7TH WCTR
SYDNEY AUSTRALIA
1995

TOPIC 12
GIS, LAND INFORMATION
SYSTEMS AND DATABASES

DEVELOPMENT AND USE OF AN INTEGRATED SAFETY DATABASE

KEVIN P. AUSTIN
Halcrow Fox and
Institute for Transport Studies
University of Leeds, UK

MILES TIGHT
Halcrow Fox and
Institute for Transport Studies
University of Leeds, UK

HOWARD R. KIRBY
Halcrow Fox and
Institute for Transport Studies
University of Leeds, UK

Abstract

The notion of joining computer files relating to road safety has been prevalent for many years, although few additional sources of information have been used in any one study. This paper describes the development and application of an operational integrated safety database for use by researchers and practitioners.

INTRODUCTION

Most countries collect information on road accidents. This is usually collected by the police who attend the scene of an accident or who have accident details reported to them. Such data provides the basis for most road safety investigations, both in terms of studies of causation and as the basis for the design of remedial or preventative treatments. The information collected routinely in this way is by its very nature limited in terms of its coverage and scope. In addition a large number of accidents are not reported on the database and of those that are a considerable number of errors exist (see for example papers by Bull and Robert, 1973, Mills 1989, or James 1991). The most commonly used dataset of information has limitations both in terms of the quality of the information present and in terms of the range of variables present. This paper suggests a methodology for improving both the quantity and quality of information on safety available to researchers and practitioners by the creation of an integrated safety database, making use of information from a wide variety of other sources of safety data. The work was based in the County of Humberside in England and used data from that authority to demonstrate the feasibility and potential of such an integrated system.

SOURCES OF DATA

A number of extra sources of data exist which could be linked to police accident data to improve the information available. Nelson (1974) stated that insurance, ambulance and vehicle registration data could be linked to police records, but at that time few sources of data were computerised and little money or experienced personnel were available. In the United States several studies have suggested (Chatfield 1985, Briggs and Chatfield 1987) that the linking of files would give a better picture of the safety situation on the highway network.

Harris (1986) identified similar data sources and linking variables in the Netherlands, but extended the range to include casualty data. His proposal for an integrated database is shown in Figure 1. Such a system was never actually produced, but provided an indication of its potential for improving our understanding of safety. The principal sources of accident data are as follows.

Police accident reports

In Britain a national database of road accident details is created using injury accidents which are reported to the police. The information is collected by use of a form called 'Stats 19', a name which is now used to describe the database itself. The database contains information at three levels: accident, vehicle and casualty levels.

Detailed police accident files

These provide further information about the accidents included in Stats 19 database, but are not usually in a computerised format. These is a lot of variability in the quality of individual reports both within and between areas. There are also some restrictions on availability to individuals due to confidentiality rules.

Damage only accidents

In Britain the amount of damage only accident information collected varies from area to area. In some authorities the police keep records on all damage only accidents which are reported to them. However, there is no systematic storage of this information, nor is any of the information regularly passed on to accident investigators at the Local Authority. The information collected is of limited quality, consisting usually of a very brief description of the accident and (often inaccurate) locational details. Given the legal requirements for reporting certain accidents to the authorities there is considerable bias in the nature of the accidents included in such databases.

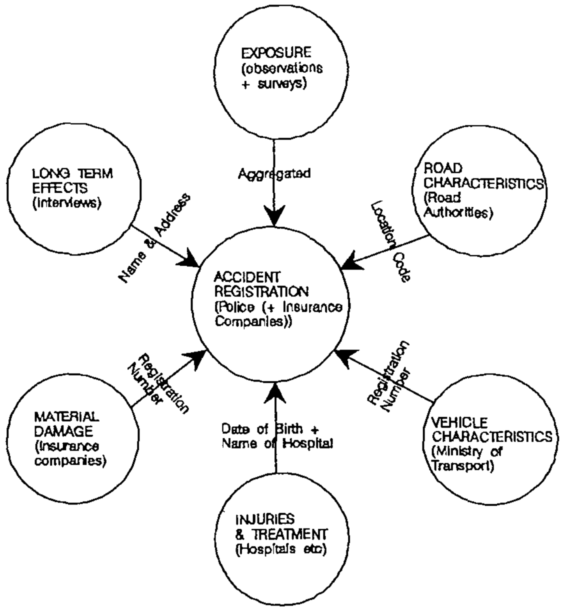

Figure 1 An integrated road accident recording system (Harris, 1986)

Insurance company data

Potentially this could be a very useful source of data, particularly with regard to less easily available information, such as that for damage only accidents. In practice, however, such data has proven very difficult to obtain in a useful format by researchers and practitioners. Partly this is because of difficulties in persuading insurers to provide the data, partly because all of the information completed by insurance claimants is not computerised and partly because of the huge difficulties in obtaining complete data for a particular geographical area due to the large number of individual insurance companies. Were such data made available its most useful function may be as a supplement to existing data from other sources, as its use alone could lead to bias problems.

