LINKING ROAD ACCIDENT DATA TO OTHER FILES

An Integrated Road Accident Recordkeeping System

Contribution to: Road Safety, Proceedings of Seminar P held at the 14th PTRC Summer Annual Meeting, University of Sussex, England, from 14-17 July 1986. Volume P 284, pp. 55-86. PTRC Education and Research Services Ltd., London, 1986.

R-86-20
S. Harris M.A.
Leidschendam, 1986
Institute for Road Safety Research SWOV, The Netherlands

ABSTRACT

The road accident data which the police collect is of great value to road safety research and is used extensively. This data increases greatly in value if it can be linked to other files which contain more detailed information on exposure; road, vehicle and road user characteristics; and consequences of accidents such as injury, medical treatment, and material damage. Such linked files make is possible to know what sort of people are involved in what sort of accidents in what sort of vehicles, under what circumstances and with what consequences. This knowledge is of immense importance when tackling road safety problems.

Prerequisites for such linkage are the following: a police accident registration of a high, or at least known, representativeness for all road accidents; preferably centralized files of the other relevant data; and common key variable between the police accident data and the other files.

In the Netherlands the police data meets the prerequisite, and there is a National Travel Survey for exposure, a centralized vehicle registration, a central hospital registration of in-patients, and a central registration by insurance companies of material damage. The linkage with the vehicle data has taken place; a trial linkage with the hospital data has commenced and will be completed in 1986.
Linking road accident data to other files results in what we call an Integrated Road Accident Recordkeeping System in which the combined value of the linked files is greater than that of the sum of their individual values.

CONTENTS Page

1. INTRODUCTION

Practically all countries of the world collect road accident data and in most cases this is done by the police. This police registration of road accidents was started, and still has as primary aim, to be able to prosecute those who 'cause' accidents and the resulting injuries and damage by disobeying traffic legislation. In this respect it can be compared in many ways to the police registration of crimes. This data is usually the basis of "official" national road accident statistics.

The information which the police record also proved to be of use to those wanting to study road accidents in order to prevent them or reduce the injuries and damage caused by them. What the police were in fact doing was writing a history of each accident they recorded by describing the situation at the location; both in time and space, what happened immediately proceding, during, and after the accident. Of course this is usually a reconstruction, just as it is in the case of a crime, because the police are very seldom at the scene at the moment the accident occurs. Nevertheless, researchers became interested in what the police recorded and started to use it as research data. This has continued to the present day.
Researchers will continue to use this data because it is a reliable and continuous source. There are other sources such as insurance companies but they are to be doubted because they are compiled by involved parties who are biased. The police, however, are disinterested in that they have nothing to gain from determining whose fault it was (if it was anyone's fault).

Police accident data are probably the most commonly used data in road safety research and although they often are not on their own sufficient for problem analyses or problem solutions they remain invaluable as background orientation material for many road safey problems. They are also extremely useful when looking for problems which seem to be of importance but have not yet been researched.

2. OBJECTIVES

The police accident data contains a great deal of information which is of use to the researcher. This is information about:

- the time and place of the accident;
- a description of the location;
- who was involved;
- what each of the involved did; prior to, during, and after the accident (and therefore who was 'guilty');
- what the result was in terms of injury and damage.

Although the police accident registration differs from country to country as regards what is recorded and the amount of detail, the above list indicates the common factors of practically all countries, at least in the western world (Andreasen, 1975) and in many developing countries (Jacobs; et al., 1975).

In most countries the form which is used to record accidents on consists of one sheet (sometimes more) of approximately A4 format. It is therefore obvious that the amount of detail which the police can record is limited. They record something of everything, or something for everybody. The data is therefore somewhat superficial which is why the police data is often referred to by researchers as a 'macro' file (comparatively little information about a large number of accidents) as opposed to 'in-depth', on-the-spot studies which are considered to be 'micro' files (a lot of information about a small number of accidents).

Researchers use both types of files in their work and both have their own specific value. At the same time researchers try to gather in-depth information from a greater number of accidents and increase the amount of information known for each accident in the national police files.

The most common way of achieving this latter objective is by getting the police to record more details of the (hundreds of) thousands or even millions of accidents they register each year. From the researchers point of view this would also seem to be the most efficient and reliable method because the police are at the scene anyway and it's "only a few extra questions to be asked or observations to be made".

As most or all of us have experienced, the police are often not very keen on additions to the accident registration form. In the first place it always means more work for them (it is very rarely suggested to drop information from the form). In the second place the information which the researchers require (whether it is national or local, temporary or permanent) often seems or is irrelevant for the purpose and needs of the police, viz. to prosecute. Of course the police realize the value of the information for research purposes, but unless they themselves are actively involved in research, the information for them is ballast.

There is a second way of obtaining extra information on accidents recorded by the police which does not entail any extra efforts from the police themselves. On the contrary it can and should lead to a decrease in the amount of information which the police have to record for each accident, and therefore to a decrease in their workload. It should, however, be said that is not the objective of this second way to decrease the work load of the police.
This second way involves combining the police accident data with data gathered elsewhere and of relevance to road safety research and policy. This is known as "linkage" because each of the datafiles is organized in such a way that they are intertwined and form a sort of chain.
In linked data files it is usual, but not essential, to have one central file (in this case the police accident file) with other sub-files each linked directly to the central file (but not necessarily linked to each other). In this way information from all the sub-files can, as it were, be 'sucked' into the central file, whereby the central file can contain all the data of the combined sub-files. This is diagramatically shown in Figure 1 (a full description is given in Chapter 3, Organization).

In orders to link files to each other there has to be at least one 'common key variable' between the central file and a sub-file. That is a piece of information which is identical in both files and unique, such as the identification of time, place, people, and vehicles. The common key variable for each sub-files does not have to be the same: for linking sub-file 1 with the central file a different variable can be used than for linking sub-file 2 with the central file.
It will be clear that such linkage, although theoretically possible manually, is only practicable if all files are computerized.

