Table 2 shows that up to 2008 more people in Ireland died in road accidents than died on the Air India flight. Full of vital and interesting data the paper fills an important gap in how we address safety in transport. I hope that it will become required reading for researchers and policy makers in academies and agencies in fields such as transport, health and safety and law enforcement.

The items that I have chosen from the paper reflect an interest in transport economics, welfare economics and institutional economics. The findings which interested me most were, for example, that Ireland has some 40 persons injured for each fatality and that there are "many injuries not reported in official data." Table 2 shows massive differences in injuries reported by the police at 6,760, hospitals at 4,343 and the Injuries Board at 18,648 in 2013.The paper notes that in Ireland almost two thirds of cyclist hospitalisations are categorised as non-traffic accidents" and that "no national data on reason for attendance at A and E units is collected." A further problem is that "Insurance Ireland did not provide data on personal injury claims, declaring that they do not collect this information from their members." There were 6,400 clinically seriously injured but not reported to the police between 2005 and 2013 in a total of 36,000 hospitalised with road injuries not notified to the police. Table 10 indicates that in 2013 there were 4,500 persons badly injured or almost nine times the police reported total. We learn also that "the number of people making claims is significant, more than are seen in the hospital or police datasets", whiplash accounts for up to 80% of claims and that the total compensation awarded for transport incidents is around €150m per annum." Many of whiplash settlements "will not involve the police or hospitalisation and in general are relatively low speed rear-end collisions." Table 12 estimates that injury accidents in 2013 cost €912m that is almost three times greater than the €309m based on official police data. The paper concludes that "road traffic injuries are both more serious and structurally different than previously considered." We underestimate the safety and injury costs of road transport in general and in particular in the case of cycling where the injury costs have been seriously underestimated by cycling advocates as this paper shows.

The standard of cost/benefit analysis of transport projects has to be improved in the light of this paper and other research. Three major projects, Metro North, DART Underground and the Cork/Limerick motorway have not been adequately assessed. Ireland needs independent assessment of projects and agencies by bodies such as DPER, IGEES and the C and AG. More published independent analysis and less spin doctoring is required. The first Nobel Prize winner in 1969 Jan Tinbergen wrote of the efficiency of instruments in regard to targets. We need to revisit that literature.

It has been obvious for some time that we have a major problem in policing in the Republic of Ireland. The casual treatment of l.8m driver breath tests with nobody held responsible rivals the financial sector in regard to moral hazard or the successful transfer of responsibility for inefficiency from those making decisions to others in society. We know from Michael Clifford's recent book on Sergeant McCabe that Brian Farrell ·of the RSA listened to him when senior police figures did not. There are two other examples where the legal system does not adequately address the social costs of road deaths and injuries. It is not compulsory to carry a driving licence when driving or when attending court. People thus escape penalty points and disqualified drivers do not surrender their licences. In the Seanad in February 2014 the then Transport Minister, now Taoiseach, Mr Varadkar, agreed with these points but was persuaded by the Garcia authorities to preserve a ten day gap between a Garcia request to produce a licence and its actual production. He said that he was working hard to ensure that the Courts Service records penalty points. Three years and some six hundred road deaths later Fiona Gartland (Irish Times, April 18, 2017) cited" new legislation specifically requiring a presiding judge to ask a convicted driver to produce a licence in court (which) would tighten up existing procedures." Only Gardai of the rank of inspector or above, that is 2% of membership, have powers to set up alcohol testing stations for road traffic. An amendment to allow sergeants to establish such stations was rejected by the Minister because the Gardai had not requested it. The amendment would have increased to 14% of Gardai powers to implement this road safety function. Tonight's paper must be required reading for the Kathleen O'Toole committee reviewing the entire structures of An Garda Siochana.

The work of Buchannan, Baumol, Niskanen and others provides useful insight into the problems indicated in tonight's paper. Bureaucracies don't wish to have their outputs measured. Indifference to data is no surprise. They don't like regulation and set out to achieve regulatory capture. The dominant goal of the institution becomes budget maximisation. A major task in Irish public administration will be to address these problems in two sectors central to this paper, policing and the health service. DEPER, IGEES and the C and AG will have a vital role in these reforms.