Hospital data

Until recently in the UK hospital data sources have been difficult to gain access to due mainly to reasons of confidentiality and patient ethics. If such data was made available it was normally made anonymous and hence it became difficult to link individual accident details from the hospital database with other databases such as Stats 19 data. A number of attempts at linking were made (see for example Nicholl 1980, Stone 1984, or Tunbridge 1987) with varying degrees of success and certainty of results. Recently hospital authorities have become more relaxed about providing data in the United Kingdom as long as the requirements of the Data Protection Act are not breached and the data is made anonymous once the match is made. Part of the reason for this change of heart is a realisation that currently used data sources are limited and that the hospital data could potentially provide a very useful means of improving both the quality of the data already collected and the quantity of accidents which are known to researchers and practitioners. The need for making such links was recognised by the World Health Organisation (1980):

> "Public health authorities should co-operate actively in measures which provide a linkage between their data and those of the transport and police authorities, particularly data relating to road traffic accidents resulting in injury."

In depth studies

The purpose of in-depth studies is to obtain comprehensive data from a number of sources so the causes and effects of accidents and injuries can be assessed. Several in-depth studies (Tunbridge et al 1988, States et al 1990, and Hopkin et al 1993) using just hospital and police data have been undertaken to investigate the effects of injury. Tunbridge et al (1988) obtained all relevant information from hospital questionnaires but used the STATS 19 records to check the accuracy of this information. Maximum Abbreviated Injury Scale (MAIS) scores were assigned to each casualty so that the effects of user group, age, gender and seat belt on severity could be made. Hopkin et al (1993) also assigned MAIS scores to casualties so assessments of injury type and severity by age and user group could be made. They also recorded the length of stay in hospital to establish the cost of road accident injuries. Police and hospital records have also been used to determine the injury effects following seat belt legislation in New York State (States et al 1990).

An investigation into the contributory factors of accidents was undertaken by Carsten et al (1989). They examined police reports for over 1000 urban accidents and interviewed the individuals involved. Both Brison (1988) and Lawson (1990) investigated the contributory factors for accidents involving children, and in addition to the sources specified by Carsten et al (1989), coroners' data were used. However, in-depth studies are expensive to undertake and so have to be restricted to relatively small samples, which reduces the transferability of the results.

LINKING DATA

In order to identify the capability of a linked system a study of its potential was undertaken in Humberside. This work focused on the linkage of police and health data using the name and address details contained on both sets of records and using a Geographical Information System (GIS) to link locationaly referenced data. Other potential data linkages are also considered.

Linking Police and health data

For this study a total of 3391 Stats 19 casualty records were obtained from Humberside County Council for the period May to December 1991. Humberside police store data on the same accidents, although they include the personal details of drivers, vehicle owners and casualties. The database containing this information is linked to Humberside County Councils accident database system and so it was possible to access the records via a modem. Several variables (accident, casualty and vehicle reference numbers, casualty surname, forename and address and vehicle make) were collected and were transferred into a Dataease file in a similar way to the STATS 19 records. Because unique accident and casualty reference numbers were contained on both files, the two data sets could be merged.

Hospital data were obtained from the Accident and Emergency Department of Hull Royal Infirmary (HRI) and included all ambulance delivered casualties from the catchment area of the hospital as well as those that were transported privately. When a road accident casualty arrives at the department, details of the patient and the accident are coded. The medical data are entered later by a doctor and include the diagnosis of injury and the treatment given. In total, 1593 casualties that were involved in road traffic accidents during the study period were recorded on the database.

Matching variables

The variables that were used in the matching process were surname, forename, first line of address, age, gender and accident date. An initial study found that only 16 per cent of those records that should match did so using all these items. This was because one or more of the variables were mis-coded by either the Police or the hospital. To improve the level of matching the length of the text based variables were shortened. Forename was reduced to three characters (and for one part of the matching process to two characters), surname to six characters (and for one part of the matching process to two characters) and address to four characters.

Both date of admission and date of accident were coded onto the hospital database. There were 43 records where date of accident was omitted, whereas date of admission was coded for all records and so this was considered more reliable. The date of admission could be identical to, or one day greater than the date on the police file (which allowed the patient time to attend the hospital, especially if the accident occurred late at night) for a match to be accepted. Four correct matches failed to conform to this constraint. Two of these had the date one day less in the hospital record, one had it 2 days less and one had it 2 days more.