The reason for carrying out such linkages is that the amount of information in all the files together (although not all is relevant) is many times more than the amount of information in the central file. In the case of linking police accident data to other files much more information per accident becomes available than the police record, or could ever hope to record. The researcher therefore has much more accident and victim data to analyse whereby the national road accident statistics can play a far greater role in road safety research. It is furthermore to be expected that some road safety problems can be analysed without having to resort to collecting special data - they can rely exclusively on the thus expanded national data.

It cannot be denied that another reason for linking data is the simple fact that 'like Mount Everest it's there'. Having separate files on different subjects but being related by subject matter, it seems a terrible waste not to bring them together to see what can be done with the combined data. There is nothing wrong with this motive as long as the combined data is used or has other benefits.

Much better of course is to determine beforehand which details of accidents are needed on a large scale, temporarily or permanently, and to make an inventory of the data available in the sub-files to determine if they will suffice. Action can be considered to gather the necessary data missing.

3. ORGANIZATION

What does it look like when other files have been linked to the police accident data to form an Integrated Road Accident Recordkeeping System? This is shown schematically in Figure 1.

3.1. The Police Accident Data

As Figure 1 shows the police accident data, or as they are often referred to, the national road accident statistics, are pictured in the centre to show that they form the core data and therefore the central file. This is the file which road safety researchers and government wish to extend with the information from the sub-files.
It is therefore logical, but not essential, to use the institute or government body responsible for processing the police data as the host for all the linked files. This body should carry out the linkage and store the linked files as needed on its computer. In most countries this is not the police themselves but a branch of the Ministry of Transport or the Bureau of Statistics. In either case they are bodies interested in and capable of carrying out such a linkage.
The type of data available in the police accident data has already been lightly touched on in Chapter 2 ("Objectives"). Although it varies from country to country a list of those variables recorded by the Dutch police is presented in Table 1 as illustration and because it presents a fairly representative picture of what the police record.

3.2. The Sub-Files

These will be treated in the research order of Pre-Crash, Crash, and Post-Crash as an indication of for what sort of research and measures the data can be used for: viz. Accident Prevention, Injury Prevention, and Injury Treatment.
Some sub-files contain information also available in the police accident data but in (far) more detail. Each sub-file contains information which the police do not record at all. Some sub-files record accidents which the police do not know of.
Each sub-file contains at least one variable which is identical to a

variable in the police data - the common key variable. The content of each potential sub-file in the Netherlands is also to be found in Table 1.

3.2.1. The Exposure Sub-File

Exposure is the extent to which certain groups of road users and vehicles travel or conditions occur whereby they are exposed to the dangers of road traffic. It is usually measured in time and/or distance or in the number of occurrences so as to measure relative safety such as death rates, and to be able to compare groups and periods of time.
This sub-file should include the extent to which those involved in road accidents travel, of course subdivided according to modal split, time of day, age, etc. etc. The exposure of the accident group can then be compared with the total population (or a representative sample) or groups without accidents to see what the similarities and differences are, as an indication of contributory factors to accident involvement.
The common key variable is a name and date of birth of each person involved irrespective of whether injuried or not, or a vehicle number to determine its mileage.

3.2.2. The Road Characteristics Sub-File

Road characteristics are kept on file by Ministries of Transport or Road Authorities who have jurisdiction in a particular area such as a borough, country state, or region; or have jurisdiction over a particular type of road throughout the country such as motorways or harbour roads. Each road authority registers certain characteristics of the roads under its jurisdiction such as the length, width, camber, road surface, and the amount and composition of traffic on them.
If this kind of detail is linked to the police accident data it is possible to compare the characteristics of roads or crossroads with many, or many serious accidents with those with relatively few accidents.
The common key variable is the exact location of the accident to within a few metres for crossroads, and within a few tens of metres along a stretch of road. This location is usually expressed as a grid reference or some other numerical definition of a crossroads, or for a stretch of road, a road number in combination with a milepost system.

3.2.3. The Vehicle Characteristics Sub-File

Practically all countries have a central vehicle registration which in some cases goes further back than the accident registration. This is usually maintained by the Ministry of Transport or indirectly through local authorities. Details of the present (and sometime past) owner is known as well as various characteristics of the vehicle itself. These characteristics include vehicle identification data such as chassis number and registration number together with variables such as type, weight, length, width, colour, age, make, and model.
Having linked these characteristics to the accident data it is possible to analyse which sorts of vehicle have the highest, or lowest, involvement in accidents.
The common key variables are either the (unique) chassis or registration number. In countries where a registration number is allocated to an owner rather than a vehicle the chassis number is to be preferred.

3.2.4. The Injuries & Treatment Sub-File

Victims who are seriously injured, or where serious injury is suspected, are sent for treatment to a hospital where they are examined and if necessary admitted, or treated as out-patients.
Hospitals keep records of their patients: how they came by their injuries, what their injuries were, and which treatment they recieved. In some countries this and other data are collated regionally or nationally under the auspices of the Ministry of Health.
Linkage of this data to the police accident data means that it is possible to study the relationship between the type of accident and the resulting types and seriousness of injury; i.e. why certain types of accidents result in certain types of injuries or in extremely severe, or light, injuries.
The necessary common key variable is the name and date of birth of the victim.

3.2.5. The Material Damage Sub-File

The main source of information on damage to vehicles, lampposts etc., and

buildings are the insurance companies. They of course record only those accidents for which a claim is made or otherwise comes to their knowledge. The number of these accidents is in The Netherlands, and probably in other countries, considerably greater than the number of accidents recorded by the police.
In Western Europe there is even a standard accident claim form, the European Damage Form, which should always be used even if the individual insurance company also has its own form. This form contains information about the owner, the vehicle, the accident, a description of the damage together with (an estimation of) the repair costs.
After having linked this information with the police accident data one will know which types of accidents result in what types of damage and in extremely heavy, or light material vehicle damage. It is not certain to what extent it can be used for the damage to objects such as lamp posts and buildings.
The common key variable is the registration or chassis number of the vehicle, or in some cases the name and adress of the owner.

