Final Report

Public

STAIRS
Standardisation of Accident and Injury Registration Systems

Contract N° : RO-96-SC.204

Project Co-ordinator	:	(FR)	INRETS/LEAT	-Gilles VALLET
Partners	:	(FR)	INRETS/LEAT	-Bernard LAUMON -Jean Louis MARTIN
		(FR)	CETE (SO)	-Philippe LEJEUNE
		(GB)	ICE	-Pete THOMAS -Rob ROSS
		(GE)	BAST	-Ingo KOßMANN
Associated partner	:	(GE)	MUH	-Dietmar OTTE
Sub-contractor	:	(GB)	TRL	-Barry SEXTON

Reference period : 1 October1996 to 31 December 1998

Report N° : FR 1

Date : February 1999

PROJECT FUNDED BY THE EUROPEAN COMMISSION UNDER THE TRANSPORT RTD PROGRAMME OF THE 4th FRAMEWORK PROGRAMME

Table of Contents

1 Executive Summary

A single European-wide crash injury database, or a compilation of individual, comparable databases, would be of exceptional benefit to the legislation process at an EU level. A direct, data driven approach would allow identification of any safety problems at an early stage, and facilitate a quick and accurate evaluation of any new or remedial measures, including legislation, that may have been implemented. The aim of STAIRS is to make the first moves towards such a goal. It involves the standardisation of in-depth road accident data collection and methodologies that would provide the core data and basic framework for crash injury studies. It is intended that existing groups will be able to adopt the protocol within their existing studies but it is anticipated that the groups will normally add to it to reflect special research concerns. The protocol is intended to facilitate rather than restrict any investigations, so it does not prescribe a particular approach that must be adopted in every case. Rather it provides a set of requirements, for the final datasets in particular, in terms of the core variables and the case selection criteria that will lead to comparability of datasets. This will bring about a European-wide agreement on how to collect this type of information, and if followed, will lead to comparable databases being held in each country that adopts and follows the protocol developed.

The initial objectives focus on ‘secondary safety’ or injury prevention. This means that it looks at methods to prevent injuries after a crash has occurred and not methods that try to prevent the crash from happening. However some tasks such as data quality, confidentiality and ethics, data collection methodologies and Work Package 2, Linkage with National Accident Databases, all have direct relevance to accident causation studies.

The work is divided into four packages :
Work Package 1 - In-depth Data Collection. This includes such areas as: A review of existing studies, the information to be collected, information collection methods, information accuracy, validity of the process, and confidentiality and ethics.

The main topic within this area is agreement upon the specific information to be collected as core elements in in-depth studies. The three studies that are part of the STAIRS consortium, all have differing aims and objectives; determined either by their sponsors, their governments, or both. For them to come to an agreement as to the minimum information requirements for a crash injury study, is a great achievement. This has resulted in the detailing of approximately 300 variables (and if necessary, their associated values), and agreement upon a meaning and context of the variable that was acceptable to all. In some cases there was no system in place that would accurately give the information the partners required, and a new system had to be developed. Other parts of the package dealt with important associated tasks including the question of the techniques to be used to collect the data, and why certain techniques are better for collecting certain data in some circumstances than others. Next came ensuring the accuracy of the data, the accuracy of the data collection, how it is transposed into electronic format, and how that information is managed. Finally, the problems of confidentiality and ethics was explored. The purpose of conducting crash injury studies is to prevent further suffering. However, this information can be very sensitive, and sometimes is not accessible to the researchers for their work. EU data protection regulations, as they are implemented in many countries, can restrict access to and the collection of data considerably. This poses a grave problem for crash injury research. It is recommended that such work as this be given a special exemption to these restrictions.

Work Package 2 - Linkage with National Accident Databases. This area concerns the linking of in-depth accident databases to Regional, National accident databases. The approach has only been applied to the 1st step of the linkage, between In-depth and Regional accident databases. For most in-depth studies, strict statistical requirements are not being met, but some linking to Regional or National databases may be possible. However, it is essential that the constraints of the data and the underlying assumptions are stated if Regional or National estimates are produced. Methodologically, the estimation approach should first be to determine if the sample of in-depth accidents are representative of the Regional population from which they are sampled. If the Regional level estimation was satisfactory then National estimates might be made. If the National level estimates also check out satisfactorily then the data may be suitable for pan-European purposes.

Weighting is required because the in-depth database is probably a biased sub-set of the Regional database and the Regional database probably a biased sub-sample from the National database. The selection of appropriate weighting variables will then reduce the bias. It cannot however, eliminate the bias caused because some specific accident data were never included. If estimates are to be made on a Pan-European basis then certain conditions will have to be met. The first is the acceptability of the National estimates, which in turn require the acceptability of Regional estimates. Only once these pre-requisites have been made can a Pan-European estimate be considered. Any constraints which apply at a Regional or National level will also apply at a Pan-European level. Further, all the constraints from different Countries must be taken into account, hence if one Country only has data on new cars then a Pan-European estimate including that Country can only be based on new cars. This does not necessarily mean that different sampling schemes used in different Countries cannot be combined at a Pan-European level, but it does mean that some care should be exercised before combining such data.

Work Package 3 - External consultation. The processes of external consultation that members of the consortia developed are the main topic of this Work Package. These range from: conference papers at prestigious events throughout the world; presentations to interested groups and organisations involved in, or associated with, in-depth crash investigation, a workshop attended by a selection of delegates from the areas of research, government and manufacturing; and the development of a World Wide Web site (www.ice.co.uk/stairs) in order to disseminate the groups work and findings to as large an audience as possible.

Work Package 4 – Workshop. The idea of the workshop was to begin a two-way discussion between interested parties and the STAIRS partners. It was hoped that from these discussions the protocols developed could better reflect the wants and needs of a wider target audience as well as being a test of the validity of the work done so far. The feedback from the delegates was excellent and any changes required were on a detailed level. This also shows the care and dedication the group as a whole has given to the work, in that it was acceptable to manufacturers, governments and other research establishments with little or no changes required.

2 Project Objectives

The overall aim of the STAIRS project was to define the fundamental requirements for a Pan-European In-Depth accident and injury database. In order to achieve this, certain objectives had to be met by each Work Package:

2.1 Work Package 1 - In-Depth Data Collection

The objective of this Work Package was to develop a protocol for the collection and processing of injury and vehicle data from road crashes, that specifies the criteria necessary for data to be included in a harmonised European database or used to support EU policy. The protocol defines procedures to be used for case selection, data collection and the link to national accident populations. It also defines a core data set with reference to the needs of the Commission and provides a glossary that indicates the data collection methods to be used. It examines the manner in which data from several groups can be combined and analysed. A code of practise for accessing and processing confidential data has been drawn up. The protocol has been validated using a small number of existing crash injury cases in the UK, France and Germany. Recommendations for future action that can be used to lead to the construction of a combined crash injury database are made. The main deliverable is a Work Package report on each section relating to the tasks in the Work Package.

The Work Package concentrates on the systems and data necessary for the evaluation of injury prevention measures, not accident prevention measures. Some aspects of accident investigation, such as case selection and the link with national samples will also be equally relevant to primary safety studies, these sections will be highlighted in the Work Package final report.

2.2 Work Package 2 - Linkage with National Accident Databases

The objective of this Work Package was to produce a protocol for the establishment of a strong relationship between a future Pan-European harmonised injury database and the national databases. The aim of this protocol was the generalisation of the results obtained from in-depth investigations at first at a national level and then a European level. The final objective was to supply a method of evaluation of different measures and legislation and to provide a solid base for future legislation.

This protocol is made up from three main themes:

1. The elaboration of the recommendations for sampling; i.e. the selection of crashes to be studied and those which will be introduced into the database, in order that they will be representative of the national databases or a carefully defined subset of the national databases.
2. To determine the necessary conditions for using the statistical dimension of the national databases in order to generalise the results obtained from an in-depth investigation.
3. To determine and describe common variables with the same definition in national and in-depth databases.

The Work Package also validates the protocol by evaluating the statistical link between the three data sets represented by the partners. This Work Package has to take into account the work corresponding to the task in Work Package 1.ii, with particular attention to the possible incompatibility between the objectives; clinical or statistical. This Work Package focuses only on statistical aspects.

2.3 Work Package 3 - External Consultation

There are several crash injury data collection studies being conducted by car manufacturers and other organisations. These groups each employ data collection methodologies appropriate to the needs of the funding group, it is probable that there will be a number of areas of incompatibility within the harmonised procedure. The objective of this Work Package was to ensure that these groups involved in crash injury research are informed about the progress of each task in Work Packages 1 and 2 as milestones are reached, and that they were invited to comment and suggest amendments as necessary.

Upon the completion of each task, groups who were involved in crash injury research and those that have an interest of substance in this area will be circulated details of the draft harmonised protocol. A variety of means were used to disseminate the results of the project. These included special mailings, e-mail, the construction of an Internet Website and, where possible, academic publications. The external groups were encouraged to comment on the proposals which were then modified in the light of the comments received.

There were nine stages of consultation altogether following the completion of each task in Work Package 1 and two from Work Package 2. The completed process of consultation constituted a deliverable which was reported on completion of the Project.

2.4 Work Package 4 - Workshop – Presentation of Results

In order to be useful, the recommendations of the project had to be widely diffused. As a compliment to Work Package 3, the objective of this Work Package was to inform the interested people about our completed results and conclusions.

The method of dissemination chosen for the complete results that was the most efficient was a workshop that could be attended by all the interested parties. The idea of a workshop means that there could be two-way discussions.

A workshop also allowed the Commission to review the results and reflect on the further work to be carried out on the Pan-European harmonisation of accident and injury databases. This Workshop constituted a deliverable at the end of the project.

3 Description of Project

The project consists of five main work packages; two main areas of research and a further three areas of dissemination and implementation. The research areas comprise of the major problems identified as requirements for a successful harmonised data collection system. These are: the information collection and handling processes, and the statistical methodology to enable linkage to other databases. The final three workpackages deal with disseminating the findings of the research areas and suggest further work required to enable implementation.

During the life of the project, the aims and objectives initially set out in each Work Package evolved due to the on-going finding of the group. Any significant shift in the focus of the individual Work Packages has been mentioned within the text.

The following chapters will give a description of each of the Work Packages along with any conclusions and recommendations that may be appropriate.

4 Work Package 1 – In-depth Data Collection

4.1 Work Package 1.i – Review of Existing Studies

4.1.1 Table of Contents

4.1.2 Executive Summary

The main purpose of this task was to review, evaluate and itemise the degree of comparability that exists between the real-life crash studies carried out by independent groups in the UK, France and Germany.

Areas studied are: the objectives of the three studies, their case selection, accident investigation procedures, the data collected; its use as well as the quality control and error checking systems in place. The differences in local and national ethical considerations are also discussed.

There are many areas of comparability throughout all three studies but only in the area of secondary safety. At a conceptual level there are many areas of similarity, but at the detailed level there are many differences.

Agreement on the objectives of collecting the data should point to which aspects of the crash are to be used and at what level. Consensus will then be required as to the methodology of investigation. The common focus concerns car occupant injuries and there are many areas of comparability. The main areas of lower comparability concern case selection procedures and the inherent differences previously discussed. The CCIS does result in a more cost effective and statistically efficient model; whereas the random sampling procedures of INRETS and MUH allows for primary safety studies. As all partners do post-crash investigations on at least some of their cases, this would seem to be the best method to employ. Some on-site, post-crash recordings may be integrated into the system to support an expansion into accident causation data collection.

Next, agreement as to the mathematical model or software used for calculating crash severity would define at least some of the variables that are required for collection (unless the programs are non-contradictory). As STAIRS is fundamentally a secondary safety study, analysis of the performance of the vehicle and its safety mechanisms are obviously important factors. These two points go some way to providing a lead as to the type of methodology to be employed collecting this essential data. It was, then, not too difficult to focus on key measurements required by all three partners. The problems met were the precise understanding of terms used. This identified a further task dealing solely with crash severity terms and their meanings, and is given on the CD Rom.

Differences in methods are likely to remain as they are a consequence of local practicalities as well as technical precision, but these can be overcome. Some movement from all partners will have to occur if the project is to be successful. It will be the drive and professionalism of each team that will allow this to happen, as well as this ability to be flexible. The aims and objectives are achievable, if at a cost to each partner, and a European-wide database will be beneficial to all.

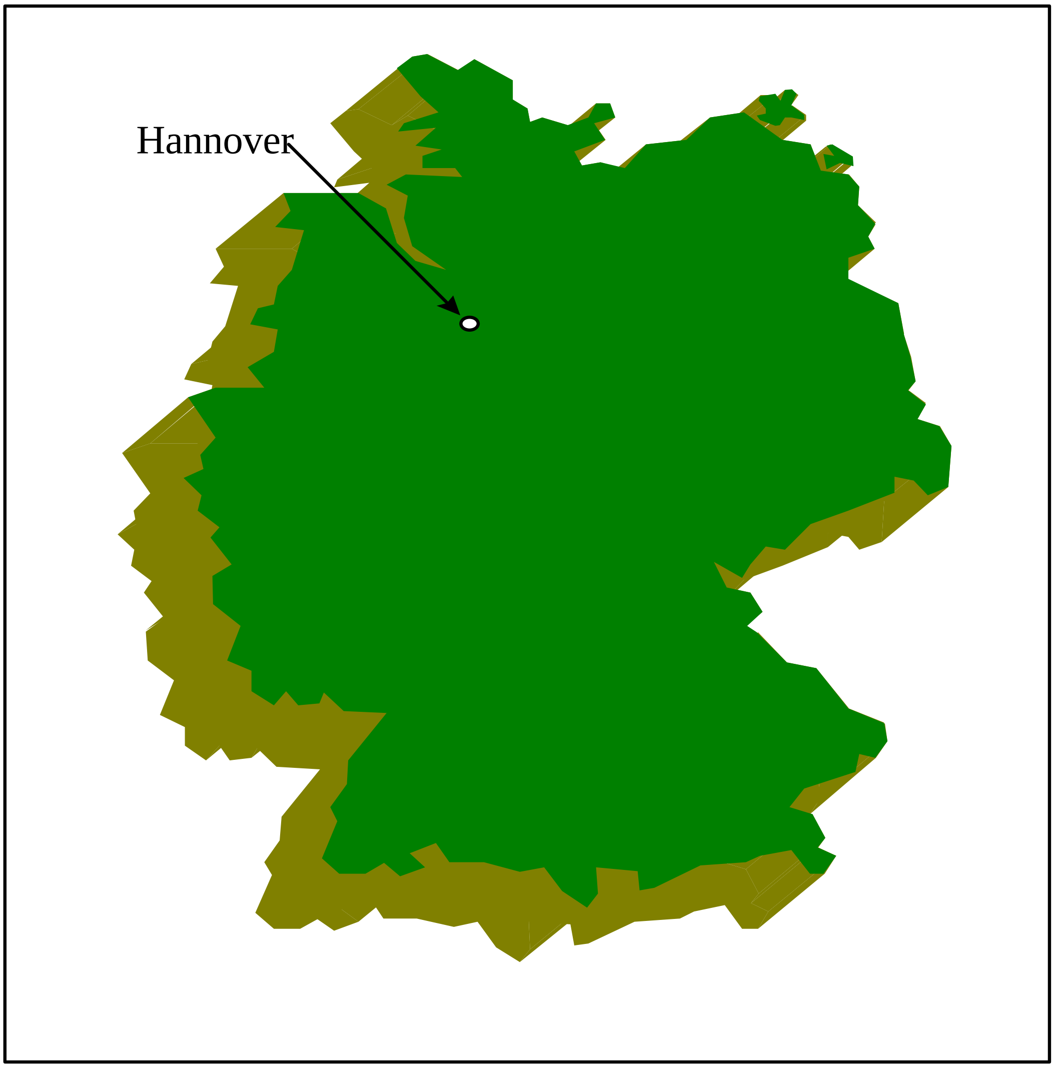
4.1.3 Introduction

Comparing any three disparate systems can be difficult, especially over three countries. Each has differing legal, ethical and practical considerations to be taken into account. Direct comparisons can be made in few cases, but indirect links and inferences are possible that give a general idea of the comparability of each. To begin with some background knowledge of each member would be helpful.