Matching method

The hospital records that fail to match with a corresponding police record can be attributed to either failure to report accidents to the police; or failure in the algorithm to match police and hospital records. To produce consistent data for analysis, the number of correct records that the algorithm failed to match must be minimised. In addition, any mis-matches, that is establishment of links between different individuals or for the same individual from different accidents, could also bias the data and should be minimised. The objective of this method was therefore to develop an optimum trade off between the number of correct matches identified automatically and the number of mis-matches.

A matching algorithm was written in Dataease Query Language (DQL) using the Dataease 4.2 database. The correct matches that were not linked were examined to identify if any alterations to the program could improve the level of matching without increasing the number of mis-matches. In the same way, the mis-matches were analysed to assess if they could be reduced without adversely affecting the number of correct matches. In practice, combinations of alterations were allowed, such that on their own each might not produce the required effect but together they would reduce the number of mis-matches and increase the number of correct matches. The combinations that produced the optimum trade off is shown in Table 1.

Table 1 The matching algorithm for linking Police and hospital casualty records

	Date	Surname (6)	Surname (2)	Forename (3)	Forename (2)	Address	Age	Gender
1	X	X			X			X
2	X					X	X	X
3	X			X		X	X	
4	X	X					X	X
5	X		X	X			X	X
6	X			X		X		X

Accident date was considered an essential item in the algorithm since it would be required to identify the appropriate records, especially when the same individual was injured in more than one accident. If several members of the same family were injured in the same accident then accident date, surname and address would usually be common to all of them. It was essential to have some way of differentiating these individuals, whilst making allowances for any transcription errors. Therefore, either gender and forename or gender and age or age and forename were required to be included in each matching level. In some families, forenames are passed down through generations and so the gender and forename condition may be inadequate in some instances. However, the level of mis-matches resulting from its inclusion is likely to be small. Both six and two character surnames were used, since in 8.7 per cent of the matched records where the six character forenames were recorded differently by the police and hospital the two characters were identical. The other variables were generally accurate, and so only 1.1 per cent of additional matches were identified. The algorithm used forename with both two and three characters, since in 1.8 per cent of cases where the three character forenames were different between the two matched records, the two character ones were identical. This led to an additional 0.4 per cent of additional matches being identified. Gender was included in all but one of the matching levels because it was assumed to be a more certain variable, although the possibility of coding errors prompted this to be omitted on one level. Age was allowed to vary by one year to account for small discrepancies in the coding of this variable between the two data sources.

The six levels of matching allowed all the various combinations of transcription errors to be catered for, given the constraints stated above. If there were fewer constraints made in the matching process then a greater number of correct links would be obtained, but at the expense of a higher number of erroneous matches, which could not be justified. The procedure for matching was to compare the first hospital and police records using each of the six matching levels. If any of these conformed to the criteria, then both records were combined and written to a separate file. The next police record was taken and compared with the first hospital record and this proceeded until all police records were checked. The second hospital record was then selected and the process continued until all hospital records were compared with all police records.

Results of using the matching method

A manual match between the police and the hospital database was first undertaken and this identified 1067 hospital records that had a corresponding police record. This meant that 67 per cent of the casualties attending the Accident and Emergency department were reported to the police. A total of 1050 of the manual matches were automatically linked (a total of 98.4 per cent) and there were 12 mis-matches (equating to 1.1 per cent). Three of these mis-matches were because an individual involved in the same accident was recorded twice on the hospital file and the remainder had at least four variables matched between the two records, which always included accident date and gender.

To enable a comparison with the matching methods that involved only in-patient data, the records of patients that were stated by the hospital as being admitted were placed in a separate file, resulting in 287 records. A total of 256 of these were also contained on the police file, a reporting rate of 89.2 per cent. The automatic matching method shown earlier enabled 251 of these to be linked, a success rate of 98.0 per cent. In addition, there were only 2 mis-matches amounting to 0.8 per cent.

The development of an improved algorithm

The matching method described in the previous section took a considerable amount of computer run time. This initial match was undertaken to identify the maximum number of matches and mis-matches that would be realised by the system, although a more efficient matching method is shown in Figure 2.

The first hospital record is taken and all police records are searched to identify those that have accident date the same or one day less than that on the hospital record. Those that correspond to these criteria were placed into a temporary file. The police records in this file were then compared with the hospital record for each level in the algorithm. When a match was identified, the two records were combined and written to a separate file and the next hospital record taken. If the police record did not match, then the next police record was selected and this continued until all the records in the temporary file were searched. The next hospital record was then taken and the process repeated.

Other types of data

Much data collected and stored by local authorities can be referenced by geographic location. For this reason a GIS, capable of automatically combining data sources based on a common geographic location, was used to provide a further computer based link between the combined police and hospital accident database and various other locational variables. This information included a base map of the area to be studied, injury accident data from the police system, non-injury accidents reported to the police, information on school location and journey characteristics of school children, highway inventory data, traffic flow information, land-use data and census data.