3.2.6. The Long Term Effects Sub-File

Convalescent homes and interviews with the victims themselves are the main sources for long term effects in terms of physical or mental disability. If the patients have not moved house since treatment, hospitals and doctors are also a source for long term effects. The most important information to be gathered is the type, severity, and duration of the disability.
Once linked we would know which types of accident lead to long-lasting or permanent disability and which do not. The common key variable is the name and date of birth of the victim.

3.2.7. Other Possible Sub-Files

There are other possible sub-files which do not appear in Figure 1 because they are either not considered important, or are non-existent in the Netherlands. A drivers' licence file could be useful if it contained all known accidents in which they had been involved in to see if recidivists exist and if they do, what their distinguishing characteristics are.

Other possible sub-files are court cases, public health files and especially that of the cause of death, weather and climate, medical records of drivers, and garage records.

4. ADVANTAGES

4.1. For the Police

The main advantage for the police is that they can record less details of accidents than they do now: all that is needed, apart from that data necessary for their own purposes, are (see also Table 2):
(a) that data which can only be gathered at the scene and time of the accident;
(b) common key variables (between central and sub-files);
(c) control variables (relevant for (a) and (b)).

Examples of data to be gathered at the scene and the time of the accidents (a) are:
- the road and weather conditions at the time, whether it was busy, if the traffic lights were working, whether it was light, dark or twilight, if the street lighting was on, whether there were men at work, if it was raining, and if the road was wet or icy;
- who was involved in the accident (details of drivers, pedestrians, and vehicles), and the manoeuvres each one made;
- the outcome of the accident in terms of injury and damage.

Examples of common key variables (b) are: names and addresses for linking accidents to exposure, hospital records, and insurance companies; vehicle registration or chassis numbers for linking accidents to vehicle files and insurance companies; location identification for linking accidents to road files.

Examples of control variables (c) relevant for (a) are time of day and date to check the light conditions, and day to check the date.
Examples of control variables (c) for (b) are date of birth to compare with name and address, grid reference to check with street names and numbers, and insurance policy number to compare with vehicle registration numbers.

4.2. For the User

Linking data files to each other is a classic example of the Law of

Increasing Returns: the combined value of the sum of the linked files is far in excess of the sum of the individual files.
The advantages to the user be he researcher, in government, or in the police itself is that an accident file per accident is created with detailed information on the circumstances of the accident itself, who was involved, what they did, and with what results. This file always contains far more information than the police can ever hope to record. Questions can be answered concerning the road conditions of accidents in which certain types of road users in certain types of vehicles carry out certain types of manoeuvres and sustain certain types of temporary or permanent injury with certain damage to their vehicles.
A concrete example could be: do older men driving large cars sustain other knee and chest injuries on wide wet roads than on wide, dry roads after braking suddenly for a child crossing the road.
These sort of questions, involving various aspects of accidents, can rarely be answerd from the police accident data on its own.
Being able to answer more questions, the combined linked files will be used more than the individual files thus making it easier to introduce further improvement or additional information in one or more of the files. It is easier to improve a much-used than a less-used data collection.
The increased use will not only come from those engaged in road safety research but from transport planners for whom the number and type of accidents is just one of the aspects they have to take into consideration. Instead of having to stick pins in a map they can get print-outs of particular roads or road sections giving details of the accidents along the route together with other road data such as surface and width.

A longer term advantage is that those responsible for the individual files will have the tendency to take the other linked (sub) files into account when planning any changes. In his own interest if he introduces, for example, a new characteristic he will check that the subdivisions match any ones existing in the other (sub) files. This will lead to a greater amount of standardization whereas with unlinked files one has the tendency to think only of one's own purposes and needs.

5. REQUIREMENTS

5.1. Quality

The most important requirement if police accident data are to be linked to other files is that the quality of the police data itself be high. Quality can be defined here as a combination of validity, accuracy, and representativeness. Validity means that if the police record that a car was involved, that it was not in reality a van and that if they say it was red it was not yellow. Accuracy means if the driver was stated to be 35 he was not 36 or 53 and that the skid marks were reported to be 10 metres long whereas they were in fact 15 metres. Representativeness means that the police accident files contains the same proportions of particular types of accidents and victims as in reality, or that if for example serious accidents are overrepresented, that this is known together with the extent in which it occurs.

If these quality requirements for the police accident data cannot be met it is doubtful whether it is worth going to all the trouble of linking other files to it. Of course linkage will result in more information but the reliability of the information, especially for research purposes, will be questionable in the least.
The quality requirements also apply to the sub-files but is of the utmost importance in the case of the police accident data because they form the main or central file of the Integrated Road Accident Recordkeeping System.

5.2. Continuity

The returns on linkage are in general higher when all the files and sub-files are continuous collections and not ad hoc. The fixed costs and effort are the same for both but the variable costs are lower and the benefits higher in the case of continuous files.
Continuity of what each file contains and the amount of detail and the meaning of each detail is also necessary. It is extremely awkward if contents and meanings change every year especially if time series are important in the safety studies (as they often are). New improvements in

a file should not of course be impeded by the desire for continuity, but the past coding should be taken into account when introducing an improvement.

5.3. Uniformity

It is of the utmost importance that terms and subdivisions of terms have the same definition in each file. If that is not possible then all users should know the differences. Unless otherwise stated a "motorway" must mean the same in all files, and the method for locating a particular spot on a motorway must be the same in all files. If one file makes no distinction between cars and vans whereas others do, then it should also introduce the distinction, using of course the same definitions for car and van. If age groups are used instead of the actual age (or date of birth) the groups in all files should be comparable. In general, however, such groups are to be avoided - disaggregated data gives more information and is more flexible.

5.4. Common Key Variables

Without these unique identifications linkage cannot by definition be carried out. As we have seen the most important for road accident data are: location of accident; name, address (and date of birth) of those involved or injured; vehicle registration or chassis number.
The ideal common key variable is the hotly - disputed personal identification number which everybody would carry with him from the cradle to the grave.