Germany has, at the Medical University of Hanover, the longest running, consistent single programme having begun in 1973, but only began collecting ‘full’ in-depth data using a statistical random plan from 1985.

The collection area is bound by the commune border of Hanover, approximately 2,289 km^2. It has a population of 1.2 million and is closely representative of the national population in terms of its percentage of urban to rural areas. The current rate of crash investigation is estimated at 1,000 per year within which 1,600 vehicles, 2,400 individuals and 4,500 injuries are examined.

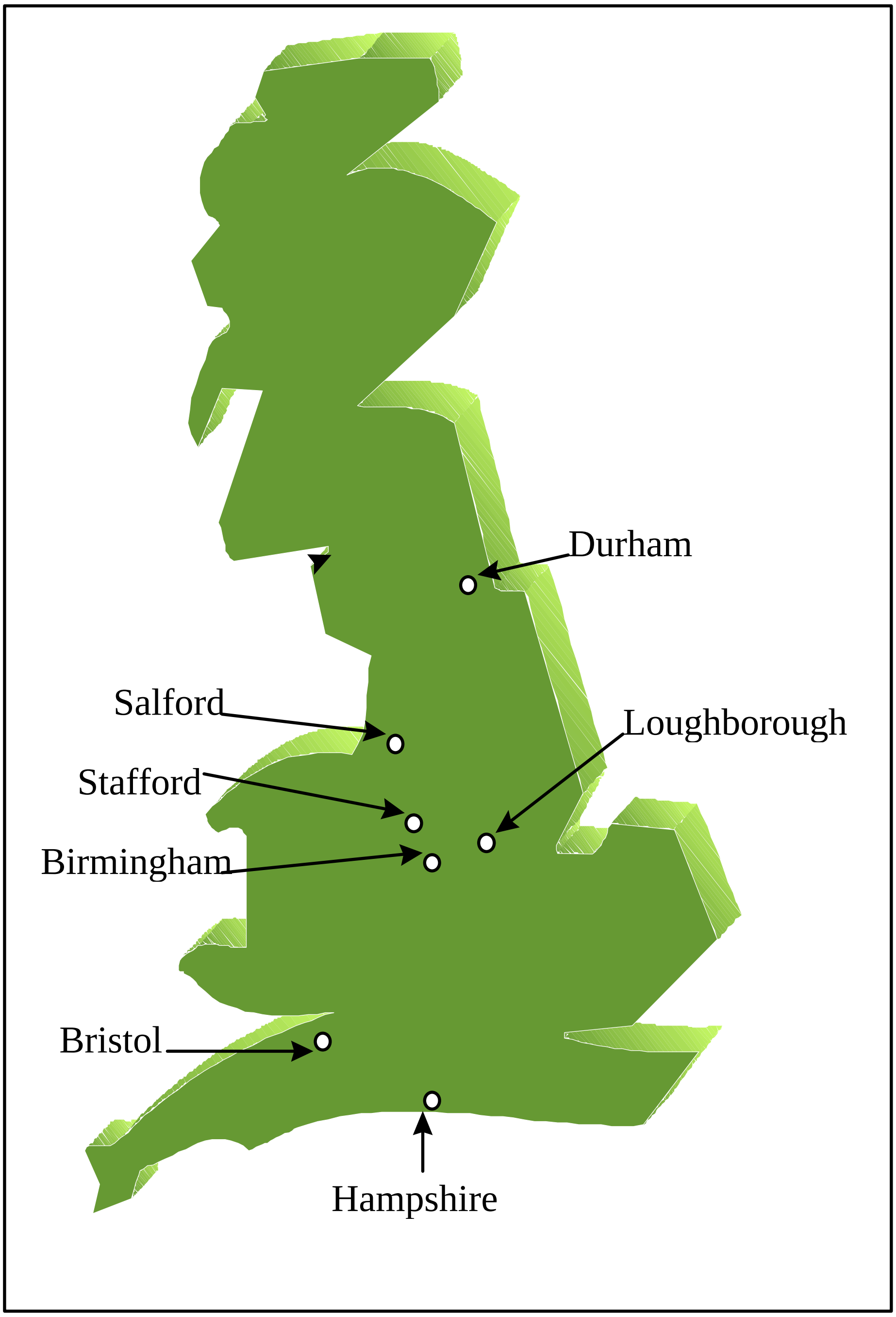
Map 1 : German collection centre.



The investigation is funded by the Federal Highway Research Institute (Bundesanstalt Für Straßenwesen - BASt).

The **UK**'s Co-operative Crash Injury Study (CCIS) project began in1983, although previous in-depth studies had been in existence since the 1960's.There are two main collection teams based in Loughborough (Vehicle Safety Research Centre - VSRC) and Birmingham (Birmingham Accident Research Center - BARC). There are a series of five smaller groups, the Vehicle Inspectorate Executive Agency (VIEA) in other areas around the country. All information is collected in the same manner and centrally managed by the Transport Research Laboratory (TRL). The total number of crashes investigated per year is approximately 1,600, involving 2,080 vehicles with 3,120 occupants sustaining 12,480 injuries.

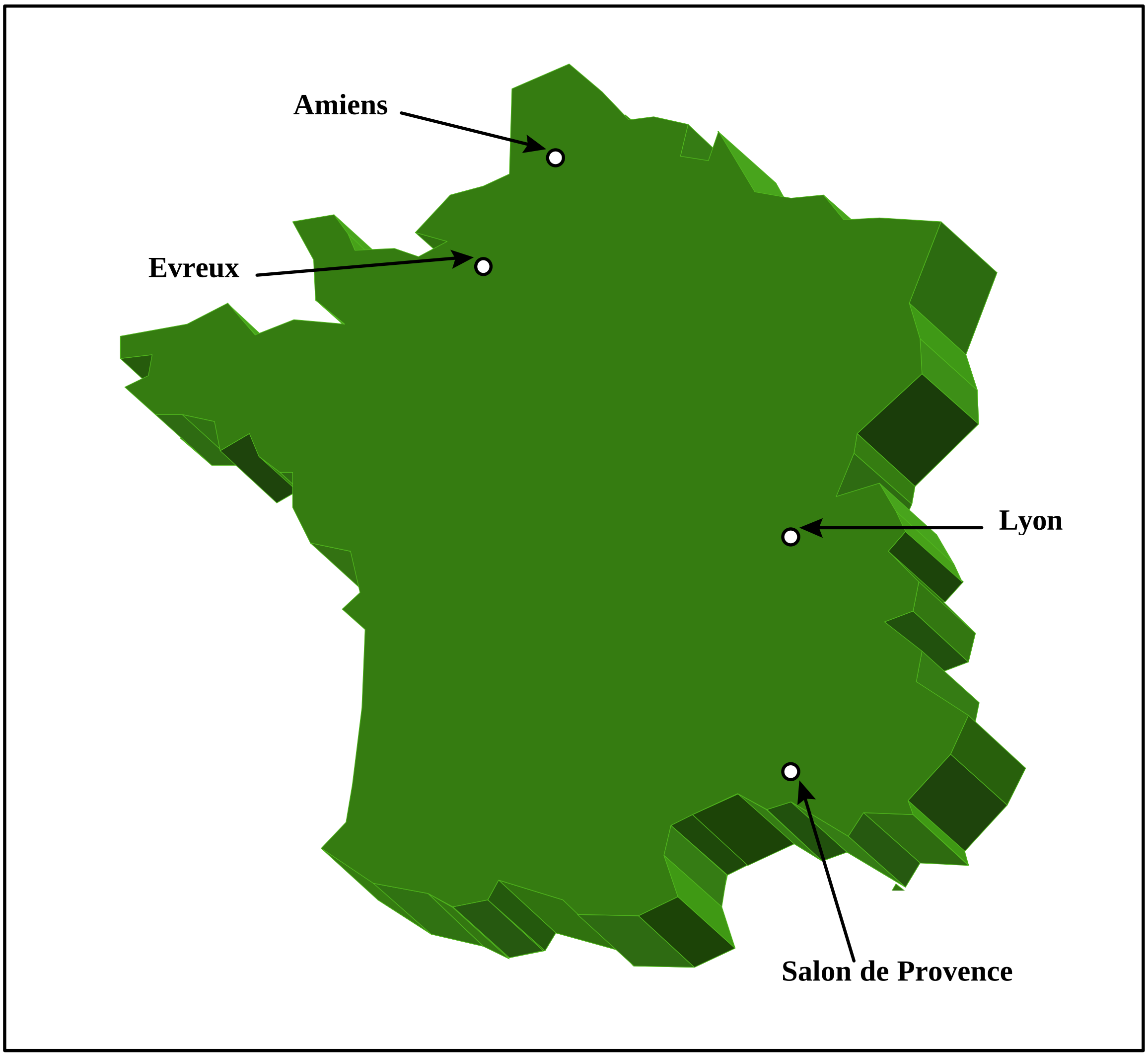
Map 2 : UK Collection Centres.



The study is funded jointly by the Department of Transport and six motor manufacturers: Ford, Honda, Nissan, Rover, Daewoo, Volvo and Toyota.

France has had a variety of accident data collection systems in place ever since the 1960's. However, the current large scale project only began in 1993 and consists of four teams: Two belong to the Institut National de Recherche sur les Transports et la Sécurité (INRETS), at Salon-de-Provence and Lyon. The other two belong to the Centre Européen de Sécurité et d'Analyse des Risques (CEESAR), the joint laboratory belonging to the two French automobile manufacturers, PSA Peugeot, Citroën and RENAULT that carry out the data collection and some analyses on behalf of the LAB[1]. All the teams involved collect their data in the same way and co-ordinate the results. Each team collects 50 to 60 crashes per year, however there is no information on the number of vehicles, individuals or injuries investigated.

Map 3 : French collection team centres.



This project is funded jointly by the French Government and two motor manufacturers: PSA Peugeot-Citroën and RENAULT.

There is a second in-depth research study in France. However, it is primarily an injury register and is not comparable to the other studies. A description of this register is given in Appendix 1.

[1] LAB: Laboratoire d'accidentologie et de biomécanique.

After comparing the study objectives of each country, the headings will then be split into separate areas of discussion and comparison relating partly to the work package one sub-sets. The technical specifics of each countries study will be found in the in-depth reports they have produced, details of which can be found in the references section. All mention in this paper to the German collection system relates to that used by the Medical University of Hanover, France to that of INRETS and CEESAR and the UK to CCIS.

4.1.4 Objectives of each study

The main objectives of each study defines the framework they are required to work within. The Medical University of Hanover (MUH) study is used as a legislative tool, INRETS are concerned with accident prevention, and CCIS, injury causation. All the three studies however, do have similar objectives in the areas of :

4.1.4.1 Legislation

- Producing a tool which can answer questions posed by their respective Governments; evaluating both the need for new legislation or the effectiveness of it.

4.1.4.2 Research

- Formulate a database of the primary safety aspects of crashes (except CCIS)
- Formulate a database of the secondary safety aspects of crashes
- Research injury problems
- Research vehicle performance
- Research the frequency of different types of accident

There is a final area within this topic. It relates to INRETS and CCIS and their collaborative efforts with private car companies. Indirectly the companies in Germany use Hanover's information, but it is bought in and they have no direct influence on what data is collected.

4.1.4.3 Manufacturers

- How do their vehicles perform in real-life crashes?
- Identify good/bad systems
- Requirements for improvements/new systems.

4.1.5 Case selection and definition of local accident population

The sampling systems employed by each partner differ. The MUH system uses a random procedure based on notifications received by the local emergency services and includes all categories of road user. The INRETS systems use a similar procedure, but filters out any multi-vehicle crashes that are too complex to investigate. The CCIS system is stratified and selects just car crashes.

The only criteria found across all three projects is that of selecting crashes that involve at least one injured person, although people and crashes not involving injury are sometimes investigated if a random sampling system is employed.

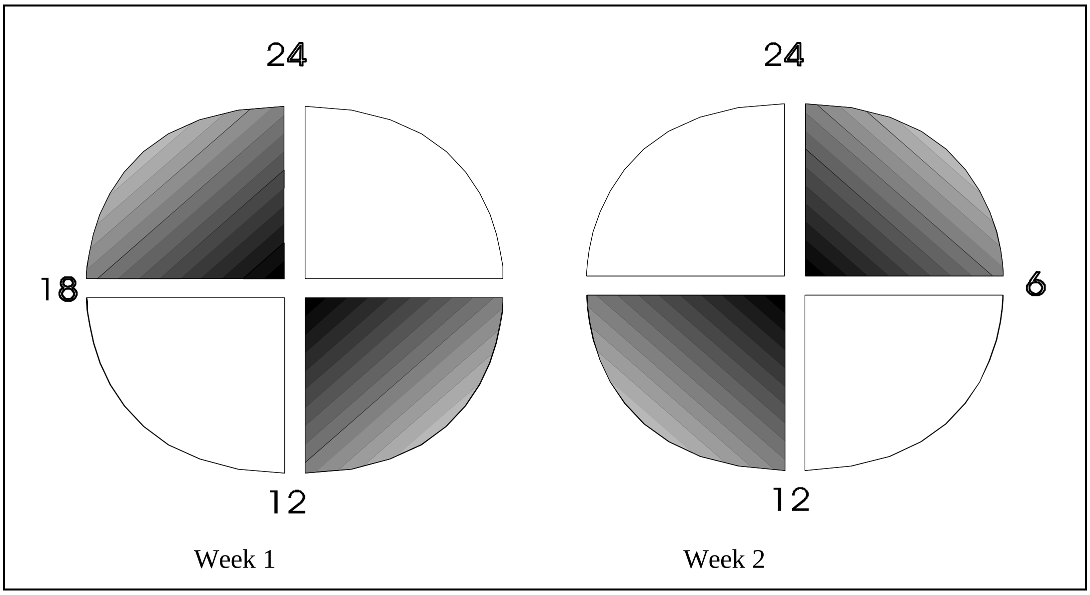
Selection of the cases to be investigated is dependant upon the type of notification used by each team.

The CCIS study allows for retrospective investigation. Notifications are received via Police traffic reports which are then sorted via a strict sampling criteria, allowing tight control over the ratio of killed, serious and slight injury cases selected. The criteria for assessing the injury severity in each case is the same. This enables greater accuracy in the later weighting process, as any cases that turn out to be non-injury are subsequently re-sampled so that adjustments for reality in respect of the local population can be made.

Further filtering is imposed by selecting only car or car derivatives which are under seven years old and were towed away (this allows only new designs to be studied in order to reflect current legislation and design practise and avoids lower energy level crashes). The obvious drawback to this system is only car related incidents are investigated; no pedestrians, motorcyclists or heavy goods vehicles are included. Problems can also be caused by the removal of the vehicle out of the sampling area or by an owner or recovery yard who will not allow access, but this is rare.

The INRETS and MUH systems are notified via the emergency services, either the Fire brigade or Police. Samples are random and attended by the teams in a fixed order based on time of notification. In MUH all times of day are covered by a series of shifts over a two week period. This ensures no bias is present in the procedure.

Fig 1: Medical University of Hanover shift system.



Teams work in two, six hour shifts which are changed weekly

The INRETS system relies upon one shift per day covering a six hour period and splits these periods over four weeks. There is no cover at weekends and public holidays and all primary aspects of the incidents during this time have to be derived from Police records. Weighting is done by comparing the cases seen to the total number of incidents reported to the Police.

Fig 2: INRETS (Lyon) shift system.

WEEK	1	2	3	4	5	6	7	8
DAY	MON	TUES	WEDS	THURS	FRI	SAT	SUN	MON
TYPE OF INVESTIGATION	ON SCENE	ON SCENE	ON SCENE	ON SCENE	ON SCENE	RETRO	RETRO	ON SCENE
HOUR	12 6 AM	12 6 PM	12 6 AM	12 6 PM	12 6 AM	12 6 PM	12 6 AM	12 6 PM

The sequence continues logically

The drawback of these type of systems is that at peak times not all crashes can be investigated, and the team may be called out to a non-injury case that cannot be used in the database, although this is rare. Also, the system misses people who make their own way to hospital and so cannot include every low injury crash. Furthermore, definitions of the severity of the accident between the MUH and INRETS are not consistent with each other with the MUH criteria based around the casualty being ambulant or stationary and INRETS on admission to hospital for more than six days or requiring medical treatment. The strength of these systems is in selecting a random sample, but they do require a more complex method of weighting. It is also expensive and manpower intensive and relies heavily on the co-operation of the emergency services, especially at the scene. However, it does have the advantage of having all the primary safety information available to the investigators.