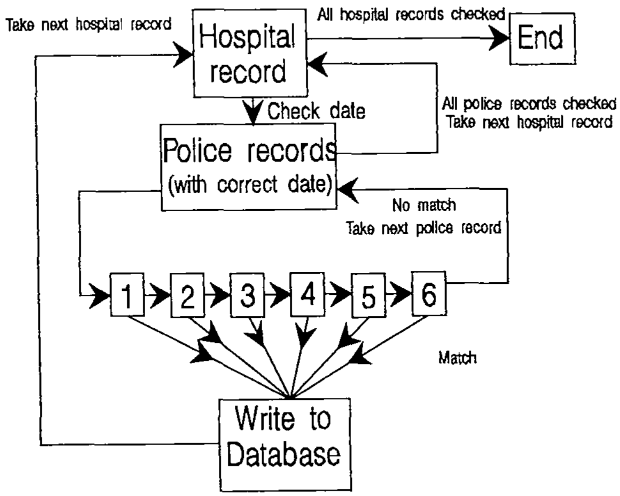

Figure 2 A procedure for matching Police and hospital records

A twenty-five square kilometre area of the town of Hull was chosen for the study which contained a considerable number of accidents and which had a wide variety of different land uses, population, infrastructure and traffic conditions. PC ARC/INFO was used, which is the most popular GIS used by local authorities in Great Britain (Campbell and Masser 1991). The data that were entered onto the system included a map, injury accidents, non-injury accidents, schools and school journeys, highway inventory, traffic flows, land use, census and district boundaries.

Geographical base map

The digital map used for this project was the Ordnance Survey Centre-line Alignment of Roads (OSCAR) provided by the Ordnance Survey. This performed two important functions. Firstly, it was used as a base to overlay other data sources and to identify certain locations. Secondly, it enabled the buffering of specific links and nodes so that analyses could be undertaken on the accidents that were contained within these zones.

Injury accidents

Police reported injury accident data were obtained from Humberside County Council for all accidents occurring between 1987 and 1991 within the study area. The GIS required two files to be created, one for spatial data, containing Ordnance Survey Grid References (OSGR) and unique identity numbers, and the other for the attribute data, containing the remaining information and unique identity numbers to enable these two files to be linked. A file containing the spatial data was created and converted into an ARC/INFO coverage showing the locations of all accidents, each with its own unique identity number. A separate ARC/INFO data file was created with the remaining accident data, which was then merged with the spatial data based on the unique identity number.

Police reported non-injury accidents

A record of damage-only accidents was contained on the police database, although this only related to 1991. Considerable data cleaning was required in order to remove accidents on this database which occurred outside of the study area, as no grid reference existed on the computer files. Details of 431 accidents were included in the database.

School and school journey data

Only thirty-three schools were located within the study area and so the spatial and attribute data could be entered manually. The digital map was displayed on the screen and the positions of the schools were located using the cursor. An attribute table was created, which included the variables of school name, postcode, pupil numbers and age range. This was then joined to the locational information. Information on the route and travel characteristics for pupil's school journeys and their perceived safety problems was obtained from questionnaires at eight of the schools.

Copies of the questionnaire were given to each school and a complementary database containing identical questions about the personal and journey information and accident history was installed onto a microcomputer within each school. The accident history section asked pupils to give details of the accidents that they were involved in and the injuries they received. The information provided was used to trace the accidents to the STATS 19 records so that estimates of the level of under-reporting and comparisons between police and children's accident descriptions could be made. The pupils entered the information onto the computer and when the task had been completed the records were copied onto disk.

The accurate coding of children's routes to school is an extremely time consuming process (as demonstrated by Tight 1987) and the burden can be reduced considerably by the application of GIS technology, whilst still providing an accurate assessment of each child's journey. This project used the routeing program within ARC/INFO, which was designed primarily for public transport or travelling salesmen, in that it automatically follows the shortest path between two selected points. The unique number for the pupil coded on the questionnaire was entered as the route identity number and the node closest to the pupil's home was selected with the cursor, as were a number of points on the way to school depending on the level of diversion from what seemed to be the shortest route. The process of entering data was found to be extremely efficient with over 200 routes coded into the system per hour.

Both routes to and from school were coded and so those pupils whose route differed on the way home had this journey coded into a separate file. Pupils who walked to school were asked to mark on the map the points where they crossed the road to enable a calculation of the number of roads crossed for each journey. All possible crossing points on the map were coded as 1 and so when the route was traced, a key was pressed at the points where the child crossed a road and 1 was added to the total. The total number of roads crossed were recorded in the attribute table for that route and the locations of the child's crossing points shown. The information on each pupil was merged with the route information so that analyses based on distance and roads crossed by the pupils personal characteristics could be undertaken. The safety concerns of parents and their children about particular locations were entered into the database and totalled for each site. The numbers of concerns at each location were then entered onto the GIS.