If they do not appear in the police accident data or the sub-files they must be introduced or surrogates must be used (see below).
If common key variables exist but are not available, for example for privacy reasons, this can be overcome in one or two ways:
1. confidential data is handled by those authorised and is removed from the file after linkage; this can be done in the case of names and addresses and licence or vehicle registration numbers;
2. surrogates are used in the form of a combination of characteristics in which the characteristics themselves are not unique but the combination

is, or so rare that mistaken indentity is virtually impossible. An example of this is "a 1982, red and white Volkswagen Golf GTI, with a sun roof, smoked glass, fitted for LPG, with a double exhaust, and wide white-wall tyres".

5.5. Other Requirements

Not essential, but useful requirements for successful linkage are:

1. Centralized sub-files:

Just as the police accident data is centralized in practually all countries it is much more efficient if the sub-files are also centralized. If they are not, each local or regional sub-file will have to be linked separately to the main file - it is possible but more complicated. It is better if the local or regional sub-files can first be amalgamated into one national sub-file. This also applies to the police accident data if it is not centralized.

2. Accessibility

Data from the sub-files should be linked to the police accident data as soon as possible and regularly: for example once a month, quarter, or year so that there is a rapid updating.

Each contributor to the Integrated Road Accident Recordkeeping System should have instant and easy access to the complete data base and not only his own sub-file, either by on-line access or by placing orders from the data base host.

It goes without saying that the information on the sub-files, when linked to the police accident data, has to be relevant for road safety research and policy.

6. THE SITUATION IN THE NETHERLANDS

6.1. The Quality of the Police Accident Data

In the Netherlands the police send their road accident registration forms to the Road Accident Registration Office of the Ministry of Transport which processes the data. The data is published by the Central Bureau of Statistics.
For as far as is possible because of the differing definitions, it has been established (Maas & Harris, 1984) that is no reason to doubt that the police record every fatal road accident and every road accident fatality. Furthermore the Central Bureau of Statistics conducts a continuous check on the police fatal accident data by comparing it with records from the Public Prosecutors' offices, the wire service, and newspaper reports.

It has also been shown (also in Maas & Harris, 1984) that the completeness of the police accident data as far as traffic victims admitted to hospital (in-patients) are concerned, is stable throughout the years making it reliable for longitudinal research. Also was shown that the completeness was high (83% in 1979) decreasing the chance of bias in comparison with all road accident in-patients. It was however not completely representative - motor vehicle occupants from 15 to 34 years old were slightly overrepresented and child pedestrians and child cyclists were slightly underrepresented. The bias of the police accident data was however established and is taken into account when using the data.

The completeness and representativeness of police accident data on those injured but not admitted to hospital is as yet unknown but will be measured in a nationwide continuous survey from July 1986 to June 1987.
Only rough guesses are possible as to the situation regarding material damage only accidents.

The validity and accuracy of the police accident data has never been systematically examined. Ad hoc comparisons have been made between the type of vehicle as registered by the police and vehicle registration file, between the date of death in comparison with the cause of death

statistics, and between the numbers of roads intersecting at a crossroads in comparison with the road authority data. In none of these cases were many mistakes found and sometimes, where the files disagreed, it was not possible to be sure which was correct.
Ad hoc controls are also carried out within the police accident data such as: agreement between date and day of the week, agreement with light conditions and the combination of date and time of day, agreement with type of road and speed limit, agreement between details of drivers and driver-victims, agreement between weather and condition of road surface, agreement between manoeuvre and point of impact, and comparison between the date of accident and the date of death (that the latter later is than the former) and in relation to the 30 day definition. Discrepancies were found of course - no file is perfect - but not to such an extent that they influenced research results. Discrepancies also lead to extra controls built in to the input programmes.

6.2. Linkage of Police Accident Data and the Vehicle Registration File

Having established that the police record all fatal accidents; that their was a centralized, computerized vehicle registration (except for mopeds and bicycles); and that the registration number of motor vehicles is to be found in both files, a trial linkage of the two files was successfully carried out by the SWOV for all fatal accidents in 1981 (Lindeijer, 1986).

The linkage had to be technically possible because they were both centralized, computerized files and there was a unique common key variable. Furthermore, in the Netherlands, the vehicle registration number stays with the same vehicle all its life. The trial linkage was aimed at determining whether it was also practically possible. It was furthermore aimed at determining the completeness and reliability of both files. The relatively small number of registration numbers (c 2000) because of the restriction to fatal accidents, made it possible for the SWOV to also carry out a manual control on make and model. The police often record it but this is not coded on to tape.

Based on the conformity found between the two files, both the validity

and reliability were found to be approx 95%. It was therefore concluded that linkage between fatal accidents and the vehicles involved should be continued annually and that there should be a trial linkage between accidents in which at least one victim had been hospitalized, and the vehicles involved.

The vehicle characteristics which can now be added to the police accident data is shown in Table 1. The list is not all that long but does contain some essential data for crash research such as weight, length, and wheelbase which the police do not record. By establishing the make and model many more detailed characteristics can be obtained from the government body issuing production and import licences because these are based on the compliance of a vehicle with government regulations. Further details are also available from the manufacturer or importer. These details which are available, but not in the vehicle registration sub-file can be regarded as a sub-file of this sub-file, or a sub-sub-file!

One of the first uses of the linked accident-vehicle file was to look at the age distribution of cars involved in fatal accidents (Lindeijer, 1985). Of course, the age of the car is not the only important variable - the age of the driver is also of great importance. These tables are presented to illustrate the possible results of linked files It was found that their average age was about one year older than the vehicle population weighted for annual mileage (Table 3). It was also found that the involvement of cars in fatal accidents increases with the age of the car: from c 1 per 10^8 vehicle kilometres for cars less than one year old to c 4 per 10^8 vehicle kilometres for car 10 years and older. The fatal accident involvement of cars decreased for all ages of car during the period from 1978 to 1982. The decrease was the least in the case of cars less than 2 years old and older than 7 years (Table 4).