4.1.5.1 Conclusion

All three groups operate a close degree of control over the case selection procedure which offers a good basis for future comparability. Each group also has to use weighting procedures so that the samples become representative of the local population. In general it can be said that the MUH study is representative of all injured people involved in traffic incidents and taken by emergency vehicle to hospital in the Hanover collection area. The CCIS study is representative of all car involved incidents within their area and the INRETS representative of a selective band of incidents across all crash types. There is no indication that the initial populations are comparable.

The random sampling is a consequence of the need to get to the scene of the crash. It is much more difficult to operate a selection procedure in these circumstances due to the fact that all the details of the incident are not known until arrival at the scene.

A stratified sample is more efficient statistically, but cannot be operated realistically for an on-scene study.

Finally, in their unweighted form the samples are not directly comparable. After weighting to the local population they may become comparable, and a comparison of the distribution of standard variables will clarify this. All three groups select from a sub-set of the national population and work package two will develop a process to handle such a crash sample.

4.1.6 Accident Investigation procedures and limitations

There are two methods used within the consortium: retrospective crash investigation carried out by CCIS, and on-scene, in-time investigation used by INRETS and MUH.

4.1.6.1 a. Retrospective procedures

The accident investigation procedure covers two stages, vehicular and injuries, each having their respective specialists:

1.- Vehicular -

On receipt of the relevant Police notifications a stratified sampling procedure is put into effect. The details of the crash are logged and then given to a team of two Accident Investigators who may be given a number of these to complete throughout that day. This will depend upon the recovery yards the vehicles have been towed to, how close they are to each other, and the complexity of the crash itself. On average the inspection takes approximately one to two hours per vehicle.

2.- Injuries -

Note is taken from the Police information as to which hospital the injured occupant has been taken to. A request is then made to the hospital for access to the individuals file. At the same time, a questionnaire is sent out to all occupants of the vehicle (unless it is a sensitive case). If the crash is fatal, a request is made to the appropriate coroner for a post-mortem. At the hospital all details of the injuries are recorded to be converted at a later date to the Abbreviated Injury Scale 1990[2] (AIS). Maximum Abbreviated Injury Scale[3] (MAIS) and Bakers Injury Severity Score[4] (ISS) are also

[2] Abbreviated Injury Scale, 1990 revision, Association for the Advancement of Automotive Medicine (AAAM), Des Plaines, Illinois, USA.

[3] Maximum Abbreviated Injury Scale - The highest AIS of all body regions.

[4] Bakers Injury Severity Score - The sum of the squares of the highest AIS code in each of the three most severely injured ISS body regions (Head or neck, Face, Chest, Abdominal or pelvic contents, Extremities or pelvic girdle. External).

calculated. The questionnaire is used to back-up or clarify data collected elsewhere and is not used for coding, however it does give other details such as non-hospital attendees, minor injuries not observed at the hospital and use of child restraints.

There are various limitations of this system. One is lack of on-scene information as this restricts the options for collision analysis. Also, the reduced information about the movement of the vehicle within the crash may restrict assessment of some of the injury causes in a complex collision sequence.

The advantages in using retrospective investigation is control of the sample for scientific accuracy and efficient use of manpower and time.

4.1.6.2 b. On-scene, in-time procedures

INRETS and the MUH systems of data collection involve going to the scene of the crash immediately after the event to take the relevant information required, although both do follow-up with a more detailed investigation at a later date. Each is notified of an incident by either the Fire Department or Police. The INRETS team waits for this notification at the relevant station and attend the scene in the emergency services vehicle. As Hanover have worked for so long with the Police they have their complete co-operation, and are able to attend the incident using a blue light and preferential driving techniques to get through the traffic, normally reserved for the emergency services. This is justified since each team has a medical student trained in emergency first-aid procedures. Both systems have people within the teams whose job it is to record: the environment, the vehicle and the injuries. INRETS have the additional aspect of a psychologist. Each area of investigation has been split up and are:

1. The Environment and Vehicle -

   At the scene of the crash all the necessary details are taken in order for a reconstruction to be made. The teams from MUH use a stereo measuring camera which allows true-to-scale drawings to be made, whilst INRETS rely upon photographs and video to record the scene. Environmental details, road structure and traffic controls are also noted. Both systems use a camera to record the vehicle damage and other features of interest, as well as written records. The INRETS psychologist interviews the people involved at the scene if possible (unless it is a sensitive case), with a follow-up interview at a later date.
2. The Injuries -

   Both systems look in detail at the medical data to correlate contact points in the vehicle with injuries sustained by the occupant. In the MUH system the medical surgeon integrated into the team arrives at the crash scene in a separate vehicle. They immediately begin to document the care the patient receives throughout their initial treatment and relays the injury details back to the investigators. This allows the investigators to focus on the areas suspected of causing injury and find contact points that otherwise may be missed. Also within the MUH system, everyone involved in the crash is sent a questionnaire so that information on the low level injuries and other information can be correlated. The return rate is about 40%. In the INRETS system, it is the psychologist attending the scene who collects the low level injury information.
3. Behaviour –

   INRETS collect one further area of information, that of the psychological aspect of the incident. A psychologist interviews all participants in the crash and some witnesses to get an immediate idea of what they thought of the incident and how it happened. This initial interview is followed up at a later date for a more in-depth discussion. Care is taken in cases that may be of a sensitive nature, including interviewing children. This information is of obvious benefit as to the thought processes involved in the lead-up to the crash.

Both on-site/in-time systems allow for easier reconstruction of the incident and help in the estimation of the speeds and energy involved. However, sample selection is difficult and at peak times incidents are missed which would otherwise be included.

This type of system is labour and cost intensive, with much of the time spent waiting for notifications of an incident, but does allow the collection of important information which is only available at the scene.

4.1.6.3 Conclusion

The CCIS study focuses on secondary safety only, whilst INRETS and MUH have some degree of primary safety aspects integrated into their systems. This means that they study cases on-scene and in-time, going back to further investigate if the particular crash is too complex for all details to be taken at the time of the incident. The CCIS version is retrospective and vehicles are inspected away from the incident and usually at a later date. Both methodologies have their own strengths and weaknesses in terms of time, manpower, costs and the tasks demanded by each project.

All three studies are notified of incidents by the emergency services in one form or another. Each also conducts post-crash investigations using trained vehicle inspectors. Photographs and drawings are also taken. In each system various determinations of energy and speed are used. The means of classifying the injury is consistent across all three studies; use of the Abbreviated Injury Scale 1990 (although other supplementary scales are also used). When all the information is known, assessments are made as to the points of contact that caused the injury. Thus, there are many areas of comparability, and with modifications there could be large areas of harmonisation and common practises.

4.1.7 Data collected

This particular section can again be broken down into further sections: Environmental, vehicular, and individual.

4.1.7.1 Environmental

Each study either collects its own information on the road environment around the crash location or links to other databases. With the MUH and the INRETS investigators being on-site, measures are taken at that time, while the CCIS study has links to the data recorded by the Police in attendance which are held on their national database. Varying levels of information are taken, from the very detailed INRETS measures entailing road geometrics, the MUH relatively detailed measurements, to the lowest level of the CCIS which the other two encompass and consists of the fundamental weather conditions, lighting, type of road and so forth.

4.1.7.2 Vehicular

The data gathered in this section by the CCIS study centres on: the structure of the vehicle, description of the damage, performance of safety related features and contact details. It does not include bicycles or motorcycles its data. The INRETS and MUH systems however, deal with all vehicles and pedestrians and the only area of consensus between all partners is that of investigating car or car derivatives.

All partners take general details of the vehicle; it's type, colour, size, age and so forth. They also note other, more in-depth features. Did the vehicle have power-assisted steering or ABS? Was there a fuel leak or a fire involved?

In all cases the profile, maximum depth, length and locality of crush on the car is measured but the method of description differs for all. The MUH system uses a plan view grid whereas the remaining two use a series of recorded measurements of all planes of the vehicle, however the points of reference for these measures are not the same in all instances. Similar methods of measuring intrusion from defined points within the vehicle are used by all partners, but again the position of these points are not exactly the same.

Other aspects of the interior safety systems of the vehicle are also covered by all partners: the state of the windscreen, sunroof, steering wheel, seat belts, air-bags, child restraints and the position and condition of the seating.

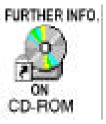
4.1.7.3 Individual

The general details of the individual are collected within each system: age, sex, weight and height. Details of the injuries vary within each system with the access to the casualty throughout the treatment period giving a high degree of detail to the information recorded in the Hanover study. INRETS and the CCIS method of relying upon hospital records leads to a similar level of detail between the studies. The CCIS study has an added bonus of access to the routine post mortem reports of fatalities, so full injury details are available. In France and Germany post mortems are not carried out routinely, but injury details may be available if the casualty died in hospital after a period of treatment. Reports on any post mortems may be available with the consent of relatives.

All partners use the injury descriptor system AIS 90. Each has enhanced the scale with additional descriptors and these vary between each group.

4.1.7.4 Conclusion

Much of the data here is at least described in the same manner, even if the collection and recording methods differ. There is a high degree of comparability at the general level, with similar core information types, but this may break down at a more detailed level and is yet to be resolved. There are some areas not covered by all three systems these being; the MUH rescue data whereby the emergency services methods are examined and recorded, and the INRETS psychology data. Based on the situation in each country access to post mortems information also differs between each partner: in the UK it is 100%, Germany 33% and France none at all.



4.1.8 Crash severity estimation methods

The type of investigation procedures again dictate the method used in this area. With on-site investigation, a backward reconstruction of the incident can be done. By using not only the damage to the vehicle but also other available information such as skid marks on the road, markings on street furniture and final vehicle resting positions, a decision on the series of events within the crash can be estimated. The damage algorithms and detailed measurements of all traces and the final positions of the vehicles can be used to calculate Delta-V (mathematically calculated alteration of speed) and Equivalent Energy Speed (EES - comparison of energy speed reduction in similar crash configurations).

Within the current area of vehicle safety research there is discussion surrounding crash severity and the most appropriate methodology to use in its calculation. Many of the problems associated with these conflicts rest upon the misunderstandings surrounding them. In crash investigation there is currently no method of always knowing the pre- and post- velocities of the vehicles involved at the actual time of the crash. However, there are methods of calculating the change in velocity using two basic methodologies. One we will call *work energy analysis* and the other *deformation energy analysis*.

The *work energy analysis* method utilises trajectory data and is based on work-energy relationships and the principal of conservation of linear momentum. Most of this type of information is ‘volatile’ and has to be collected at the scene (skid marks on the road surface and debris that can indicate the point of contact) so that the vehicle direction and point of impact can be measured as accurately as possible.

The other method mainly used is *deformation energy analysis.* For this type of calculation the principles are based on the principles of the conservation of energy, the conservation of momentum and the vehicle damage.

Both methods use a set of assumptions within their calculations based on vehicle crash testing under laboratory conditions (stiffness) or field testing (road friction). W*ork energy analysis* methodologies use the work energy dissipation due to friction (using the length of residual skid marks), pre- and post-crash trajectories and the distance travelled to rest to make their Delta-V calculations, but can also use the deformation of the vehicles as an adjunct to this. The *deformation energy analysis* method uses vehicle deformations only (i.e. the deformations from all vehicles involved in the collision).

Another problem encountered was that the method of recording the crush damage of a vehicle was not consistent throughout each study, although the differences are minor. INRETS and CCIS use the Collision Deformation Classification (CDC[5]) along with other descriptors, and are very similar. The MUH model is of an earlier type, the Vehicle Deformation Index and is closely comparable to the others.

4.1.8.1 Conclusion

There is no consistency throughout all three studies as to the use of a particular piece of software to determine crash severity. Crash 3 and PC Crash are the most common software in use. PC Crash is used more for reconstruction modelling but both programmes use the same algorithms in their calculations. All groups use Delta-V and EES and are fully comparable with each other. Crush damage is measured in a similar manner throughout and would need very little modification to become comparable.

4.1.9 Data quality and error checking

Quality of information collection is maintained throughout all three studies by use of highly trained staff. Definition of variables and measuring points to be used are taught throughout in order to maintain consistency. There is also the manual cross-check that occurs when data is collected on a single incident by more than one person. Each area can highlight the plausibility of the other.

[5] SAE Recommended Practice J244 MAR 80

All teams are able to further check their data by use of photographs, and in the case of the INRETS system, video. Pictorial evidence of the vehicle or the scene can be used at a later date to check the validity of a piece of information that may be in doubt.

Further checks are done by all partners when the case is compiled. A manual check is made on the logic of the information recorded throughout the case in all three systems. Periodic reviews are also done in order to get consensus on certain types of cases and to retain the quality of the information collected. A final manual check is carried out by the CCIS team, whereby a person totally unconnected to the incident checks the whole of the case for consistency, whereas the INRETS psychologist does a synthesis of each incident.

All systems check their data on logical input, input error and plausibility.

4.1.9.1 Conclusion

All partners make efforts to retain data quality and are in the main comparable to each other. All use more than one person to collect the data on the vehicle which allows for a cross check to be made manually. A final check via computer is also made by all concerned, but in various ways and to varying degrees. There are other checks in place in each system, but they are not fully comparable with each other. However, the specification of a core data set can lead to the identification of a common set of data checking procedures that can be implemented by all.

4.1.10 Use of data

The data collected by each member is used in three areas: research, legislation and by the manufacturers.

4.1.10.1 Research

All partners use their data to investigate and analyse various areas of their study. However, within these confines, areas of study from the frequency of certain kinds of accidents to the performance of the safety systems within vehicles take place. Trends of injuries and injury causation can be followed and the impact of new vehicle design or new safety systems can be measured. Areas of further research can also be identified which may point to necessary developments of other systems to offer higher safety standards.

4.1.10.2 Legislation

The Governments of each respective country utilise the data collected to answer any questions that may arise within the field of crash injuries and their causation. It is also used to identify or formulate new laws and evaluate their effectiveness as well as point towards any further regulation or legislation efforts that may be necessary.

4.1.10.3 The manufacturers

Information collected in real-life crashes is of value to the manufacturers in that it can show the strengths and weaknesses of their designs, identifying at an early stage any defects that may occur in their vehicles or the safety devices within them. The projects can then be evaluated over the short, medium and long-term. This allows in-lifetime changes to the designs and development of safety strategies around successful projects. In this way the data is used to enhance and focus their own laboratory crash study research and provide them with information that would otherwise be out of their reach.

4.1.10.4 Conclusions

The owner of the data is different in each system. This determines the access and use of it. This area is one that is constantly changing, and full comparability will never be achieved. However, the general thrust of each system is the same.

4.1.11 National/local ethical considerations

This section has the least comparable information in it due to the major differences within the structure of the governing bodies concerned.

MUH has the least amount of trouble in obtaining some of its information due to the fact that access to the files has been authorised at a high level, both for the Police and the medical data. INRETS and CCIS have had to build up a strong local relationship with the necessary officials in order to obtain their information.

Due to both the INRETS and MUH studies being on-site, and the personal nature of some of the information recorded, agreement of the individuals concerned is required before any of data can be used. At MUH this means obtaining a signed release at the scene, or trying to obtain it at a later date. In INRETS the psychologist has the responsibility of obtaining the persons permission. Until this permission is given, the data may not be used.

The access and use of medical data obtained from a post-mortem seems to be an area of digression. As it stands, only the CCIS has access to post-mortems on a regular basis. The MUH study may see the details of a case that is judged to be of forensic interest, but only with the agreement of the relatives, and the INRETS have some access if a post mortem has actually been done. Considering this provides the most useful data on the most serious of crashes, improvement in this area would be beneficial for this project as a whole.

4.1.11.1 Conclusion

All groups collect the data according to strict ethical principles. These principles concern the preservation of anonymity, computer data anonymity, and constraints of medical confidentiality and where necessary seek the individuals permission for its use. Discrepancies in this area occur due to the differing political nature of each member country. Conducting post mortems and gaining access to the valuable information they contain is an area that can only be solved by a shift in policy of the governments concerned.