Highway inventory data

The data that were collected and coded into the computer included: road class; road number; speed limit; pedestrian crossing facilities; carriageway type and markings; junction detail; and junction control. The correct type of feature at each location was obtained from a number of sources. Ordnance Survey maps provided information on road class and road number, whilst the location of speed limit signs, pedestrian crossing facilities and junction control type (ie signalised, stop or give way) were obtained from paper maps within the Accident Investigation Section of Humberside County Council. Information on junction detail (ie cross-roads or T junction) and carriageway type were also obtained from Ordnance Survey maps although a field survey was required as a validation check. The locations of pedestrian crossing facilities were manually plotted onto the digital map.

For each variable, the computer drew a buffer around all the links or nodes relating to it. This was necessary because the grid references for each accident are calculated from paper maps, which do not necessarily correspond to the road centre-lines or junction nodes of the digital map. All accidents falling within this zone could then be considered to be associated with that feature. For the variables of road class, road number, junction detail and junction control the radius of the zone was 24 metres. This distance was based upon:

- the average distance between the centre-line and the edge of the carriageway (3.6 metres);
- the maximum error in the Ordnance Survey maps (0.4 metres);
- the distance from a junction in which an accident is considered to be at the junction (20 metres).

The distance of 24 metres was also used for the variables of carriageway type and speed limit to retain consistency in the investigation method. For pedestrian crossing type, any accident located within 50 metres of a facility should have the crossing type coded and so the area of influence around each facility was set at 50 metres. This would include some roads that were not associated with the road that the facility is located on, for example, those running parallel. The sphere of influence was therefore narrowed to include only those accidents that occurred within 24 metres of the centre-line of that road. The use of too wide a sphere of influence could lead to some of the carriageway type and speed limit codes to be shown as incorrect even though this was not the case. This would occur when accidents were located on roads that existed within the boundary derived for another road.

The width of these zones corresponds to the 24 metre radius from the centre-lines as explained above. The dates that features were altered were also supplied so that accidents that were incorrectly coded but occurred before the feature was altered would not be considered as in error.

Traffic flows

Most local authority safety sections use traffic flows as an additional source of data, hence this element would be an essential element in an Integrated Safety Database. Because of cost constraints, only the major routes within the study area have had traffic counts undertaken. For each location, the twelve hour flow (0700 to 1900) and the date of the count were recorded. The counts at the various locations were made in different years and so growth rates were used to standardise them to the 1989 figure (which was the median year that the accident data were collected). No adjustment was made to the data for counts undertaken in different months, since this would have little effect on any analysis.

In ARC/INFO the map was displayed and the traffic flow values were added by selecting each road section in turn and typing in the relevant value.

Land use

The most recent land use data for the study area were for 1986 and were contained on paper maps. This information was used, although to account for changes since then a site inspection was made. The number of different land use categories was limited to just six, namely: residential, industrial/offices, commercial, open space (such as recreation grounds, allotments and cemeteries), agriculture and services (such as schools and hospitals).

The boundaries of the different land use categories were digitised from the paper map allowing these to overlap the roads. Therefore, if there were different land use types on either side of the road the boundary between them would be along the centre-line. The land use codes were added to each area by selecting each polygon in turn and typing in the relevant value. The land use and roadway boundaries were joined, which retained the road network but split it according to the land use boundaries. The line dividing different land use sections along the road centre line was removed and each of these sections coded with the value relating to the land use on both sides of the road. This was because a certain level of inaccuracy exists in locating accidents to the correct side of the road, and because the land use on both sides of the road affects the accident frequency.

Census data

The census boundaries were digitised into ARC/INFO in the same way as for the land use boundaries. The respective codes for each enumeration district were then added to the attribute table by selecting each polygon in turn and typing in the relevant value. The following variables were produced for each enumeration district:

- total population
- total number of households
- numbers of persons per household
- percentage of households with 0, 1 and 2 or more cars
- percentage of the population in the age ranges 0-4, 5-9, 10-15 and 60 and over
- socio-economic group
- percentage of owner occupied and council households
- percentage of households with greater than 1.5 persons per room
- percentage of economically active people that are unemployed
- mode of travel to work.

These data were transferred into an ARC/INFO and linked with the respective enumeration districts.

Data that were not included

Insurance data relating to claims made against the council were not included since the data was not made available. Insurance claim data were provided for the rural areas of Humberside so that an assessment of its value could be made. Only 26 of the 410 claims related to vehicles striking an object, and half of these were with road works, which was related more to site practice than accident investigation. In addition, the locational description refers to just the town or village where the incident occurred and so this source of data is unlikely to be useful.