6.3. Linkage of Police Accident Data and Hospital In-patient Data

In 1986 a trial linkage will be carried out between those 1984 road accident victims recorded by the police to have been hospitalized (c 80% of all such cases, see 6.1) and the records of the Centre for Health Care Information (S.I.G.), formally the Medical Records Foundation (S.M.R.).

The hospitals participating in and contributing to the centralized, national file of the Centre for Health Care Information accounted for 99% of all hospital admissions in 1984. Their file can therefore be regarded as complete.

1984 has been chosen as the trial data year and in that year the police reported c 15,500 hospital admissions among road accident victims. These will be linked, where possible, to the correct patient among the nearly 21,000 admissions of road accident victims in the hospital data.
It is not so certain that the trial linkage will be a success, as in the case of the linkage with the vehicle registration file (see 6.2). This is because there are no unique common key variables available to SWOV. They do exist, in both the police and hospital files viz. name, address, and date of birth of the victim, but of these the name and address are confidential. This means they are not present on the copy of the police accident registration form which is made available for coding, and although of course known by the individual hospital, are not made available on the discharge form which is sent for coding to the Centre for Health Care Information.

We will therefore have to create a surrogate consisting of a, hopefully unique, combination of characteristics. These are: date of birth, sex, and the name of the hospital. If this combination proves to be insufficient we will then use a number of 'reserve characteristics' which do not have quite the same meaning in the police file as in the hospital file. These are 'time of accident' vs. 'time of admission' and 'date of accident' vs. 'date of admission' (there can be a delay of hours and sometimes days between the accident and being admitted). We will also use the "category of road user" code in both files but the coding in the police file is much more detailed (124 categories in 12 main categories) than in the hospital data (10 categories). Furthermore not all the 12 main categories of the police fit into the 10 categories of the hospital, and in more than 10% of the cases the hospital does not know the category of road user.

If after having used the 'reserve characteristics' a number of victims has still not been linked, an attempt will be made to locate them under

the out-patients (the police may think someone was admitted because the hospital intended to, but later decided that the injuries were not of the type necessitating admittance). Out-patient data is however not yet centralized.
Using the hopefully unique combinations, the 'reserve characteristics', and the out-patient records we are confident that we will be able to link practically all the victims.

The linkage itself will be carried out, for privacy reasons, by the Centre for Health Care Information so that there will be no leak of identifiable patient data. The Centre for Health Care Information will recieve a copy of the 1984 tape of accidents with non-fatal injuries from the Road Accident Records Office which processes the police accident data. The inclusion of victims who according to the police were not admitted to hospital has been agreed upon because the Centre expects to find that some of these were after all admitted and therefore on their file and linkable.

Although it is not the goal of the linkage, the results will be used to check the validity and accuracy of both files. Where characteristics disagree between the two files it will be determined which file was 'right'. If for example the police say the victim was a cyclist and the hospital says it was a pedestrian the police version will be preferred because they were at the scene of the accident. If the police say the victim was admitted to hospital A whereas the hospital data say it was hospital B then the hospital data will be believed. If the files disagree for example on the sex of the victim, further investigation will be necessary to determine who was right. The same applies if the ages differ or the date of admission is earlier than the date of the accident.

The additional information which becomes available when linkage has been completed as shown in Figure 1. The length of the list is not impressive but the value for road safety research purposes is considerable. This is especially so for the injuries (the so-called E-code of the International Classification of Diseases) and the 'length of stay' as a measure of the seriousness of the injuries (in conjunction with the E-code). The 'type of discharge' indicates if the patient died in hospital, was sent on for

further treatment or rehabilitation elsewhere, or went home. This is extremely useful for studies on further treatment and any long term disabilities.

6.4. Exposure

The Netherlands has a National Travel Survey, carried out on a continuous basis since 1978 by the Central Bureau of Statistics (CBS, 1985). The information it contains can be found in Table 1.
It is however carried out among a sample of the Dutch population of 12 years and older irrespective of whether they have been (ever) involved in an accident. It is therefore not suitable for linkage and was set up as a measurement of mobility for transport study purposes, and exposure for road safety purposes. It's main purpose for road safety purposes is to be able to relate types of accidents and victims to their mileage and time spent in traffic.

6.5. Road characteristics

There is no centralized file of road characteristics in the Netherlands. Each road authority: state, province, borough, and some others has its own file. They are often manual files (lists, maps) and often not compatible with each other. The Central Bureau of Statistics has aggregated records with type of road and road width. The road typology is however different from that used by the police accident data.
The Road Accident Records Office which processes the police accident data has a computerized file of roads and road locations but no road characteristics.
An improvemant in this situation is not to be expected in the foreseeable future.

6.5. Material damage

The Dutch insurance companies have a centralized, computerized file of road accident claims covering the great majority of insurance companies. A file based on claims means that less serious accidents with little material damage are not recorded. Neither are accidents with uninsured

vehicles or vehicles such as bicycles which in the Netherlands do not have to be insured.
This data, whatever its value may be for road safety purposes, is closed to all except the participating insurance companies themselves.
We hope that this will change in the future especially considering the fact that the insurance companies recieve a copy of every police road accident form.

6.7. Long Term Effects

The information from convalescent homes are collated by the Ministery of Health but only in the form of the situation on January 1st of each year. It is therefore not suitable for linkage. A possibility however is that the Centre for Health Care Information expands its registration to include convalescent homes. This is already in the trial phase but may take a long time to become operational.

6.8. Other files

As stated in 3.2.7 these are either considered unimportant, or are non-existent in the Netherlands. Further investigation may lead us to be more active in this field, especially in the case of some sort of drivers' file.

7. FINAL REMARKS

In 1985 IDBRA (International Drivers' Behaviour Research Association) conducted a survey of 12 OECD countries to establish how far each country was with linking other files to road accident data. There has, however, as yet, been no report of the findings.