4.1.12 Concluding remarks

There are many areas of comparability throughout all three studies but only in the area of secondary safety. At a conceptual level there are many areas of similarity, but at the detailed level there are many differences.

Agreement on the objectives of collecting the data should point to which aspects of the crash are to be used and at what level. Consensus will then be required as to the methodology of investigation. The common focus concerns car occupant injuries and there are many areas of comparability. The main area of lower comparability concern case selection procedures and the inherent differences previously discussed. The CCIS does result in a more cost effective and statistically efficient model; whereas the random sampling procedures of INRETS and MUH allows for primary safety studies. As all partners do post-crash investigations on at least some of their cases, this would seem to be the best method to employ. Some on-site, post-crash recordings may be integrated into the system to support an expansion into accident causation data collection.

Next, agreement as to the mathematical model or software used for calculating crash severity would define at least some of the variables that are required for collection (unless the programs are non-contradictory). As STAIRS is fundamentally a secondary safety study, analysis of the performance of the vehicle and its safety mechanisms are obviously important factors. These two points go some way to providing a lead as to the type of methodology to be employed collecting this essential data. It was, then, not too difficult to focus on key measurements required by all three partners. The problems met were the precise understanding of terms used. This identified a further task dealing solely with crash severity terms and their meanings, and is given on the CD Rom.

Differences in methods are likely to remain as they are a consequence of local practicalities as well as technical precision, but these can be overcome. Some movement from all partners will have to occur if the project is to be successful. It will be the drive and professionalism of each team that will allow this to happen, as well as this ability to be flexible. The aims and objectives are achievable, if at a cost to each partner, and a European-wide database will be beneficial to all.

FURTHER INFO. ON CD-ROM

4.1.13 Another approach to in-depth investigations : Road Trauma Register

In France, the development of a new approach has been the result of two observations:
The first concerns the accident which is studied in far more detail than the victim (the latter, however, justifying the major part of the interest which one shows in the former). In other words, the 'health' dimension of the problem is overshadowed. As this is not studied elsewhere, one ignores an important facet; the statistical distribution of the different injuries which these casualties may have.

The second part is on a disciplinary level. The accident research field is divided into two main approaches: case study and statistical analysis of the crashes. There is however, room for investigations with a statistical aim concerning the medical aspects of a crash. The epidemiological tool is made for this purpose.

It is in this context that INRETS has been lead to propose and set up a Register of road accident casualties in the Département du Rhône.

4.1.13.1 Register of road accident casualties

Grouping public, private, civilian and military structures together, and aiming to be a real register of morbidity, this original network offers, for research, indisputable and recognised advantages. The Rhône department is interesting, in that it has within its boundaries a large urban metropolis and a rural area, along with a sufficiently diversified departmental road network.

The register rests on the participation of the different intervening services within the area of health care of accident casualties. As illustrated by the map of the Rhône below, this network currently encompasses the services which, although outside of the department, are likely to receive eligible casualties.

4.2 Work Package 1.ii – Variables and Values

4.2.1 Table of Contents

4.2.2 Specification of Variables

The objective of this task was to define a set of variables along with detailed data fields and a glossary of terms that will form the basis of a harmonised crash injury database. In order to achieve this goal, it was decided to determine a nucleus of variables, which were seen as essential for the realisation of a secondary safety orientated, data collection system. The group tried to specify the different values required for these variables, but of course, one must almost systematically add values such as “unknown” or “inapplicable”. Although only a small selection of road users are used as examples, the basics principles are transferable with little or no modification. This document is seen as a starting point and it will evolve over time.

The work has tried to capture within its remit other major contributions to safety, such as the CREST program on child restraints, the COST 327 project on research into helmets, and Euro NCAP. The requirements for these studies have been assessed and variables included that are compatible with them.

As much as possible the variable list has tried to use commonly accepted systems for data collection, such as AIS90, but it has been found necessary to expand on this in order to obtain some of the in-depth details required. Wherever possible, backward compatibility has been retained.

In some cases a new system had to be devised due to the incompatibility of those already in place. Wherever possible these systems have been kept very simple with the hope that they will be easily understood and still collect the necessary details. They will also allow for the ‘mapping’ of other systems onto a common framework.

The document as a whole is designed in a modular fashion, in order to allow the relevant sections to be extracted for whatever aim or objectives a study may have. It is expected that other variables will be required to be added. The list within this document is seen as the minimum requirement and has limitations.

The work package includes the following sections:

- A general data module,
- A vehicle data module (for each type of vehicle),

- Pre- and post-crash data,
- Seating data,
- Restraints (a description of the action and deployment of safety systems),
- Child restraints (type, position and performance),
- Intrusion (description of the interior crush),
- Description of the casualty (all types of road users),
- A pedestrian data module,
- Description of the injuries.

Each section sets out in detail the information that is required to be collected, and if necessary how it is to be collected, along with explanatory notes.

4.2.3 Variables, values and notes (Full version on CD ROM)

Due to the length and complexity of this section a full version has been included on the CD Rom and only the introductory pages to the sections are given here. The same headings are used on the CD version, but the variables are greatly expanded upon to include values, diagrams, explanations and examples.

Although STAIRS was not envisaged as a coding system, it was felt necessary to include codes for certain variables by way of identifying their meaning in a more comprehensive manner. For some (either completely new systems or variables devised specifically for STAIRS), they were required for completeness. Further explanations or development of the codes are given in the relevant notes section.

Note was taken of the views expressed by parties outside of the STAIRS consortium as to the modular structure of the protocol, and it has been adjusted accordingly. Initially, the emphasis is upon a global view of the crash and the elements contained within it. Later, more and more details are required of these elements until a complete picture is built up of the incident. There has been some duplication of the variables necessary for each section due to this modular format being adopted.

4.2.4 General data module

A brief description of the crash, vehicle(s) and casualties involved are given to allow an overview of the incident.

There will be some common variables within this section that relate to the National Accident Databases of each country involved in order to allow for possible sample weighting.

The vehicle categories are broken down into separate sub-groups; however the collision partners are divided into International Classifications. Each vehicle also has an International Classification in order to allow for a more global view if analysis is to be broadened to take into account the International situation.

Variables:

Crash data	Total number of articulated trucks	Total number of persons injured
Date of crash	Total number of bus/coaches	Total number of persons killed
Time of crash	Total number of agricultural vehicles	Total number of pedestrians
Number of vehicles involved	Total number of trams	Total number of two-wheeled users
Total number of pedal cycles	Total number of railway vehicles	Total number of occupants type
Total number of motorcyclists	Total number of other vehicles	Collision partner
Total number of cars	Numbers of people involved	Main collision
Total number of vans	Total number of persons not injured	Secondary collisions and following
Total number of small trucks		
Total number of trucks		
Total number of rigid trucks		

4.2.5 Vehicle data module

This section goes into more detail concerning the vehicle's identification: size, weight, body style etc. It can be applied to all types of road vehicles involved in crashes including trucks, cars and motorcycles. More specific variables are described within this section so that a higher degree of detailed analysis can be made and some of the more subtle differences between the vehicles explored. This level of detail also allows certain groups of vehicles to be identified that may be involved in a higher proportion of crashes than should be expected and may highlight shortfalls in licence grouping, training or safety related features. It can be used to determine the interaction between vehicles and vulnerable road users, such as pedestrians and cyclist. Other issues such as compatibility can also be investigated.

Various attributes of the vehicle that may cause injury are registered, such as the engine orientation, the kind of steering column involved, fitting of bull bars, etc. The obvious benefits of collecting this information is that it can help to identify a source of injury to occupants as well as to the vulnerable road user. The identification and placement of each element can, in certain circumstances, be attributed to the injuries caused and strategies designed in order that the offending item can be removed or re-designed to reduce the injury. This is an important consideration with injuries that cause long-term disabilities or impairment and the cost to society as a whole. It will also allow evaluation of the new safety measures or products that may be introduced as a consequence.

The glazing of the vehicle is of interest, especially in cases of ejection, in that occupant kinematics can be discerned for the various crash configurations via contact points and a deeper understanding gained of the movements of occupants in real world crashes. This can help in dummy design, especially the evaluation of side impact movement and any interaction with an air bag.

Variables :

Identification :
Make
Model
Vehicle Identification Number
Age of vehicle
Class of vehicle
Odometer
Body style :
Motorcycle
Three-wheeled vehicle
Car
Van
Truck
Wheel configuration
Trailer
Driver position
Dimensions :
Length
Width
Height
Number of doors
Bumper height
Bumper width
Weight :
Kerb weight
Weight at time of crash
Permitted gross weight
Load weight
Suspicion of overloading
Technical features :
Capacity
Power
Turbo
Speed limiter
Location
Orientation
Fuel type
Gearbox type
Gearbox position
Braking system
Steering column type
Steering wheel
Steering wheel hub
Driven axles
Towing
Permitted maximum towing weight
Towing weight at time of crash
Carrying - Inside
Carrying - Outside
Bull bars
Winch
Load barrier type
Folding rear seat
Load barrier position
Glazing :
Type
Retention
Damage
Position
Cause of breakage

4.2.6 Pre-crash and seating data module - Car

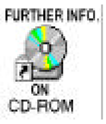
Within this section are details concerning the seating position and attitude of an occupant. There are also descriptions of the seat itself and the different design attributes associated with them.

Internal pre-crash measures are collected to describe the seating attitude of the vehicle occupants; the distance to the steering wheel, head restraint, etc. This can be related to the normal design position of the seat belt and airbags.

Another reason for collecting this information is to assess the available space for the occupant within the vehicle. From this, occupant kinematics can be derived and associated with the actual injury outcome.

The measures described are not necessarily the ones that can be taken in the field. Factors such as the position of the vehicle or intrusion into the area to be examined may make it impossible to take them. In cases such as this, the measures can be derived from others directly related to them (i.e. the distance from the roof of a vehicle to centre of the steering wheel, can be used to deduce the distance from the centre of the steering wheel to the floor, by using standard measures taken from an undeformed vehicle of the same make and variant), and it is seen as the investigators job to decide, using their expertise, which is the most appropriate to take. This may change due to the circumstances of each crash configuration. However if the measure described can be calculated, by whatever means, and its accuracy is within the acceptable limits of the study, then it can be used for analysis.

The positioning of occupied seats follows the International Standard (Preliminary FDIS version ISO 13218) and can be used for any number of seats in a vehicle, including coaches and buses.
Head restraints details are collated under this section, rather than under the 'restraint' section, as they are integral to the design of most seats and function as part of the whole. New developments within head restraint and seat design and investigation into there effectiveness, especially concerning the new rear impact orientated seats, would fall within this area for the same reason.



Variables :.

Horizontal distance from the top of the seat back to the D-point
Horizontal distance from the top of the head restraint to the D-point
Vertical distance from the centre of the steering wheel to the floorpan.
Minimum horizontal distance between the seat backs.
Horizontal distance from the middle of the centre of steering wheel to seat back
Minimum horizontal distance between the seat bases.
Horizontal distance between the foremost part of the seat to the mid-point of the 'B' pillar
Angle between the seat base and the seat back
Seating Data (Car) Module
Seats (coded for each seat position) :
Seat position
Seat occupied
Seat direction
Seat back hinge
Seat back loaded
Folded before crash
Type (for rear seats only)
Head restraint type
Head restraint position
Head restraint damage

4.2.7 Post-crash Data Module - Car

A more detailed view of the accident configuration is given within this section, and can further help to refine any analysis that may be required.

The necessity for both sets of measures are to evaluate the performance of the vehicle in a crash. It is the difference between the pre and post crash measures that define the effectiveness of the vehicle in a collision, but this does have to be balanced against the type of injuries, if any, sustained. A stiffer vehicle could perform better in a crash than its counterpart, but may produce more injuries due to the higher deceleration rate of the occupants. These measures can also highlight other factors within the area of compatibility.

The post-crash measures also define the way the vehicle and the relevant passive safety systems (e.g. seatbelts and airbags) have behaved in the crash, and give some information as to its crashworthiness. They can be related to the level and kind of injuries that may be expected to be seen in a certain type of crash. Over a period of time this can also be an indicator as to the effectiveness of any new legislation or safety systems.

The amount of overlap, mass of the vehicles and energy contained within the crash is also detailed and can be used to compare real-world data with that obtained in laboratory controlled impacts. This will identify if the results were reflective of the legislation or directives that developed from the tests. If necessary, they can then be further refined and monitored by using this method.

The Collision Deformation Classification developed by the Society of Automotive Engineers as defined in SAE J224b has been enhanced to allow a better understanding of which structures have been involved in the crash. A higher level of detail is obtained concerning the influence of a vehicle's longitudinal stiff structure and any road wheel involvement. This can affect the dynamic load path of the crash forces involved and influence the amount of intrusion within the vehicle along with the injury outcome to the occupants. At another level, this information is helpful in identifying if certain software can be run to calculate crash severity.

The area of crash severity is examined, and measures suggested for defining this complex subject. It is due to this complexity that a separate section has been written on it, describing definitions and terminology along the practical difficulties in using different methodologies.

The detailed measures for two-wheelers is set within a separate section to allow for a modular approach.

Variables :

Collision condition :
Number of collisions
Chronological collision number
Collision number (by severity)
Collision partner
Mass of collision partner
Overlap
Collision angle
Direction of Force
Rollover
Under-run
Final position
Fire
Water submersion
EES
Delta V
Collision Deformation Measurements :
Direction of Principal Force
Deformation location
Specific deformation area
Vertical deformation
General type of damage
Deformation extent
Bonnet
Edges raised
Contacted screen
Penetrated plane of screen
Longitudinals
Direct loading
Crumple
Multi-bending
Rearward movement
Front projects
Engine
Direct loading
Rearward movement
Contacted/displaced bulkhead
Mount failure
Other movement
Side impact
Height of minimum impact
Sill loaded
Sill buckled
Front wheel strutted
Front pillar moved back
Others
Loss of stability of 'A' Pillar
Loss of stability to front header rail
Loss of integrity or damage of the dashboard/firewall to the 'A' Pillar
Floor buckling
Location - Front left
Location - Front right
Location - Rear left
Location - Rear right
Footwell ruptured
Footwell movement
Doors (for each door)
Opened in
Opened out
Jammed
Missing
Door aperture reduction at waist level
Overrode sill
Latch damage
Latch failure
Hinge damage
Hinge failure
Tailgate
Door beams fitted
Height from ground
Damaged
Loaded occupant
Intruded
Jammed door

4.2.8 Restraint Data Module

This section applies to all vehicles fitted with seat belts or airbags. All details relating to vehicle restraints are included within this section. Each is related directly to the relevant seating position concerned. This will provide valuable information not just on the effectiveness of the system, but also on usage rates and other factors relating to legislation.

The list is not exhaustive and with new systems being developed all the time, variables relating to them can be added to take into account the changes.

Any problems relating to the restraint can be noted, either user related (which might be solved by the use of either legislation or education strategies), or manufacturer related (again this can be identified and solved by legislation or even consumer pressure).

Variables :

Restraint details:
Seatbelt used
Seatbelt incorrectly used
Seatbelt overloaded
Seatbelt other damage
Pretensioner
Pretensioner type
Pretensioner location
Load limiter
Load limiter type
Load limiter activated
Webbing lock
Airbag (for each airbag)
Size
Activated
Disarmed
Location
Damaged
Deployment problems

4.2.9 Child Restraint Data Module

This section deals specifically with child restraints, their use and ultimately their effectiveness. The cost to society, as well as the pain and suffering of the families, mean that this area of safety is crucially important.

The variables given here closely follow those of the CREST program (Child restraint system for cars). As the project will not produce it's final findings until after the completion of STAIRS, this section may have to be changed to reflect the final outcome of the group.

The information that can be accessed from this section will enable the practical aspects of using a child seat, from installation and correct usage, to be assessed. It can also be used to validate current and future legislation.