An overview of the system

The data contained within the Integrated Safety Database is shown in Figure 3.

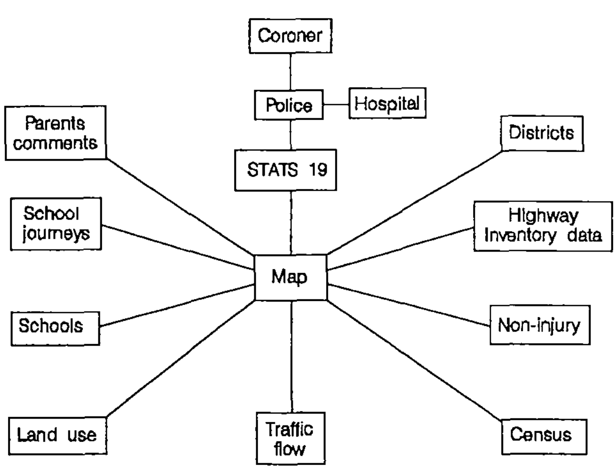

Figure 3 Data sources included on the integrated safety database

This contains much of the information that Harris (1986) and Briggs and Chatfield (1987) suggested should be included. The major exception to this is the absence of vehicle and driver characteristics. This information was not made available by the Driver and Vehicle Licensing Agency, although links between Stats 19 records and vehicle data have been undertaken using registration number, which is common to both records. Links between Stats 19 records and the driver licence file could be undertaken using name and address. Insurance company data was not obtained because many companies would have to be involved to obtain a complete sample, and due to commercial considerations most would be reluctant to disclose any information.

USES OF A LINKED SAFETY DATABASE

There are many uses of such a system. This section aims to give an idea of the capabilities of the system by reference to a number of examples of analyses undertaken to date. The principal findings of these analyses are summarised in order to enable the reader to gain as much idea of the potential scope of the system as possible.

Improving the quality of safety information

Underreporting

The initial use of the system was to check the level of reporting of the police accident data system, compared to the information contained in the hospital database and various other sources. The main findings on this work were as follows.

- The level of reporting increased as the injuries become more severe, although the difference between fatally and seriously injured casualties was not significant.
- The oldest and youngest age groups had the highest level of reporting, although this may be due to the interaction of other factors such as user group and severity.
- Casualties that were injured on the journey to or from school had the lowest level of reporting, whilst those injured on social trips had the highest level.
- Casualties that were referred by 999 (emergency) calls had the highest level of reporting. The same was true for casualties transported to hospital by ambulance as opposed to other modes. This would be because the police are always alerted when a 999 call is made requesting an ambulance.
- For employment status, retired casualties had the highest level of reporting and the under 5 group the lowest.
- Dislocations and fractures had the highest level of reporting whilst sprains had the lowest level of reporting. This related to differences in severity (perceived or real). Casualties with upper limb injuries had a significantly lower level of reporting than other body locations.

Errors

The system was also used to identify errors on the Stats 19 database. Comparisons were made between verified locational data and hospital records and the Stats 19 variables. The major findings of this analysis were:

- For the variables of road class, road number, speed limit and district the level of mistakes was less than 10 per cent. This low figure is probably because the features are unambiguous and are recognised by people with some local knowledge. The level of mistakes for pedestrian crossing facilities, junction detail and junction control was between 10 and 20 per cent. All require the estimation of distance and so it is probably inevitable that a greater number of mistakes will be made.
- Carriageway type was the most inaccurately coded variable, mainly due to difficulties in coding this variable when the accident occurred in the vicinity of bus or turning lanes. This

may be a peculiarity of the sample and a larger study should be undertaken to identify if this phenomenon is unique.

- One of the most striking results was that accidents at pedestrian crossings were underestimated by almost 40 per cent, showing them to be considerably safer than they actually are.
- The level of mistakes in the casualty variables was 0.3 per cent for gender, 14.1 per cent for severity and 17.1 per cent for age. The number of records where age was omitted or incorrectly recorded was not significantly greater for seriously injured casualties than those that were slightly injured. The number of years difference in age between the police and the hospital records was small and so it would be unlikely to affect safety studies greatly.
- The number of seriously injured casualties was underestimated by over 30 per cent and this could have a significant impact on the economic evaluation of schemes. Consideration should therefore be given to using hospital data to code casualty severity, since only a small proportion of seriously injured casualties would not attend accident and emergency departments.