In conversations and correspondence with road safety researchers in a number of countries it would seem that although the benefits of linking are plain, the possibilities are often less than in the Netherlands. Either there are no centralized sub-files, or they are not computerized, or they do not contain common key variables, or the files are not available for privacy or other reasons. Sometimes doubt is cast on the reliability of the sub-files - or even the police accident data. This means that linking the police accident data to other sub-files to form an Integrated Road Accident Recordkeeping System will take a lot of time and effort.

It is worthwhile because apart from the additions to the police accident data there are all sorts of secondary benefits: one learns much more about other data collections relevant to road safety and about the police accident data itself, uniformity between files is achieved or at least improved, and the files can be used to check each others' quality.

LITERATURE

Literature cited

Andreasen, E. (1975). Road Accident Statistics on a National and International Level. E.C.E. Symposium, Brussels, 11 December 1975.

CBS (1985). The National Travel Survey in the Netherlands. Central Bureau of Statistics, Heerlen, 1985.

Jacobs, G.D.; Bardsley, M.N. & Sayer I. (1975). Road Accident Data Collection and Analysis in Developing Countries. No. LR 676. TRRL, Crowthorne, 1975.

Lindeijer, J.E. (1985). The Age Distribution of Cars involved in Fatal Accidents in 1978-1982. SWOV, Leidschendam, 1985. (Available only in Dutch. SWOV No. R-85-16)

Lindeijer, J.E. (1986). A Trial Linkage of the Road Accident and Vehicle Registration Files. SWOV, Leidschendam, 1986.
To be published in Accident Analysis & Prevention, 1986. (Published by SWOV in Dutch in 1983. SWOV No. R-83-47)

Maas, M.W. & Harris, S. (1984). Police Recording of Road Accident In-Patients. Accident Analysis & Prevention 16 (1984) 3 (June) pp. 167-184.

Other Useful Literature

Blokpoel, A. An Integrated Traffic Element Registration System. SWOV, 1978. (Available only in Dutch. SWOV No. R-78-8.)

Nicholl, J.P. The Usefulness of Hospital In-Patient Data for Road Safety Studies. In: Foot, H.C. et al. (ed.). Road Safety, Research and Practice. Praeger, New York, 1981.

Transportation Research Board: Introduction to Comprehensive Computerized Safety Recordkeeping Systems. TRB Research Circular No. 293, July 1985.

TABLES

FIGURES

ACCIDENT (POLICE)	ROAD (ROAD AUTHORITY)	VEHICLE (V. REGISTRATION)	VICTIMS (CENTRAL HOSPITAL REGISTRATION)	CLAIMS (ASSOCIATION OF INSURANCE COS)	LONG TERM EFFECTS (INTERVIEWS)	EXPOSURE (CBS SURVEY)
TIME OF DAY			TIME ADMITTANCE	TIME OF DAY		TIME OF DAY
DATE			DATE ADMITTANCE	DATE		DATE
LOCATION	LOCATION			LOCATION		
URBAN/RURAL	URBAN/RURAL					
SPEED LIMIT(S)	SPEED LIMIT(S)			DRIVING SPEED		
ROAD AUTHORITY	ROAD AUTHORITY					
ROAD SITUATION	ROAD SITUATION			ROAD SITUATION		
SPECIAL SITUATIONS	GEOMETRY					DURATION
TEMPORARY CIRCUMSTANCES	CYCLE PATHS/ LANES	No. OF WHEELS				LENGTH
LIGHT CONDITION		No. OF DOORS				PURPOSE
STREET LIGHTING	STREET LIGHTING	WHEELBASE				
WEATHER	RIGHT OF WAY	COLOUR		VISIBILITY		VEHICLE OWNERSHIP
ROAD SURFACE CONDITIONS		NET WEIGHT				CAR: NET WEIGHT
RD. SURFACE TYPE	R. SURFACE TYPE	FUEL TYPE				:FUEL TYPE
TYPE OF ACCIDENT		No. OF CYLINDERS	TYPE OF ACCIDENT (ICD)	TYPE OF ACCIDENT	TYPE OF ACCIDENT	:ANNUAL MILEAGE
VEHICLE(S): TYPE	VEHICLES:	TYPE				
:COUNTRY OF REGISTRATION	:PERMITTED TRAFFIC MIX INTENSITY	MAKE & MODEL		MAKE & MODEL		
REGISTRATION No.		REGISTRATION No.		REGISTRATION No.		
DRIVER(S): (NAME)				DRIVER: (NAME)		
:SEX		IMPORTER				
:PLACE OF BIRTH		CHASSIS No.		USE OF SEATBELT		
:DATE OF BIRTH		YEAR				YEAR
:(ADDRESS)				(ADDRESS)		
:DRIVING LICENCE						
:USE OF ALCOHOL				USE OF ALCOHOL		
OWNER(S): (NAME)		OWNER: (NAME)	OPERATION(S):DATE(S)	OWNER: (NAME)	THERAPY: DATE(S)	
:(ADDRESS)		:(ADDRESS)	:TYPE(S)	:(ADDRESS)	: TYPE(S)	
:NATIONALITY		:DATE OF BIRTH				
:TYPE OF INSURANCE			TYPE OF INSURANCE	POLICY No.		
:NAME INSURANCE Co.				NAME INSURANCE Co.		
DESCRIPTION OF MATERIAL DAMAGE				DESCRIPTION OF MATERIAL DAMAGE		
VICTIM(S): (NAME)			VICTIM: (NAME)	VICTIM: (NAME)	VICTIM: (NAME)	(NAME)
:(ADDRESS)			:(ADDRESS)	:(ADDRESS)	:(ADDRESS)	(ADDRESS)
:SEX			:SEX	:SEX	:SEX	SEX
:DATE OF BIRTH			:DATE OF BIRTH	:DATE OF BIRTH	:DATE OF BIRTH	DATE OF BIRTH
:MODE OF TRANSPORT			:MODE OF TRANSPORT		:MODE OF TRANSPORT	MODE OF TRANSPORT
:TAKEN TO HOSPITAL			:MAIN INJURY (ICD)		:MAIN INJURY	EDUCATION
:ADMIT. TO HOSPITAL			:SECOND. INJURIES(")		:SECOND. INJURIES	OCCUPATION
:NAME OF HOSPITAL			:HOSPITAL CODE	NAME OF HOSPITAL	:NAME OF HOSPITAL	INCOME
:PLACE OF DEATH			:HOSPITAL BOROUGH		:NAME OF NURSINGHOME	
:DATE OF DEATH			:DATE OF DISCHARGE		:DATE END OF THERAPY	
			:TYPE OF DISCHARGE		:DISABILITIES(ICIDH)	
ACCIDENT: BRIEF DESCRIPTION				ACCIDENT: DESCRIPTION		