Variables :

Child restraint - car related measures :
Minimum horizontal distance between the seat backs (left side)
Seat back angle of position 1.1
Minimum horizontal distance between the seat backs (right side)
Seat back angle of position 1.3
Child restraint type :
Location
Use
Orientation
Type
Manufacturer
Model
Approval number
CRS fixation
Restraint of child
Inclination
Specific adjustment
Additional features
Misuse - Incorrect use
Misuse - Inappropriate use
Child restraint recovered
Description of damage

4.2.10 Intrusion Data Module

This section describes the intrusion measures to be used for cars. The amount of intrusion that occurs in a vehicle is a good indicator of the level and type of injuries to expect. Due to this relationship, a high degree of importance has been attached to this area and many different systems have evolved to measure it. Due to the complexities and divergent nature of these methods it was felt that a new, simplified system was required. It is based on a matrix that can be further enhanced to allow for a more detailed investigation, and also identifies common contact points so that injuries can be coded to certain areas of the vehicle.

It is hoped that other systems can be mapped onto this method as a collection process has not been stipulated. This will allow for a wider analysis of data from different studies to be undertaken.

Variables :

Measurement of Intrusion :
Position of intrusion
Maximum intrusion
Intrusion measurements at common contact points - Frontal Impact :
Roof
Windscreen frame
Windscreen
Dashboard left - Outboard
Dashboard left - Inboard
Dashboard right - Outboard
Dashboard right - Inboard
Footwell left - Outboard
Footwell left - Inboard
Footwell right - Outboard
Footwell right - Inboard
Central console
Brake pedal moved by intrusion - Up/down
Brake pedal moved by intrusion - Left/right
Brake pedal moved by intrusion - Forwards/backwards
Evidence of braking
Steering wheel movement - Up/down
Steering wheel movement - Left/right
Steering wheel movement - Forwards/rearwards
Steering wheel damage
Intrusion measurements at common contact points - Side Impact :
Roof
Glass
Glass frames
Doorcap
Door
Door fixtures
Sills
Brake pedal moved by intrusion - Up/down
Brake pedal moved by intrusion - Left/right
Brake pedal moved by intrusion - Forwards/backwards
Evidence of braking
Separating device at bracketry
Steering wheel movement - Up/down
Steering wheel movement - Left/right
Steering wheel movement - Forwards/rearwards
Steering wheel damage

4.2.11 Casualty Data (Car) Module

This section deals with the type of casualty involved in the crash. Due to the nature of collecting volatile information of this type, it is more conducive to an on the scene study than a retrospective.

Variables that are known to affect the capabilities of individuals, such as alcohol, are included within the list and are seen as an important indicator of legislation and enforcement efforts.

Aspects that may affect a persons abilities, such as vision aids or other physical disabilities, can be used to monitor the exposure of these groups to collisions and injury. Identifying if certain problems are occurring within a statistical sub-group can prompt remedial action that may require new or revised measures. When these are put into affect, they can then be monitored using the same, or revised, variables.

Variables :

Type	Weight	Alcohol
Behaviour pre-crash	Gender	Clothing thickness
Age	Vision aid	Car occupant ejected
Height	Physically disabled	Ejection route

4.2.12 Pre and Post-crash Data (Two-wheeler) Module

As with the module for the car, this information will give an insight into how the vehicle has performed in the collision. It can also be used to identify injury causing mechanisms both to the rider and other vulnerable road users.

The seat, handlebars and other measurements will allow for a reconstruction of the crash and help to determine the stance or seating attitude of the rider. This is especially important when the kinematics are being investigated.

Variables :

Pre-crash measurements - Two-wheeler :
Total length
Distance from front axle to rear axle
Distance from handlebars to rider’s seat
Height of handlebars
Height of seat
Width of handlebars

Post-crash measurements - Two-wheeler :
Total length
Distance from front axle to rear axle
Distance from handlebars to rider’s seat
Height of handlebars
Height of seat
Width of handlebars

4.2.13 Casualty Data (Two-wheeler) Module

This section is essentially used for the same purpose as the casualty data for car occupants, with the exception that variables associated with two-wheel motor vehicle riders (helmets, special clothing, etc.)are included.

Variables :

Type	Alcohol	Helmet damage location
Behaviour pre-crash	Clothing thickness	Kind of helmet damage - inner shell
Age	Kind of special clothing	Kind of helmet damage - outer shell
Height	Rider stayed with bike	Helmet ejected
Weight	Type of helmet	
Gender	Cyclist helmet	
Vision aid	Cyclist helmet - approval type	
Physically disabled	Helmet damage	

4.2.14 Pedestrian Data Module

The objective of this section is to define data specific to accidents involving pedestrians. It is important for one to have a good knowledge of the position and shape of the impacts on the vehicles involved, in particular for light vehicles, for which regulations have been foreseen.

Likewise, information on the pedestrian's position and movements before the impact is interesting. The pedestrian's trajectory is fundamental for speed estimations.

Knowledge of the presence of braking before the impact is also very interesting because, depending on the vehicle, the height of the bumper may change significantly.

Variables :

Age
Height
Weight
Gender
Vision aid
Physically disabled
Alcohol
Clothing thickness
First impact, vehicle

Second impact
Vehicle impact - Pedestrian kinematics
Pedestrian kinematics relative to the vehicle after the first impact
Run over
Pedestrian behaviour before collision

Direction before the impact
Children
Head contact point
Thorax contact point
Pelvis contact point
Leg contact point
Braking

4.2.15 Casualty Data Module

This module has all the necessary variables listed for all types of road users:

Type
Behaviour pre-crash
Age
Height
Weight
Gender
Vision aid
Physically disabled
Alcohol

Clothing thickness
Car occupant ejected
Ejection route
Kind of special clothing - two-wheeler
Ride stayed with bike
Type of helmet (two-wheeler)
Cyclist helmet
Cyclist helmet - approval type

Helmet damage
Helmet damage location
Kind of helmet damage - Inner shell
Kind of helmet damage - Outer shell
Helmet ejected

4.2.16 Injury Data Module

This section lists all the necessary variables to describe the injuries sustained by the casualty.

Using the variables available, analysis can take place at differing levels using the structure inherent within AIS. However, allowance has been made to take into account long-term consequences and effects upon society. There is no one methodology to determine this variable, and it is expected that differing studies will use differing means to assess this item.

An item from the NCAP tests is also included within this section and others relating to the tests are distributed throughout the document

Variables :

Death
Duration of survival
Long-term consequences
Trapping

Knee contact onto stiff structure
Hospital treatment
Further treatment needed
Economic consequences
Single Injury Description :

Injury description code
Duration of hospital in-patient treatment
Injury location
Injury causing part
Influence of intrusion

.

4.3 Work Package 1.iii – Data Collection

4.3.1 Table of Contents

4.3.2 Introduction

The objective of this task was to present different ways of collecting data that we will qualify in this report as in-depth investigations. We have classified the detailed accident studies into 3 categories. The first two categories are usually classed as "in-depth investigations". This classification is normally based on the time of intervention by the data collection team. The final category is described as 'hospital based'.

The first category are the so-called "retrospective" studies. These studies normally have a secondary safety objective i.e. the principal aim was to understand how involved people got hurt or were killed in road accidents. This required the inspection of the vehicles involved in the crash and not necessarily the crash scene itself. The vehicles could therefore be inspected some time after the crash and even at a different location.

Over time, the range of research has started to cover a wider area and certain volatile but necessary data (if we are to understand the mechanism processes studied) could disappear from inspections carried out by accidentology teams using the retrospective method. Therefore certain systems have been set up allowing investigation teams to get to the scenes of an accident as quickly as possible. We can call these data collection systems "in time on scene". With this type of system the analysis of the accident starts immediately, or very soon after the crash occurs. It is difficult and subjective to precisely define "immediately", and for our purposes this category is specified as on-the-spot investigation in the presence of the people involved in the crash or at least the vehicles involved.

It is clear that these two types of studies are centred around the crash which is the event that initiates the investigation procedure. However, there are some studies that are centred around injuries of road traffic accident victims and these are classified as "in-depth data collection". These studies are referred to as "hospital based". In this case the event that initiates the data collection procedure is the arrival of the injured person at a hospital. The crash and the vehicles are then taken as secondary elements which are not always studied in the detailed manner of the previously mentioned methodologies.

We are now going to present the advantages and disadvantages of these three types of study.

4.3.3 Retrospective Studies

4.3.3.1 General points

The principle consists, as the name suggests, in collecting data some time after the accident, generally one or two days after it occurs in the case of the fastest data collections. It is obvious that this type of data collection applies only to stable data or to data collected by others and conveyed to the team later on, as is the case with medical data. This type of data collection is applied mainly to vehicles, such as general characteristics or deformations, as well as to information on the scene and circumstances of the accident.

The term "retrospective" means "after the event" but the delay between the moment of crash and the investigation process can affect the data availability. Some specialist in-depth team use their own staff to inspect vehicle, casualties or crash scene within a small numbers of days after the crash. Normally, a considerable level of detail can be reliably available. Other teams, particularly insurance based groups, may bring together the written records of the police, hospitals and insurance assessors. These groups will generally not have access to the level of detail of the specialist teams and there may be more variability in the accuracy of the data.

4.3.3.2 The crash notification system

Owing to the fact the notification process poses one of the major problems for the setting up of accident investigations, the principle of the retrospective investigation allows more flexibility than the other methods. In fact, it is generally possible to trigger the notification system from different emergency services, such as the medical and police services and in France and Germany, the dispatch centre of the fire brigade. However, it is always easier for these departments to inform the investigation team after their intervention, for instance during the writing of a report. Moreover, there may be other sources of information which improve the quality of the information on the occurrence of an crash. For example, the breakdown mechanics who transport the vehicles after the crash. and of course insurance companies. It is therefore possible for one to increase the alert systems and in so doing limit dependence on just one system, thus improving the quality of the data on the existence of the accident. This helps if a sampling regime is followed by the study.

4.3.3.3 A priori data

In retrospective investigations, time factor is not a great constraint. This does not mean that there is no “time” constraint. Obviously, if one wants to note down deformations on the vehicles, this should be done before they are repaired. However, generally there are several hours, or even days in which to do this. The person who issues information on the occurrence of a crash therefore has time at their disposal for hindsight; this enables them to describe the crash succinctly in the alert message. The team involved in this kind of data collection generally know which type of users have been involved in the crash, or even obtains information on the circumstances (such as the presence or absence of a lateral or frontal collision). It is also possible to obtain further information, either from the emergency services, insurance companies or garages. A first estimate of the seriousness of the injuries observed may also be obtained. One may therefore have a good knowledge of the crash, generally enough to allow an initial classification and selection without teams having to go to the actual crash scene.

4.3.3.4 The choice of the crash

In retrospective studies, as in all data collection, one must choose the crashes for which data is to be collected. The information given when a team is informed of the crash allows possible sampling variations; ranging from the most extensive to the most selective So, one could just as well decide to exhaust a given alert, (for example studying all the crashes reported by an emergency service) as well as carrying out sampling on a certain quantity of crashes reported. The possibilities for sampling are very large but as mentioned in the findings of Work Package Two, the choice of the sampling method depends upon the research objectives. We will mention as an example the Co-operative Crash Injury Study, which is working on crashes involving cars which are less than seven years old and are required to be towed away by a recovery vehicle (with a sampling plan based on the seriousness of the injuries).

4.3.3.5 The method’s advantages

The first obvious advantage is that of cost. In fact, the staff designated to this type of study carry out their entire task during normal office hours. Moreover, the organisation of the alert which allows a certain degree of freedom in timing, also contributes to the reduction in costs. In this type of data collection system, the notification of a crash could be slightly delayed. But a system where the researcher is informed immediately costs more in terms of man power and technology than one where the reporting is more casual. Therefore, the choice of the crashes on which it is possible to carry out a study, a priori, restricts costs by limiting the rejection rate of the files. This rejection is generally due to the absence of values for data considered as obligatory or to a crash classification error; for example one is doing a study on recent vehicles and the vehicle involved in the accident does not correspond with the required criterion.

This characteristic is also interesting in terms of how motivated the data collection team is, because it is quite discouraging to collect data in order to eventually reject the files, especially if this occurs frequently.

The quality of the data noted down is generally improved by the absence of stress in a team which does not work "under the stress of an emergency setting".

4.3.3.6 The disadvantages

The major problem here is the noting down of volatile information. In fact, with time, it becomes difficult to note down certain data. Let us take the example of marks on the ground: the more time passes, the lesser the chances are of having clear traces which can be easily attributed to the vehicle studied. It may even be impossible to note down some information, such as the exact position of the vehicle after the collision, and one has to therefore rely on the measurements taken by others, the police for example.

The problem comes up in a more general manner for data which needs to be taken down almost instantaneously. The method then commonly used, consists of using the services of other people intervening, such as the police. The problem may then become one of quality or of the reliability of information due to the lack of specific training in recording data in the way a researcher would do. The objectives of the police, for example, and those of the researcher are simply not the same.

4.3.4 On the scene, On time Studies

4.3.4.1 General points

Here again the general principle of the study is summarised by its appellation. It is a question of rushing to the scene of the accident as quickly as possible after it occurs, ideally before the vehicles involved have been removed. It seems clear that the speed of the team's intervention is essential to the quality of the data collected. This is the most sensitive aspect of "on the scene on time" investigations.

The solution is to be informed at the same time as the rescue services. This poses problems because the investigation teams do not necessarily have a status which puts them into the same category as such a service. Often one has to create very strong ties with these services so that the alert may be conveyed to the investigation team. Initially, they can be uncooperative due to the extra workload and time and effort is required to overcome this attitude.

The second major problem entails that of carrying the investigators to the scene of the crash. Generally with the members of the STAIRS consortium, the investigators do not travel in the vehicle of the emergency services (doctors, police, fire brigade) but in a separate team vehicle. Only the Hanover study has the required permission for using a blue revolving light, which obviously improves their chances of getting about quickly. For the others, arriving in time to collect the necessary data is their first challenge.

4.3.4.2 A priori data

The data is of the same accuracy as that obtained by the emergency services. That is, generally the team travelling to the scene of the accident has an imperfect and sometimes totally erroneous knowledge of the accident which is to be studied. So the seriousness of the accident or even the type of vehicles involved or their number is generally unknown at the time when the data collection starts.

Let us point out once more however, that this is very dependent on the notification system set up, the choice of the system being often imposed by the different legal or statutory constraints, or even simply due to the habits and tradition of the potential notifying authority.

4.3.4.3 The choice of crash

Contrary to the previous type of study and in view of the comments made above, it is difficult to make an a priori choice of the crash to be studied. One solution may be to choose the exhaustive intervention technique, which involves studying all the crashes reported. Moreover one may also, as in the previous case, choose the crashes after the event. This second solution having as its main problem the rejection of a certain number of cases, which has its drawbacks. It is also possible to judge on the spot whether the crash which has just occurred is eligible for the study in progress according to the different criteria defined. This causes a lot of wasted time in travel and rejecting a part of ones work is very demotivating for the team members. Moreover as in the case of the exhaustive option, the problem of the saturation of data collection teams may arise, for example two crashes occurring simultaneously or almost simultaneously, in different places.

Another solution exists. Choosing the crash directly through the notifying organisation, if this is feasible. This would entail the police or the fire brigades deciding if the accident corresponds to the criteria demanded by the study protocol; if so, the specialised team is alerted. However, this system has its problems. Having to make such a choice is not normally a task for the emergency services concerned and this can cause problems in certain borderline cases, where there is a doubt concerning their eligibility for the sampling criteria.

Experience has shown that in such cases the accident is often left aside, this of course constitutes a bias in the selection process.

4.3.4.4 The advantages of the method

It is obvious that this method, as opposed to the retrospective method, makes it possible to note down volatile data. In fact, all the primary safety data may be noted down under these conditions (in particular data required for reconstruction), and can be of the highest quality. We will obviously mention as a priority all marks on the ground, resulting from braking or skidding and which are good indications of the trajectory of vehicles. Moreover early knowledge of the medical report, even if it is brief, is of help when inspecting the vehicle. It is obviously easier to find occupant contact marks under a dashboard when one knows that there was a lower limb injury.

This method therefore clearly permits a wider range of study than retrospective data collection covering both the primary and secondary safety aspects of a crash. In fact we will mention French studies which cover even the field of driver behaviour, with the intervention of psychologists on scene, and the Hanover study which covers the field of trauma pattern and treatment effectiveness.

Another interesting point is that the data collection may thus be carried out almost entirely by the team's staff. It is therefore a lot easier to control the quality of the data collected.