If all highway authorities adopted such a system, then these variables could be removed from the police form altogether as they would be added to the computer database automatically once the location had been specified. This would mean that over 2.1 million fewer items of data need to be collected nationally by the police. This estimate is obtained by multiplying the number of reported accidents in Britain in 1992 (233,025) by the eight variables tested in the system and adding this to the number of junction accidents (143,161) multiplied by two (to account for the extra codes of road class and number when an accident occurs at a junction). The annual costs of recording changes in the road network are likely to be minimal and the information could be used for other purposes, hence it is likely to be worthwhile.

Research applications

A number of potential research applications of the database system were also considered:

Anatomical site of injury

Combination of the Stats 19 information and the hospital injury database enabled a detailed study of the kinds of injuries associated with particular accident types to be undertaken. The principal findings were as follows:

- The use of a linked police and hospital database has shown that the proportion of head injuries to motor cycle and vehicle occupants has significantly reduced over the last two decades, which have coincided with various legislative changes. Priority should now be given to reducing the incidence of leg injuries to vulnerable road users by developing pedestrian friendly vehicles and improving leg protection for motor cyclists.
- A large reduction in the proportion of head injuries to front seat passengers occurred between 1971 and 1991. This decline was not mirrored by rear seat passengers and may be because fewer people wear seat belts in the rear.
- Differences in injury location by age and gender were revealed, although for car occupants this may relate more to seating position than the personal characteristics of the casualty.
- An investigation of the anatomical site of injury by vehicle make and accident type was also undertaken although a larger study would be required before any confident deductions can be made.
- No significant differences in pedestrian casualty severity were identified by distance from a change in speed limit or by distance from pedestrian crossings. However, a comparison of injury severity by land use type revealed that commercial areas had a lower proportion of fatally and serious injured casualties than residential areas.
- Variations in the anatomical site of injury by these locational factors revealed no significant differences, although this may be because only a limited amount of data were available.

In conclusion, the automatic linking of both locational and health based data to police accident records can assist in efficiently providing large quantities of relevant information. The techniques developed in this study have shown their worth and so larger studies should be undertaken to fully explore the reasons for differences in casualty severity and in the anatomical site of injury.

Exposure to risk

Information on background levels of exposure to risk was obtained using the highway inventory data, traffic flow data and census data combined with the accident and casualty databases. Findings included the following.

- The calculation of accident rates per vehicle kilometre can be undertaken more efficiently using a GIS than by previous methods. Other factors such as speed limit, carriageway type, junction density and land use can also be included into any model, although only several of these were found to be significant. Using a predictive equation based on these items, priorities for treatment based on the difference between expected and observed accidents were derived. This produced a considerably different result to that of accidents per vehicle kilometre.
- For area wide investigation, risk factors based on the length of road and the proportion of accidents involving children were also developed using this system. The ranking of areas using this method of calculation produced different results to that of accident frequencies. The combination of a minimum number of accidents and risk factors though would probably be the best method used to select areas for traffic calming.
- The GIS automatically plotted the home location of each casualty and then related the totals to population data taken from the census survey. This produced population based accident rates which enabled priorities for educational campaigns to be directed to where people live rather than where the accidents occur. Although casualty rates differed little between the two areas studied, the disaggregation of the data into age groups revealed that one area suited a campaign directed towards children and the other a campaign for the oldest age group.

Child pedestrian safety on journeys to and from school

A combination of information obtained from children about their journeys to and from school and accident and casualty information was used to obtain details about their actual and perceived risk. The main findings were as follows:

- The use of pupils to enter personal, journey and accident information into the computer and the application of GIS technology to code the routes to school and link other related data is an efficient way of providing complete descriptions relating to journey characteristics and safety problems.
- The factors that influenced pupil's mode choice were found to be distance, crossing or following medium and high traffic flow roads, age and car ownership, whilst the factors influencing their accompaniment were age, distance, crossing or following high traffic flow roads and car ownership.
- In total, 8.4 per cent of pupils claimed to have been involved in a road accident of which 50.8 per cent stated that they were injured, although only 30 per cent of those injured were identified on police files. The accident risk per kilometre travelled using the accidents recorded on the STATS 19 records was similar to that identified by Tight (1987), although was considerably greater than this if the accidents stated on the questionnaire were used. Children's descriptions of accidents were generally consistent with the police but were not comprehensive enough to be used to ascertain the cause of unreported events.
- An investigation of the relationship between straight line and route distance revealed that primary school pupils had a more direct route than those attending secondary school. A large variation in the relationship between straight line and route distance existed among the different primary schools. Equations disaggregated by mode revealed that those who walk have the most direct route, and those who travel by bus the least direct. Although the equations fitted reasonably well, further work is required to determine whether such equations could be used to

factor straight line distance to route distance for studies where a GIS is not available or required.

- The parent's perceptions of locations that were unsafe did not correspond satisfactorily to the incidence of reported accidents and objectively measured surveys. Despite this a couple of remedial treatments have been implemented based on accident causation and the perception of danger.