(in parenthesis) = confidential; ————= linkable

Table 1. Available data in files to be linked in the Netherlands

INTEGRATED ROAD ACCIDENT RECORDING SYSTEM

1. The police should only record:

(a) that data which can only be gathered at the scene and time of the accident

(b) common key variables (to sub-files)

(c) control variables (relevant to (a) and (b)

2. Other data should be recorded in sub-files

- Exposure
- Road Characteristics
- Vehicle Characteristics
- Injuries and Treatment
- Material Damage
- Long Term Effects
- Any other relevant data

Table 2. Division of work between police and other authorities.

	1979		1980		1981		1982	
	On the road	fatal acc.	on the road	fatal acc.	on the road	fatal acc.	on the road	fatal acc.
Same year	19.5%	9.1%	15.5%	6.4%	12.4%	6.2%	11.8%	6.7%
1 year old	18.0	14.2	17.7	15.0	14.6	9.6	11.5	10.4
2 years old	15.6	15.5	16.1	12.6	16.3	13.4	12.9	8.3
3 years old	12.2	12.4	13.4	12.3	14.6	12.2	14.0	12.0
4 years old	10.1	10.8	10.4	12.7	12.4	11.4	13.1	12.9
5 years old	7.5	9.7	8.8	11.0	9.3	12.6	11.3	12.3
6 years old	5.8	8.6	6.2	8.3	7.5	11.5	9.0	11.6
7 years old	4.7	8.1	4.8	7.7	4.9	6.6	6.5	8.1
8 years old	2.9	5.1	3.0	6.0	3.2	7.7	4.1	7.1
9 years old	1.9	3.7	1.8	4.3	2.1	3.8	2.3	5.6
> 10 years old	1.9	2.8	2.2	3.7	2.7	5.0	3.4	5.2
Total	100	100	100	100	100	100	100	100
Average age (years)	3.4	4.4	3.6	4.7	3.9	5.0	4.3	5.2
N	4.197 mill.	1615	4.313 mill.	1704	4.351 mill.	1471	4.406 mill.	1436

N.B. "Cars on the road" are calculated by multiplying the numbers of registrated cars by their annual mileage as measured in the Car Panel of the Central Bureau of Statistics (this panel gives the same annual mileage for all cars of 8 years old or older).

Table 3. % distribution by age of cars on the road in comparison with cars involved in fatal accidents in the Netherlands, 1979-1982.

	1979	1980	1981	1982
Same year	1.2	1.0	1.1	1.2
1 year old	1.9	2.1	1.5	1.9
2 years old	2.4	2.0	1.9	1.4
3 years old	2.5	2.3	2.0	1.8
4 years old	2.6	3.1	2.1	2.1
5 years old	3.3	3.2	3.2	2.3
6 years old	3.4	3.4	3.6	2.7
7 years old	4.2	4.1	3.1	2.6
8 years old	4.3	5.1	5.6	3.7
9 years old	5.0	6.1	4.2	5.1
> 10 years old	4.0	4.3	4.3	3.3
Total/Average	2.7	2.7	2.4	2.1

Table 4. Involvement by age of cars in fatal accidents per 10^8 vehicle kilometres in the Netherlands, 1979-1982.