4.3.4.5 The disadvantages

The major disadvantage is the cost. In fact, intervening in real time on the scene of the crash generally requires round the clock intervention every day. Staffing costs therefore increase due to people being required in the teams for the entire data collection range to be covered.
The second disadvantage is certainly the setting up of the notification system. This is proven in the two studies of this type presented in STAIRS; Hanover and INRETS/ CEESAR. Both require specific devices involving the Emergency services whose collaboration is not necessarily easy to obtain. In particular, an eventual increase in the number of data collection sites could give rise to installation problems by increasing the chance of non co-operation, which to date have been avoided.

As for the sampling of crashes, we have seen that it is less easy than for the other, retrospective method. In the two in time, on-scene investigations studies mentioned, the choice of the exhaustive method of crashes reported has been made.

4.3.5 Hospital based

4.3.5.1 General points

This concerns a systematic register type collection, (similar to compiling a register of cancer cases or of other serious illnesses), with the basic entry being the individual victims of a traffic 'accident'. This type of collection functions for a given population which must be clearly defined. For example, in the Rhone's accident register, one will only find individuals who are victims of crashes which have occurred in the Rhone region. This department is a geographic and administrative sector covering a certain surface area of the French territory. The data collected is mainly medical. The problems encountered using this method (we will go into more details later on), surrounds the collection of more information on injuries than the vehicle or the accident itself.

A register should be systematic and exhaustive. This obviously requires the active collaboration of the different medical departments involved. The problem of setting up the system is therefore very dependent on the area chosen and on the organisation of medical care in the area. In this respect the French experience with the Rhone region is a good example since this department (chosen because of the presence of a large INRETS centre), has been working with more than eighty different hospital departments!

4.3.5.2 A priori data

With this type of register, the only data required is knowledge of the road traffic casualty. This may be known on arrival at the hospital department or later on. What is important, is that it is known at some stage and that the corresponding data is recorded. It should be noted that the information which allows one to classify the casualty, for example by the category into which he/she belongs (light vehicle, pedestrian etc.), is sometimes known on arrival at the hospital but this is not always systematic. This data is generally obtained by interviewing the casualties or their close relatives directly.

4.3.5.3 Choice of the victim

In the case of a register, the collection process is systematic. The choice of the victim is therefore simple: either the victim corresponds to the inclusion criteria or he/she does not. There are two categories of these criteria:

- The victim is the victim of a crash which has been defined in the study's protocol.
- He/she belongs to the studied population defined in the same protocol.

The exhaustive system is in principle a simple method, although difficult to carry out.
However, in order to confirm that the study is really exhaustive, you have to be certain that no particular case has escaped your attention. This is generally where problems occur because evaluating how thorough the study has been is always a difficult task.

4.3.5.4 The advantages of the method

One of the main advantages of this method is the fact that the data collection staff are also hospital staff and therefore already working at the hospital. Intensive training is therefore not required for this task. It is thus possible for one to have a large number of correspondents in a large number of hospital departments (about ninety for the Rhône register). This brings us to the second interesting point, the number of cases dealt with annually may be very significant (approximately ten thousand for the Rhône register).

This type of (mainly) medical data collection gives a picture of the casualty rather than of the accident and one must not forget that the ultimate goal of crash studies (through different means such as regulations or the improvement of vehicles or networks), is to reduce the number of deaths and generally to decrease the seriousness of the crashes. That is, the objective of these studies focuses on the user only and not the crash as a whole.

Finally, we will note that with this system exhaustivity is a realistic objective. Moreover, if we consider that the sampling is directly dependent on the question put forward, exhaustivity then becomes an even bigger advantage.

A hospital based register of injury and casualty details is generally self contained. However in some circumstances it is possible to link the data to other datasets (e.g. the police) to increase the analysis possibilities. All the same, such links between databases must be manipulated with great caution. In particular, one must respect all existing legislation, national or European and this therefore involves good knowledge of these laws. It is obvious that the setting up of a link between databases of a different nature must have a clearly defined objective and must demonstrate its scientific value.

4.3.5.5 The disadvantages

The main disadvantage, other than that which is purely medical, is insufficient information. In fact, it is generally easy to know the type of road user involved, (car to car, pedestrian to car, etc.) but on the other hand more difficult to know, from reliable sources, whether the seat belt was fastened or whether there was an air-bag present. As for the deformations and the characteristics of vehicles, one cannot obtain this information from a purely hospital collection of data.

The second problem is in conflict with the main advantage: the hospital staff are not necessarily available for an additional task and this may therefore lead to a real, but not easily detectable dysfunction of the system. Exhaustiveness being generally the goal, the loss of information is fraught with consequences. This is why the redundancy of information sources is of primary importance in this type of data collection : two sources for the same information is better than a lack of information.

Finally, and this depends on the local legislation, the medical registers are often closely regulated and this can complicate the setting up of the system.

4.3.6 Some suggestions

To choose a system you have to make an assessment according to the principal criteria of:

- the objective of the study and therefore the data that you wish to gather
- the means that you can bring to bear, on a human level as well as on a financial one
- the local organisation of the emergency services

In order to carry out of the study, you have to clearly establish the data list that you intend to collect. Keeping in mind that further you go back in time away from the moment of reference (the instant where all movement of vehicles or victims a trajectory sense have come to an end), the more difficult will became an on-the-spot study. Obviously as we saw earlier on, the setting up of on-the-spot studies is more expensive in human terms and could therefore limit the choices available.

Finally, and this is not necessarily the easiest problem to resolve, the local organisation of the emergency services may influence the choice of collection methodology. It is impossible to set up a system of accident notification for an on-the-spot study there is a categorical refusal to collaborate by the emergency services. In such a case either change the system or change the study site if that is possible.

The following table sums up the different possibilities:

	RETROSPECTIVE	**ON THE SPOT**	**HOSPITAL BASED**
Secondary safety	Yes	Yes	Very little data on vehicles
Primary safety	Partially possible	Yes	Need specific studies
Accident cause	Not suitable	Possible	Need specific studies
Injury details	Independent	Independent	Yes
Cost by case	Low cost/case	High cost/case	very low cost/case
Man power	Reference level	Higher	Lower
Specific team (gathering)	"Universal specialist" (except for medical data)	Several specialists	Except for specific studies
Notification system	Relatively easy	Difficult	not relevant
Special agreement (except for involved people)	Police reports Hospital data	Police (access to the scene) Hospital Emergency services	Hospital (public and private) Emergency services
Statistical sampling	Easy, movable	More difficult, more often fixed	Possible exhaustivity

4.4 Work Package 1.iv – Data Quality

4.4.1 Table of Contents

4.4.2 Executive summary

The objective of this task was to define a minimum set of procedures that can be used to ensure that data is collected and recorded to a sufficient level of accuracy. The areas covered are :

- Data collection,
- Compilation of data,
- Initial processing of electronic data,
- Comparison of data from several sources.

Within this structure there are three data levels :

- Collection (observation and recording),
- Coding,
- Analysis.

It was found that to ensure the least degradation of information, quality management procedures should be directed at the collection and coding stages allowing for accurate analysis to take place. The recommendations to be followed to ensure that data quality and accuracy is achieved are as follows :

1. Select a balanced team,
2. Ensure constant updating of skills,
3. Produce a glossary of terms,
4. Use an objective method of recording data (photographs, video) wherever possible,
5. A series of checks should be in place at relevant stages to allow for :

   i. A manual logic check at the collection phase,
   ii. An impartial check of the information,
   iii. A check to ensure correct conversion to an electronic format,
   iv. A self-consistency check to ensure appropriate coding,
   v. A management check.

Feedback loops should be established throughout the system. If a problem is identified, this will allow corrections to be made at the earliest opportunity.

4.4.3 Introduction

The areas covered in Data Accuracy are :

- data collection and compilation,
- processing of the electronic data,
- comparison of data from different sources.

The data collection process begins with investigator training. The differing collection methodologies will determine the areas within which training will be required, as well as the tools to use. However training should concentrate on the areas of understanding definitions and interpretation of information. This will allow for accurate coding on the data collection forms.

Data compilation will cover team make-up, visual evidence and recording techniques, as well as combining the accident, vehicle damage and occupant injury information in order to form a completed case examination. Further details surrounding electronic checking for self -consistency of the data and its coding options will also be discussed along with data checks on information obtained from different sources.

All these processes will be identified and presented in a graphical format, in order to show the points of feedback when identifying training needs, as well as quality control, within the system.

4.4.4 Data Collection

The primary source of information collection comes from the investigators who record information at the crash scene, collect data from hospitals or retrospectively inspect vehicles. In each case the quality of information is dependant upon the accuracy of the investigator when collecting both the factual and interpreted data. As this is the initial stage of the investigation system it has to be as accurate as possible. The main problem that can occur is human error due to the various pressures prevalent in each collection methodology. These pressures can be practical constraints such as the Police trying to clear the crash site as quickly as possible, to uncontrolled influences such as inclement weather. The conditions can be such that they can adversely affect the collection of data necessary to meet research objectives as well as the accuracy of the information. The answer to many of the problems is appropriate training.

There are two areas that the training will need to cover: collection of the data (which includes observation, interpretation, and recording of factual data) and coding of the information (such as the use of the Abbreviated Injury Scale[6] for injuries).

4.4.4.1 Collection and recording of factual data

The initial area is measurement; a skill whereby there is a need to know all the necessary points at which to take the measures and ensure accuracy by means of the tools used. Any theory and calculation exercises that may be necessary can be classroom based and the proficiency of the investigator assessed before being allowed into the field.

The main areas of study should be based around impact kinematics, biomechanics and vehicle examination The investigator must fully understand how to use all the tools at their disposal and the circumstances in which each one is the most efficient, practical and accurate. Training within this area can be accomplished relatively quickly, especially as the collection form can dictate where and what is to be measured and any calculations that may be necessary, thereby acting as an aide-memoir. To further enhance this accuracy, information sheets containing details of standard measurement points, Collision Deformation Classification, Abbreviated Injury Scale coding and other established systems should be available for on-site reference. All data and measurements to be collected must be very well defined and described in a comprehensive manner.

Currently, there is a shift towards the electronic collection of data, however it must be borne in mind that technical sophistication does not equate with accuracy. Training in the fundamentals of crash investigation is still necessary for data accuracy.

---

[6] Abbreviated Injury Scale, 1990 revision, Association for the Advancement of Automotive Medicine (AAAM), Des Plaines, Illinois, USA.

4.4.4.2 Collection and recording of interpreted data

The second area is interpretation; a skill that requires not just knowledge of the coding system (such as the Abbreviated Injury Scale or Collision Deformation Classification), but also experience in the use of these systems in order to be accurate and best represent the circumstances of the crash or injuries observed. For this skill to be of benefit, the research objectives and tolerance levels must be known so that undue time and effort is not wasted in over-detailed examinations that will not provide any useful information for analysis. Peer review on-scene will go some way towards minimising this problem, but cannot solve it entirely.

These tolerance levels will not be known until the collection methods are in place and the tools to be used decided upon and their accuracy noted. It is these that will dictate the level of investigation necessary.

There is also the problem of definitions, which is an area that is being covered by the core variables work package. A thorough understanding of the technicalities of the information to be collected must be inherent throughout the process along with a knowledge of the appropriate place for certain types of interpretation within the system (some interpretation is best left to the analysis stage).

Due to the nature of this specialised area, a large part of the job done by the investigators should be ‘in the field’. This is an on-going skill that must be constantly developed, especially with the ever changing conditions experienced in vehicle construction.

4.4.4.3 Coding

The coding process is a method whereby a system of words, letters or symbols are used to transmit a complex idea or message. For example, if the damage to a vehicle had to be described using a written reporting procedure, it would be long, complex and unsuitable for in-depth, aggregated analysis. Therefore a number of coding processes were devised in order to convey the approximate damage profile. The system STAIRS has decided to use is based on the Collision Deformation Classification (see Work Package 1.ii - Values and Variables). This is a coding system that uses letters and numbers to define the direction of the principal force, the location and extent of deformation and the general type of damage.

Even with a long established system (in crash research terms) such as the Abbreviated Injury Scale, coding differences do occur between the different collection organisations, who have their own wants and needs from the coding systems used. As this part of the procedure is closely linked to data interpretation, which is itself linked to the variable definitions as discussed previously, then these are required to be in place before coding can be checked.

Accuracy of the coding process is initially down to the level of training the investigator receives. This can then be backed up by regular meetings of the personnel dealing with all facets of the collection process, in order to reach consensus on any areas that may need clarification or further training (See fig 1.). These changes may occur due to a new set of requirements for the investigation, or perhaps the implementation of new technology. Whatever the reason, continuous training and updating of the coding process is necessary to retain a high degree of accuracy.

One person should be given control of the final decision in order that progress can be made in cases were consensus cannot be reached. If it concerns interpretation of the information, this should be included in any manual, glossary of terms or amendments to the coding form so that there is a constant update of prevailing definitions for the complete range of variables collected.

To ensure a further level of quality and accuracy in the coding process, an electronic validation suite can be used for a range check and self-consistency. This can occur at a later stage in the proceedings when other areas of the collected data can also be cross checked.

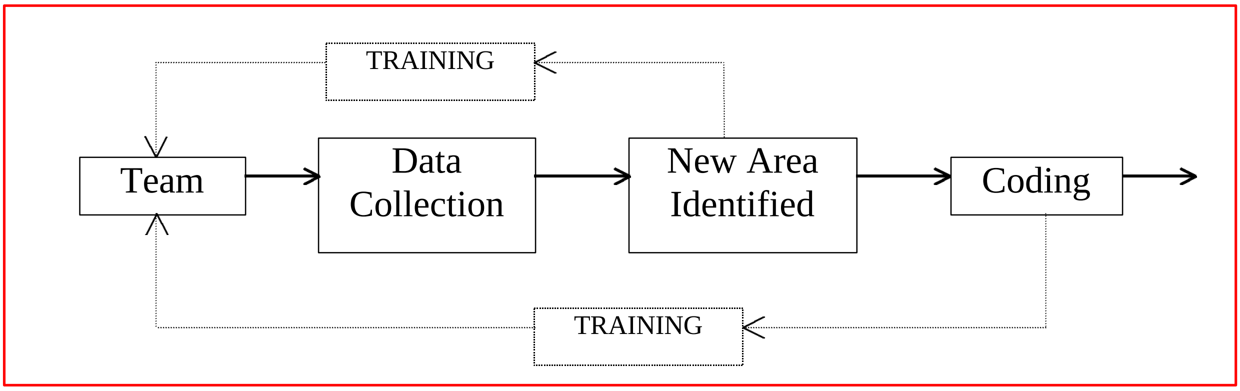


Figure 1: The Coding Loop

4.4.4.4 Measurement accuracy

The accuracy and subsequent quality of data is dependant upon three influential factors :

1. The different tools used have different tolerance levels (tape measure or laser measure),
2. Different measures give different tolerances (undeformed and deformed),
3. Different people have different tolerances (training and work ethic).

There should be regular internal evaluation to ensure the accuracy of the data is within the accepted parameters of the research objectives. The methods used to obtain this information will be wholly dependant upon the project in hand, as the acceptable tolerance levels may change with the objectives to be met.

4.4.5 Compilation of Data

The main topics surrounding the compilation of the data are :

- Team make-up,
- Visual evidence,
- Assembling the case,
- Electronic checking.

4.4.5.1 Team make-up

Each partner has a team of experienced investigators who concentrate on certain aspects of the crash collection process; from road engineering to medical specialists.

This type of team makeup is essential in order that the most can be made of a situation that is of a temporary nature. It allows a focused judgement to be made, at a level that is detailed enough to produce useful data, by an individual who is conversant in all aspects of the area in question.

This practise also allows for an in-built quality control check. With more than one person collecting the data, each will highlight the plausibility of the other. There is also the added advantage of another available individual that can give support to help identify or code a difficult element of a crash; adding to the quality and accuracy of the data collected. This level of accuracy can be compromised by a failure to have enough of the right personnel available for the collection process, either on-scene or retrospectively.