The insurance sector is now regulated by the Central Bank following a series of bankruptcies in which the customers of the surviving companies were levied to cover the losses of the bankrupt ones. The Central Bank will, I hope, remedy the data deficiencies in the sector as stated on p. 14.

Short writes that" despite several requests and discussion, Insurance Ireland did not provide data on personal injury claims, claiming that they did not collect this data from their members. These data are important in themselves as they are the key to motor insurance costs; but they would also allow improved understanding of the links between the Injuries Board data and the Insurance system. It is to be hoped that the insurance industry will begin to make efforts to contribute better to understanding the traffic injury problem." (p.14). The Central Bank as regulator might examine the cost of the whiplash problem which is large in Ireland by international standards.

Our present road safety policy seeks to overcome formidable obstacles due to defects in policing, enforcement, court procedures and inefficient insurance providers largely by marketing campaigns promoting road safety. I have concerns about this emphasis. For example, media campaigns on television may have no effect on young males who are a major target group in road safety but may not watch much television.

Our present campaigns pay little attention to the vehicle component of road injuries. For example a major independent bus operator on the intercity routes installed alcohol locks and benefitted from a lower insurance premium. Seat belt alarms might play an increased role in a country where a reported 40% of victims do not wear seat belts. Should the alarms be programmed to get louder the longer the seat belt is unused rather than go silent as at present? Technology already exists in some vehicles to spot tiredness in drivers, and obstacles on roads and may offer protection against single vehicle accidents caused by driver fatigue. Ireland because it does not have a vehicle industry is in a strong position to insist that vehicle standards make a greater contribution to road safety. It is ironic that we have policies to increase the contribution of cars to save the planet rather than save lives and serious injuries today. We also neglect the contribution of roads to road safety. We should address the contribution of poor maintenance, poor design and poor planning to road injuries and the liability for such injuries. The NRA evaluation of 298 road safety remedial schemes between 2002 and 2005 estimated that the expenditure of €5.3m on 298 schemes reduced fatalities by 52, serious injuries by 51 and minor accidents by 180. End to end motorways also reduce accidents but cost €10m per km. They may not be cost effective compared to tailored schemes of remedial works.

The statistics presented in this paper are not dilettante but have strong policy implications. John Kells Ingram would approve. It will be a test of economics in Ireland to ensure that this paper stimulates better public policies. This paper must not be allowed to gather dust.

Tom Ferris: Chairman, I would like to join the other speakers in congratulating Dr Jack Short for an extremely comprehensive paper on the topic of traffic injuries, which to date has not been given sufficient attention by policy makers in Ireland. Jack and I go back many decades to the Planning Unit of the Department of Transport. Jack went on to a distinguished international career, ending-up at helm of the Paris-based ECMT (later renamed the International Transport Forum). It is clear that he has never lost his appetite for research, analysis and policy formulation. I will confine my remarks to the crucial need for robust statistics in order to formulate good policies. Jack Short's paper clearly demonstrates that need for robust statistics as a prelude to formulating policies that will help tackle the number of road traffic injuries occurring on Irish roads. The statistics assembled for this paper show that road traffic injuries are both more serious and structurally different than previously estimated. In particular, Table 12 shows that the annual social cost of traffic injuries is at least three times greater than the official estimate in 2013, i.e. €912 million as against €312 million. The gravity of the issue of road traffic injuries requires a higher policy profile from the relevant authorities. In particular, the Department of Transport should be driving the agenda for the development of better data and the formulation of policies that will lead to action being taken at an early stage.

Paul Sweeney: I ask if Mr Short had considered the impact of driverless cars on injuries. I am of the view that driverless cars would have a massive impact on road injuries, deaths and insurance costs. In fact I think the impact of these cars would lead to substantial job reductions in the insurance industry, in hospitals, in local authorities and private car parking maintenance. I would point out that the average person does not use their car 95% of the time. With driverless cars, few will want to own cars. People will call a driverless car on an app; go to their destination; leave the car to be picked up by someone else or to wait nearby. The number of cars on the road will shrink dramatically, eventually to perhaps 15-20% of current car stocks. This will impact on taxis, delivery vans and trucks, car manufacturing, car sales, vehicle insurance and the built environment, on the city streetscape. It will lead to huge reductions in employment. But it will lead to large reductions in car accidents with the consequent gains in lives and injuries saved, in health and related expenditure and in insurance. What does Mr Short think?