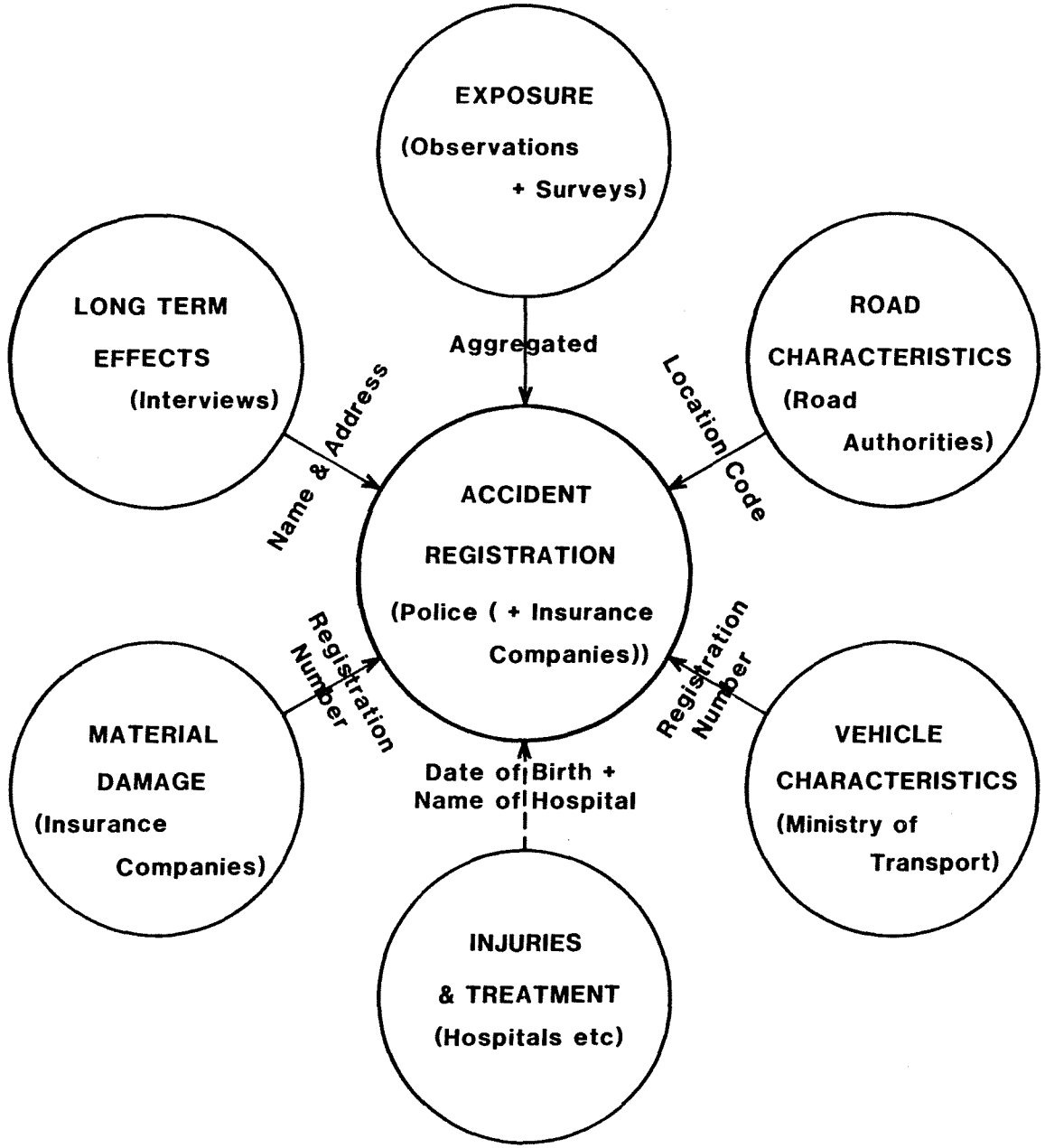

Figure 1. Diagrammatic Representation of an Integrated Road Accident Recordkeeping System.

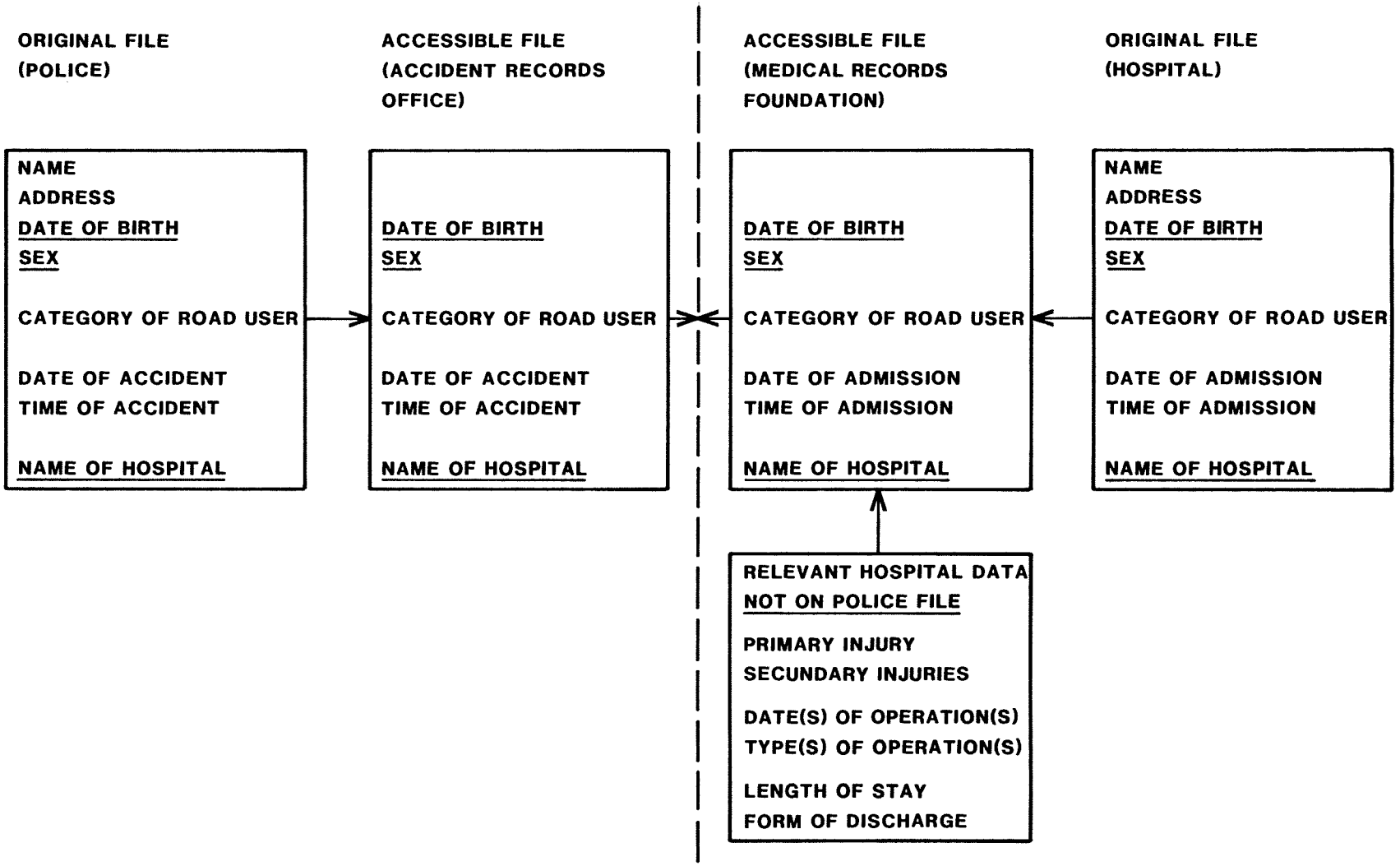

Figure 2. Linking Road Accident and Injury Data Files of Hospital In-Patients.