4.4.5.2 Visual evidence

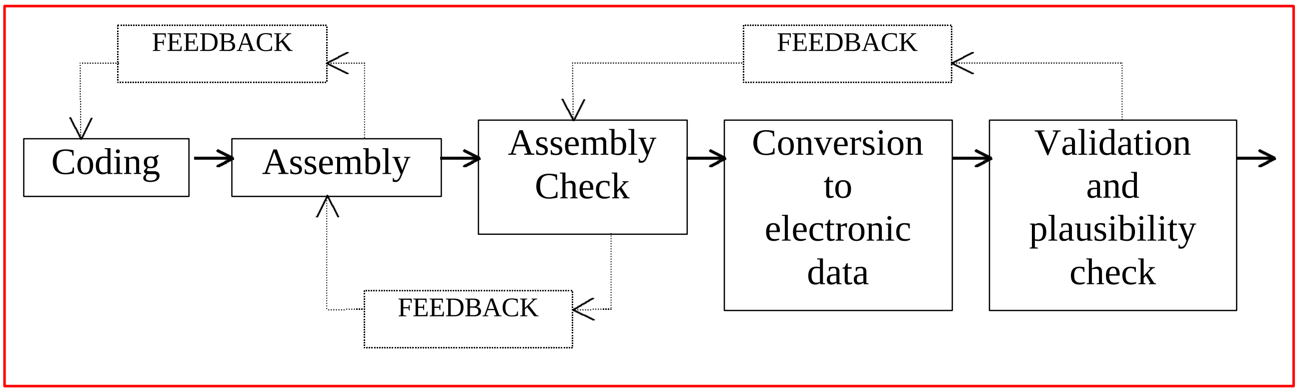
To back up and further enhance the written data, a visual record should also be made. This can take many forms, from a diagram to a video or a combination of different media. It should have two levels of information present: firstly, an overall view of the vehicle and if possible the situation at the scene of the crash, and secondly, a detailed picture of the areas of particular interest (collision partner, injuries, traces on the road, etc.). To obtain the best data, at least one of the media used should be as dispassionate as possible, such as photography. Essentially, the aim is to clarify or support the interpretation of the investigator and enable a person unconnected with the collection process to understand what has happened and agree with the coding applied. There is also the added advantage of having data that can be used in crash reconstruction.

4.4.5.3 Assembling the case

When all data has been collected, it should then be assembled together into a coherent case. There are four methods of doing this :

- An investigator from the team is nominated to perform the assembly,
- A joint effort is made by all parties involved in the collection process,
- An external person unconnected with the data collected is assigned the task,
- A combination of the above.

Figure 2: The Assembly Loop



4.4.5.3.1 Team Investigator

The benefits of having a team member who helped in the collection process to do a case assembly, is that of memory and awareness of the subtleties within the crash that may not be obvious to a person unfamiliar with the case. This can, however, lead to a more subjective interpretation of the information or worst still, omission of relevant points that may seem obvious during the collection process, but which are difficult to notice if you are unconnected to the case. These problems can be overcome in part if another, post-assembly check is made by a person not directly associated with the case. Any problems concerning data interpretation or omission can then be identified.

Although this adds another level of quality control and the attendant costs in time and money, the end product will be of a higher, more consistent accuracy.

4.4.5.3.2 Joint Effort

As teamwork is the strength within the collection process, so it can be at the assembly stage. With the differing views from each member available, a more harmonious case should result. None of the subtleties will be missed and all relevant points should be noted. The difficulties arise however, at the practical level.

Having two or more people working on a single task is not an efficient use of manpower and is sometimes not physically possible. Also, as each specialist wants their area to have the profile it deserves, there can be disputes as to the validity or inclusion of certain items of information. There is, of course, the option of each specialist writing up their own particular area of expertise. The problems that may arise from this have already been covered and once again a post-assembly check can nullify these effects

4.4.5.3.3 External assembly

Having someone totally unconnected to the case in question to do the assembly of all the differing sections that make up the product has some advantages. There will be an unbiased view and interpretation of the information that will be consistent throughout. That consistency will cover all cases and thus provide a standard product. This does not however, guarantee accuracy.

Initially the problem occurs with a knowledge base of the specialised areas required at a depth of understanding that will enable all facets of the case to be identified. There is also a requirement to update that knowledge continuously. This would mean that the individual would have to be drawn from the current staff of the collection area, and with the small size of the teams taken into consideration it would be more efficient, and probably more accurate, to use the joint approach method.

4.4.5.3.4 Combination methods

A combination of the best parts of each methodology into a single approach would gives an enhanced accuracy and interpretation to the data. Each area or team have their own strengths and weaknesses, and by utilising the appropriate strengths a high quality, accurate and consistent product can be achieved.

An example of one such synthesis would be for a single person from the actual data collection team to be assigned the task of assembling the complete case, but with the added task of communicating with their colleagues on matters that may require precise definition or interpretation. This method avoids a totally subjective view and also enhances the knowledge base of the team.

This feedback, loop system should not just be between the individuals but should be throughout the system so that any new or unusual developments can be identified, discussed and a protocol developed to handle the data that will be consistent in every case. Regular meetings should be held to pass on new or improved information or methods of handling cases to enable everyone to work to the same standard.

One of the main strengths of this system is that an initial manual logic check can be made of the whole case data at one time. However, due to the problems of a single individuals subjective approach, it is advantageous to have a second opinion; a post-assembly check.

This is a second level accuracy check, whereby a different person goes through the case checking the overall logic of the data. It is now that the problem of a subjective case assessment can be rectified. This level of checking should be done by an individual who was not directly associated with the data collection of the case. This will highlight any discrepancies or difficulties that may be encountered by someone looking at the case from outside the team or even the organisation.

4.4.6 Electronic checking

With electronic data collection not yet in place, all the information now has to be converted into a computer format. There is obviously a margin for error in this task and strategies should be in place to combat this problem. Whether the data conversion is done in-house or by a professional outside organisation at least two checks should be in place.

The first check is that of accuracy of input. Human error whilst keying in large quantities of information is not uncommon, and the normal checking procedure for this type of activity is ‘double punching’. This process involves two different operators inputting the same information, at the end of which an electronic check is made to verify that both sets of information match. If they are different, the information is manually checked and the necessary amendments then made to ensure accuracy.

The second check is on the plausibility and validity of the information. This is to ensure that the information is sound and meets all the necessary conditions for it to be a true statement of circumstances. For this to be done a specialised validation software program is used. Essentially the program confirms that the coding given in a field is within the parameters allowed, and that it cross-checks with other information relating to it elsewhere in the data. This means that the data is self-consistent.

A further stage can be added at this time. A computer crash reconstruction program. Although not all centres collect the information to enable this type of check to be run, if it is available then as a further quality control this can be very useful.

Once again, if any errors are found then a system of feedback into the assembly stage and from there into the collection procedures, should be in place to alert people to the problem.

If the information is to be presented in a special final form for an outside body, a final check by a member of the management team should occur to ensure levels of accuracy, quality and presentation are being maintained. This check can take the form of a random sample of the work output over a set period of time and need not be comprehensive.

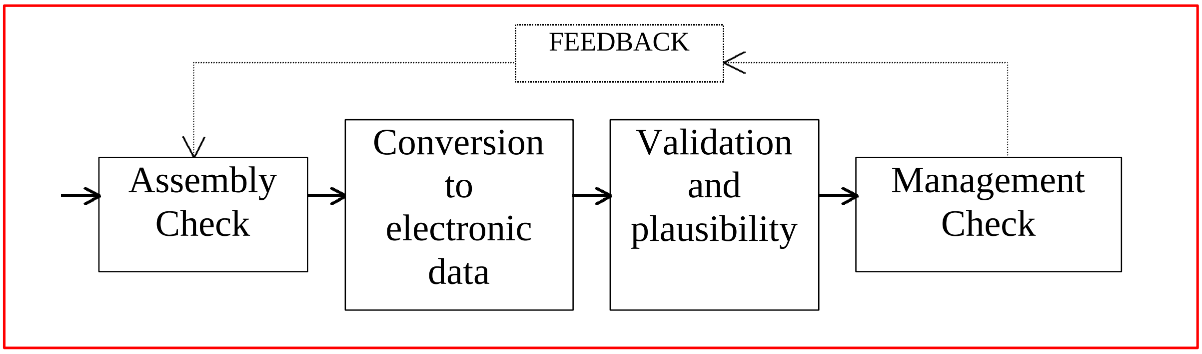


Figure 3: The Management Loop

4.4.7 Other quality issues

Although this paper is concerned with the technical accuracy of the data and quality control systems to ensure a consistent product, there is another aspect that should be highlighted.

The data has to have a certain level of 'user quality'; that is the ability to answer questions that will be asked of it. This area of quality is determined by the prevalent areas of investigation current at the time, which is in turn dictated by the overall aims of the funding body concerned. In order to be of use, a common database must have datasets from it's contributors that cover all the relevant areas of interest, both political and social, to such a level that it can provide useful information on any query that may arise. To ensure that this happens in a controlled manner rather than in a haphazard fashion, regular reviews of the core dataset should occur.

This also touches upon another area of quality; the comparison of original data from different sources. Some of the core data may be collected in a slightly different fashion by the different partners and might not be directly comparable to each other. This problem can be overcome by devising a framework of variables and values that encompass all the variations to produce a 'new' one. However, this area of investigation is more concerned with the production of the core variable set and should be dealt with by that particular section.

4.4.8 Recommendations

1. A balanced team should be selected, with the appropriate specialists in place.
2. Training and a constant updating of the skills necessary to ensure high quality collection of the data should be a main priority.
3. A similar process for the coding of the information should also occur.
4. A glossary of terms, updated as necessary, should be in place with a clear, precise understanding of the terminology and conventions used.
5. An objective method of recording data, either as the primary or secondary tool for investigation should be used wherever possible.
6. Assembling the case should have at least two stages:
    - The initial methodology to bring together all the separate parts which should include a manual logic check,
    - and a second, more dispassionate check.
7. Checks should be used to ensure conversion of the data into an electronic format is correct.
8. Validation of the coding and plausibility of the electronic data should occur to check for self-consistency.
9. Management check.
10. Feedback loops should be established throughout the system to allow for errors to be corrected and new conventions or training identified at the earliest possible stage.

4.5 Work Package 1.v – Validation

4.5.1 Table of Contents

4.5.2 Introduction.

With the variables and values, collection processes and data accuracy procedures decided upon, the next step is to validate these procedures and check if the information can be collected in the manner described.

In order to do this it was decided that each partner would collect three cases covering three different road user types. The French partner would cover pedestrian crashes, the German partner would investigate two wheeled motor vehicles and finally, the UK would look at car crashes. This covers the three main modules described in the value and variables list and should highlight any problems inherent within the scope of the document. It also allows the collection methodologies to be tested to the accuracy levels given in the latter two documents.

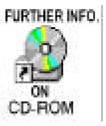
With the space available within this document, it is not possible to present the cases within this section. However, full documentation is given on the CD Rom.

4.5.3 Special circumstances

The initial problem foreseen was that of coding; the STAIRS protocol was never meant to be a coding system. However, in order to collect the necessary information codes would have to be developed. Much of the data concerning the different road users was already being collected by the partners but not in the exact manner required for the STAIRS protocol. For one partner, the French, it had been decided that the protocol would be the mainstay of their new database and a system was devised using the variables and values to make up a new collection process, but included extra information for their particular needs. This is precisely one of the purposes for which the STAIRS protocol was intended and would help to validate the process, at least with reference to pedestrians, completely.

Other problems that were identified concerned the availability of the information required due to the collection process being used. It was felt that some of the problems that might occur in this area could be attributed to the process of collecting the information and not to the variable itself.

Finally, it was decided that three cases from each partner should be enough to evaluate the procedures. Any major problems should be identified within this constriction, but obviously not all.



4.5.4 The cases.

The cases will be presented in different formats:

- The German data is derived from cases already held in their files; this should show that previously collected data can be directly transferable.
- The UK data follows the format of the document (variables, values, data), uses new cases and shows that the protocol can be followed prescriptively.
- The data from the French is in their own format. This shows that although there may be differences in the layout and other variables may be added, principally the protocol is sound.

The cases have used slightly different variables due to the constant development of the protocol over time. It is assumed that this will continue as new developments within car construction are integrated into the fleets.

Photographs of the case vehicles are also given on the CD Rom.

4.5.5 Practical difficulties.

Following the collection of the data, the partners reported on the practical difficulties they had experienced using the protocol.

Generally, the protocol was easily followed. However, there were certain problems concerning accuracy and interpretation. For example, the inclusion of areas involved in crush measurements using the enhanced CDC system. If one of the pillars is just included within the deformation to a door, should it be coded? Another example is the height of a vehicle's bumper. The moulding and angles on new vehicles make it difficult to precisely define where the measures should be taken from. The solution to this problem would be to design a training program that identifies and explains the values to be collected and the manner required in their collection to ensure accuracy and consistency.

A further problem identified within this area concerns the qualitative and quantitative nature of some of the variables. Some of these can be calculated, others even estimated, and the value given depends upon the judgement of an individual. Although this is considered normal it can be disconcerting when included within a document that aims to "standardise" the collection process. Again, training in this area would help to improve confidence in the system.

Linked to this would be the production of a collection form. The original form, although stated not to be a coding system, was seen as a check list and totally inefficient. Therefor, time and effort from each of the partners was duplicated in the design of a form that could adequately collect the necessary information. This was especially important for aspects of the vehicle that required collection repetition (seat belts, seat positions, etc).

All partners agreed that certain variables were difficult to collect due to the design of current vehicles. This involved information relating to parts of the vehicle that are totally enclosed or impossible to code without partially dismantling the vehicle (type of steering column, type of pre-tensioner, etc). The solution to this problem would be access to a database on all manufacturer's vehicle specifications.

There were other problems encountered by the collection teams, but these were related to aspects of the protocol that will be solved by the implementation of the training regime suggested above.

4.5.6 Implementation.

All three collection systems experienced similar problems, but only at the detailed level, especially concerning coding developments. This would indicate that further work is required in the development of work sheets for the coding process.

This could be facilitated by the different groups exchanging staff and allowing the free flow of information between the groups. Regular workshops could be scheduled with a different focus for each one e.g. coding of a particular variable in a specific way. Case review conferences could also be arranged for any particularly interesting or difficult crashes.

This may not be as difficult as previously thought. With the advent of computer technology and video conferencing; the exchange of information via e-mail and other means, allows for faster and more accurate feedback which can greatly aid this process. Every effort should be made to allow this interchange and further enhance the development of the protocol.

4.6 Work Package 1.vi – Confidentiality and Ethics

4.6.1 Table of Contents

4.6.2 Introduction

Within the framework of in depth investigation of traffic accidents, personal information is required. The type and volume of this information depends upon the relation to the research plan and approach. Most of the information concerns injury pattern and human individual data such as age, driving circumstances and others. There are some regulations for the receipt and use of such personal data.

The following general regulations exist for obtaining information

- Protection of personal data
  Personal data consisting of information about a living individual including expressions of opinion about him or her, must be protected.
- Professional discretion of injury data.
  Medical diagnoses are taken by doctors, who use these for treatment. No official transmission of such information for another purpose exists.
- Proprietary rights on personal image by pictures.
  It is not legal to take photographs of individuals without their consent. They have the rights to their own image.

In-depth investigations gather information about the individual and sometimes take photos of the person and of the injuries. For a scientific-statistic description and analysis of injury patterns information can be collected in a global manner (i.e. slightly injured/seriously injured/killed) and/or in a detailed way (i.e. leg fracture/arm fracture/scull fracture). For these details, knowledge about each single injury of a patient is necessary with regard to type, localisation and severity. For the compiling process, names of the victims are required. For a detailed analysis of long term consequences a full documentation of injury pattern and treatment is needed in clear and detail description. Information has to be drawn from the patient's hospital archives.