The techniques developed in this study should enable the travel and safety characteristics of children's journeys to school to be identified more efficiently, allowing more time to be spent in analysing the data in greater depth and in designing appropriate remedial measures.

CONCLUSIONS

This paper has described the creation of a safety database designed to provide a tool which can be used by researchers and practitioners to improve their understanding of safety issues within an area. The results have shown that linking a number of sources of data to create an integrated safety database is feasible. The applications described in this paper have also shown that such a database will bring considerable improvements in both the quality and quantity of data provided to the user. Such an improvement in information can be expected to bring improvements in the levels of safety on our roads due to the greater insight it will provide into the accident causation process.

ACKNOWLEDGMENTS

The support of the Engineering and Physical Sciences Research Council through a research studentship is gratefully acknowledged, as is the help provided by Humberside County Council, the Humberside Police, the Accident and Emergency Department at Hull Royal Infirmary and the Ordnance Survey.

REFERENCES

Austin, K.A. (1995) The development and application of an integrated safety database, PhD Thesis, University of Leeds.

Briggs, D.W. and Chatfield, B.V. (1987) *National Cooperative research program synthesis of highway practice 133: Integrated Highway Information Systems*, Transportation Research Board National Research Council, Washington D.C.

Brison, R.J., Wicklund, K. and Mueller, B.A. (1988) Fatal pedestrian injuries to young children: A different pattern of injury, *American Journal of Public Health*, 78(7), 793-795.

Bull, J.P. and Roberts, B.J. (1973) Road accident statistics: A comparison of police and hospital information, *Accident Analysis and Prevention*, 5, 45-53.

Campbell, H. and Masser, D. (1991) The impact of GIS on Local Government in Great Britain, *Paper presented at the conference of the Association of Geographic Information*, held at the N.E.C., Birmingham.

Carsten, O.M.J., Tight, M.R., Southwell, M.T. with Plows B. (1989) *Urban Accidents: Why do they happen*? AA Foundation for Road Safety Research, Basingstoke.

Chatfield, B.V. et al (1985) *Transportation research circular 293: Introduction to Comprehensive Computerised Safety Recordkeeping Systems (CCRS),* Transportation Research Board, National Research Council, Washington D.C.

Harris, S. (1986) *Linking road accident data to other files. An integrated road accident record keeping system (Road safety)*, Proceedings of Seminar P Held at the PTRC summer annual meeting, University of Sussex, England, 55-85.

Hopkin, J.M., Murray, P.A., Pitcher, M. and Galasko, C.S.B., with Jordan, C., Madden, M.J., Mansfield, S., Sunter, J., Kinsella, A. and Towers, B. (1993) *Police and hospital recording of non-fatal road accident casualties: A study in Greater Manchester*, Department of Transport TRL project report 1/93, Crowthorne.

James, H.F. (1991) Under-reporting of road traffic accidents, *Traffic Engineering and Control*, 32(12), 574-580, 583.

Lawson, S.D. (1990) *Accidents to young pedestrians: Distributions, circumstances, consequences and scope for countermeasures*, Automobile Association and Birmingham City Council, Basingstoke.

Mills, P.J. (1989) Pedal cycle accidents—A hospital based study, *Department of Transport TRRL report RR 220*, Crowthorne.

Nelson, P.G. (1974) *Special purpose multi-discipline mass data collection: The experience of the Royal Australian College of Surgeons' pattern of injury survey*, Presented at a road accident information seminar, Canberra 26-26, March 1974.

Nicholl, J.P. (1980) The use of hospital in-patient data in the analysis of injuries sustained by road accident casualties, *Department of the Environment Department of Transport TRRL report SR628*, Crowthorne.

States, J.D., Annechiarico, R.P, Good, R.G., Lieou, J., Andrews, M., Cushman, L. and Ingersoll, G. (1990) A time comparison study of the New York safety seat belt law utilising hospital admission and police accident report information, *Accident Analysis and Prevention*, 17(1), 7-13.

Stone, R.D.(1984) Computer linkage of transport and health data, *Department of transport TRRL report LR 1130*, Crowthorne.

Tight, M.R. (1987) *Accident involvement and exposure to risk for children as pedestrians on urban roads*, PhD thesis, University of London.

Tunbridge, R.J. (1987) The use of transport-health road casualty data, *Department of Transport TRRL report RR 96*, Crowthorne.

Tunbridge, R.J., Everest, J.T., Wild, B.R., and Johnstone, R.A. (1988) An in-depth study of road accident casualties and their injury patterns, *Department of Transport TRRL Report RR 136*, Crowthorne.

World Health Organisation (1983) *Report on the seminar on the prevention of road traffic accidents, held in Manila, February 1983*, Regional Office for the Western Pacific of the World Health Organisation.