It is regarded as important to record the following minimum information:

- personal data (birthday, sex, day of first treatment)
- hospital stay
- injury type (detailed description and operational reports)
- brief accident description

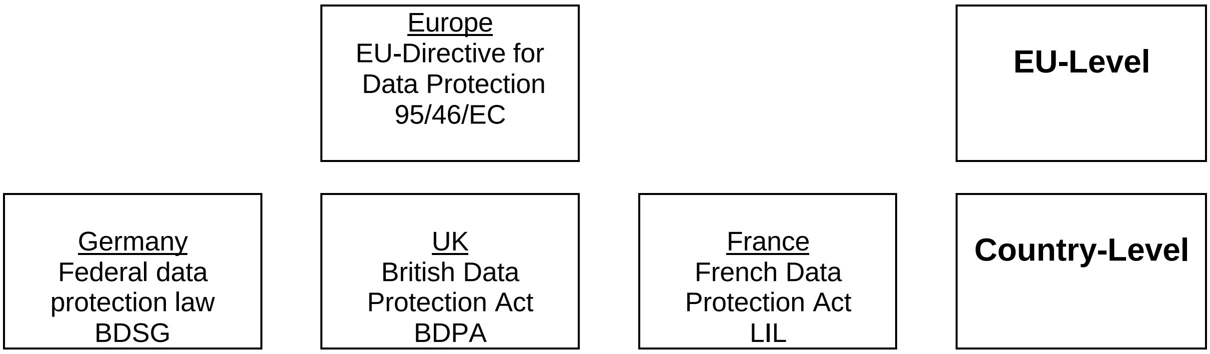
There are several ways to get hold of this information. A request could be made to the hospital or the doctor responsible for treating the casualty.

4.6.3 General Regulations for Data Collection and Storage

4.6.3.1 Rules of the data protection act

Data protection regulations are implemented to protect persons from the collection, processing and transferring of data for use by others without consent or for illegal purposes, so called data protection acts.

The differences in regulations set down by different countries with regard to such personal data also has to be observed.



For a long time, only a few European Member States possessed regulations concerning data protection. For some, the processing of personal data was handled, almost without any legal restriction. The European Commission worked out, for the first time in 1990 a series of regulations to answer questions concerning data protection, and issued a recommendation in 1992, on the basis of which the first European Data Protection Regulation was finally issued on May 1995 by the Commission of the European Union. The regulation was passed on October 1995 and published in the official paper of the European Community on 24 November 1995 (No. 95/46/EC). The regulation is valid for processing of personal data, i.e. for its collection, storing, depositing, connecting, changing, usage, passing-on, blocking and deleting. Records are not subject to this regulation, only the automatic data processing.

For many countries the EU regulations have not been integrated into their national law. Many national regulations already exist. The adoption process of these European regulations seems likely to be lengthy, particularly in countries that already have strong confidentiality regulations.

Therefore the national regulations on data protection should have priority concerning the activities on in-depth investigations in Europe. However, European regulation should be adapted if possible.

According to the EU-article as well as all national regulations, the processing of personal data is only permitted if :

- the person concerned has consented to this, beyond any doubt,
- the processing is required for accomplishing a contract position or a contract-related position of trust with the person in question.

The exception, contrary to earlier concepts and to different countries laws, is the claim that the data is derived from sources accessible to everybody and the processing is used exclusively for correspondence purposes. Furthermore it is not necessary for the person concerned to be informed about the recipient of his personal data prior to his consent.

However, a processing of such sensitive data is permissible if the person concerned has given his consent openly and publicly.

Data transfer within the EU must not be limited because of differences in data protection, i.e. the standard of data protection must be uniform within the EU. Any national limitation of the data exchange is prohibited for this reason.

So long the EU-Directive has not been adopted in the country where the accident data project is carried out, national rules apply.

4.6.3.2 Rules of professional medical discretion

The medical professional discretion is regulated in accordance with the declaration of the World Compound of Medical Profession of 1964 in Helsinki and revised edition of 1975 in Tokyo, in Venice in 1983 and Hong Kong in 1989, which emphasise the task of research projects. Written reports, explanations of patients, X-ray photographs and diagnostic reports are protected. An authorisation for disclosure can only be given if a declaration is made by the patients themselves or else, if a disclosure is relevant for the benefit of a more important cause. A transfer in the interest of science is permissible if anonymity is guaranteed. This rule of protection is also valid for non-medical staff, i.e. for secretaries, nursing staff etc.

The transfer of information is only permitted within the framework of further consulting activities for treatment between the hospitals and the treated medical professions as a whole. Permission for examinations should be given without pressure, with the stipulation that they are carried out in institutions under medical supervision. The authorised doctor serves the function of a medical expert and is bound to the code of medical discretion.

4.6.4 Application in practice

4.6.4.1 Application of the regulations for ARU Medical University Hanover

How far the research project Accident Investigations Hanover can comply with the existing legal regulations is described with the activities of the team:

1. section: Accident Registration and Data Collection on scene,
2. section: Data Storage as case description,
3. section: Computer storage of data and evaluation.

The data collected from the accident place, the hospital and retrospectively is entered into a data base. The enclosed personal data remains in an identifiable form only for the time required to process the retrospectively collected data. These reports are kept in locked filing cabinets to which only members of the accident research team have access. The anonymity of patients is assured.

The collected data are encoded onto computer files in such a manner that no individuals details can be identified. The data collected is multiply protected by pass codes against unauthorised access. The access to computer processing is only possible with an additional pass word and access to the archive exists for authorised persons only.

The following data is collected, divided into type and subject within the framework of the accident investigation:

1. <u>at the scene of the accident</u>
   photographs of the accident place,
   of the vehicles,
   of the traces,
   of the rescue measures,
   name and address of the car holder and the persons involved.

2. <u>at the recovery garage</u>
   photographs of accident vehicles as well as technical details.

3. <u>at the hospital</u>
   photographs of injuries to individuals,
   questioning of injured people,
   diagnostic report by the doctor,
   inspection of X-ray photographs.

4. retrospective
   traffic accident report and Police report,
   medical records, doctors letters and X-rays from hospital,
   judicial records from courts, autopsy photographs and reports, diagnostics.

At the accident site names of the people involved are established.

For documentation of the rescue activities the patient can for the first time be photographed.

In order to accomplish an extensive description of the accident, questioning of all the people involved will, as far as possible, be recorded on tape.

The patient, if conscious will be asked for his consent to the transfer of diagnostic reports, for scientific purposes. With an unconscious person this is done at a later time, and with children, by their legal guardian. If no consent is given, the undeveloped film, which is really not yet a picture, will be destroyed. Therefore only one film is used for each person.

The admitting doctor informs the team about the established medical condition of the patient and allows an examination of the X-ray photographs.

Medical reports and X-rays as well as accident reports from police stations and judicial records are required.

A copy of the written consent is enclosed with a letter which contains a declaration about the scientific purpose of the study and observance of the data protection regulations.

The investigation project ARU Medical University Hanover was supported by the Minister of Lower Saxony with an official letter. This allowed all in the town and rural district of Hanover authorised auxiliary accident stations of the police, entitlement to the transfer of traffic accident notices to the accident investigation Hanover. The confidential character of this information is emphasised. Furthermore, all hospitals in the region are asked for support in giving information about injuries to patients and to allow a photographic documentation of the injuries. This was done by referring to the scientific purpose of the study and in compliance with the data protection regulations. All medical superintendents contacted declared their support in writing.

All members and staffs of the Investigation Hanover are informed about the data protection regulations and have accepted the strict secrecy characteristics of the collected data and the results by filling in a special form.

4.6.4.2 Confidentiality and Ethics at Vehicle Safety Research Centre - UK

4.6.4.2.1 Handling of Confidential Material

Staff are aware at all times of the confidential and sensitive nature of the data being handled and do not discuss individual accidents with those not involved in the research programme. All those employed within the VSRC sign a Confidentiality Agreement not to pass on personal or identifiable information, verbal or written, to anyone not directly involved in the research programme.

The VSRC is a Centre within ICE Ergonomics, a company owned by Loughborough University. Loughborough University conforms to the UK Data Protection Act for handling raw and electronic data for research purposes. The data handling within CCIS is obtained from outside organisations, i.e. police and hospital records. It was therefore necessary to raise the matter directly with representatives from the police and health authorities based at the Data Protection Headquarters. A successful meeting was held between those representatives and the CCIS project managers from the UK Department of Transport and the Transport Research Laboratory to obtain the necessary clearance.

4.6.4.2.2 Access to Public Road Traffic Accident (RTA) Notifications

To access copies of all RTA's notifications from the police it was first necessary to obtain permission from the senior Police Officers in charge of the Traffic Divisions within each Constabulary.

Hard copies of RTA notifications are obtained by a member of the VSRC staff collecting these three times a week from police Traffic Headquarters and returning directly to the VSRC offices.

Supplementary data are accessed from one constabulary via electronic records. Access to electronic records are limited to the VSRC staff who are only able to read data relating to the accident, vehicles involved and those people involved in the crash. These data are added to the applicable RTA notifications where data is missing, e.g. police reference number, details of uninjured vehicle occupants, hospital attended etc. Missing data on RTA notifications might include information on injuries to specific body regions that are required for special topics of research. This information is obtained by contacting the police officer dealing with the accident. This information is also recorded on the hard copy of the original RTA notification.

Hard copy RTA notifications are sorted according vehicles involved. Any notification that is not applicable to the current VSRC requirements are immediately destroyed via shredding. The VSRC will however first record the number of these notifications received and brief details regarding the nature of the accident on an electronic database. This will not include any personal details of the people involved or vehicle registration numbers.

Those RTA notifications that are to be included for investigation are then sorted by severity of injury. A random sampling system is applied to the slight severity accidents. Those accidents falling outside of the sampling criteria are filed. The files are secured in a locked room. Those accidents that are included are placed individually into case file folders and will form the basis of a case file. Whilst the file is still 'active' the folders are secured in locked cupboards.
No electronic records are kept that include personal details of the people involved in the accidents.

4.6.4.2.3 Access to Hospital Injury Data and Post Mortem Reports

Permission to access injury data contained in hospital records is obtained from the Consultant in Charge of hospital Accident & Emergency Departments. More recently, hospital Ethics Committees are required to give their approval. In the case of accident fatalities Post Mortem reports are obtained from Her Majesty's Coroners.

Injury data are obtained in one of two ways. For the majority of the hospital the VSRC operates a system that allows our Medical Liaison Officer to extract relevant injury data directly from the hospital case notes under the supervision of a medical Consultant. In one large trauma hospital the medical Consultant dictates in the injury data onto a tape which is later transcribed by our Medical Liaison Officer. In all cases the injury data when received by the VSRC does not contain any names and addresses of the injured people.

The systems operate whereby the VSRC notifies each hospital of the names, addresses of the injured person together with date and time of accident. Against each name is a unique identification number. The hospital records are accessed by a medical secretary within the hospital. The Medical Liaison Officer records relevant injury data onto forms that only contain the identification number. In the case of tape recordings each person is only identified by an unique before the injuries are described. The hospitals themselves dispose of our written requests to access the case notes as this information contains names and addresses. No hospital records are removed from the hospital and only data relating to injuries sustained in the road traffic accidents are accessed.

A questionnaire is sent to each person involved in a 'case' vehicle. The questionnaire was given ethical clearances through the UK Survey Control Unit. No questionnaires are sent to people in accidents where there has been a fatality. In cases where people involved in the accidents have sustained serious head injuries or severe life threatening lesions the injury data is first obtained from the hospital to assess whether or not it is appropriate to send a questionnaire. When sending questionnaire an accompanying letter is included describing the purpose of the Study and assures the recipient of our awareness of the sensitive nature of the research. The recipient is also informed that any information contained that a questionnaire is strictly confidential and will not be passed to a third party in any form that will identify that person. Questionnaires are not included in raw material case files that are copied to the sponsors of the research.

4.6.4.2.4 Data Storage

Raw data: Completed data collection forms and photographs are sent to members of the CCIS sponsoring consortium and the Transport Research Laboratory (the project managers). The original file is retained within the VSRC. Before despatch to sponsors the file undergoes a rigorous checking procedures to ensure that no personal information is contained in the file and the photographs do not show registration plates. This ensures that names, addresses and vehicle registration are not contained in the files. Copies of Post Mortem reports and questionnaires are retained in the VSRC files only.

All raw data is stored in locked cupboards whilst the files are 'active'. Completed cases are locked in a central filing room within the building the VSRC occupies. The building itself is sited within a business park that is secured by foots patrols and video camera.

Electronic data: No personal data (names and addresses of vehicle registrations) are stored on electronic databases. The electronic data are therefore anonymous.

4.6.4.3 Confidentiality and Ethics at INRETS

4.6.4.3.1 General Information

Generally, accident studies carried out in France obviously respect French laws. As a reminder we will mention the primary ones; that is, the ”la loi Informatique et Libertés”, (”The Data Protection Act”), the law governing professional secrecy and the medical code of ethics.

The staff of the INRETS have general civil servant status and in this capacity are obliged to exercise a certain degree of discretion. As an EPST (Public, Scientific and Technological body) the INRETS staff are bound to a duty of confidentiality. Staff who do not have civil servant status but who work for the INRETS are also bound by the same duties. Those who do not respect these rules are sanctioned by administration or by a lawsuit if it concerns the breach of professional secrecy (especially medical secrecy, as far as in-depth investigations are concerned).

These studies lead us to use information held by the Police or the Gendarmerie, particularly during interventions at the scene of the accident. In order to be able to use this information, each investigator is provided with an authorisation in his/her name allowing him/her to consult the police statements and to work at the scene of the accident. These authorisations are requested annually, first from the Ministry of Justice, then, with the latter’s agreement, from the Home Office for the Police and the Ministry of Defence for the Gendarmerie.

As with the Hanover system, the investigator’s intervention is in-time on-scene. We then interview the people involved, a necessary requirement for our primary safety objectives. When this is not possible or in the event of a death, we interview the parties concerned or their close relatives later. Prior to this interview, the interviewer explains precisely who we are (a research institute), the goals pursued and asks for permission to collect the different information needed, whether this concerns the vehicle or is occupants. The interview is generally recorded and this is done, obviously with the consent of the person concerned. The refusal of an interview is considered a refusal of authorisation and the corresponding file is then abandoned, or destroyed if it concerns the case of a postponed interview due to serious injury, for example.

The interview is carried out in two parts, one immediate, on the scene, therefore inevitably short, and the other, later. It is during this second interview that we ask for medical information or the permission required to obtain it.

The vehicles are generally examined later, at the garage. If the investigation is carried out on the spot, as is most often done in the case of accidents involving pedestrians, permission is given directly by the user. In the case of a subsequent examination, this is generally done in a garage and then we request the garage owner’s permission. If a technical problem is suspected, we may ask the owner of the vehicle to dismantle certain mechanical parts. The places and injury traces do not generally pose a particular problem of confidentiality.

4.6.4.3.2 Coding and Storage

All information collected is then coded and the interviews gone through (in detail) and coded.
The documents input on computer are then anonymous (without username, address, etc.). The database thus created, is situated in each data collection centre.

The original documents (check-lists, interview reports, audio cassettes) are stored in locked cupboards on the premises of the accident research team. Only members of this team have access to these cupboards. Photographs are filed according to category, places or vehicles. They may be identified by the identification number given by the computer database. Therefore, if one may link together information gathered for a vehicle and its photographs, this link remains anonymous.

The coding and input are carried out by the investigators themselves. The computer data, is available for the four data collection centres. Data processing is carried out internally through the different partners. (INRETS, CEESAR[7]).

[7] CEESAR: Centre Européen de sécurité et d'analyse des risques.

4.6.4.4 The Rhône register

The structure of the register is a lot more complex because it requires numerous contributors. In particular, all the hospital staff associated with the register in the different hospital departments which receive road accident casualties. We would like to point out here, that in the department of the Rhône, there are about ninety different hospital departments, which participate in data collection. The personnel are grouped within an association, governed by the 1901 law, which manages the register. This association is only made up of hospital and INRETS's staff. It has a scientific council which is responsible for defining and checking the analyses and research carried out from the register.

The data collected is essentially medical but does contain a few additional points such as the road user type (pedestrian, cyclist, occupant of a car etc.), date and time of the accident, and the place of the accident. More precisely, it is an injury report established in each department. This data is noted down by the medical team when the casualty arrives in the department and also during his/her treatment. The hospital staff are bound to medical secrecy.

4.6.4.4.1 The data collection method

For each road accident casualty, and during his stay in each hospital department, a form with a detachable slip is established. Only this slip contains the patient's name and address. It also comprises a number which is repeated on the other part of the form. The slip is kept by the medical department. The anonymous form, which has the same number as the slip, is then transmitted to the register for coding and input. No data, other than that on the form from the register is transmitted. No copy of the radiological documents or the various biological examinations carried out on the patient is collected.

4.6.4.4.2 Data acquisition

The processing of the data is carried out by the INRETS on a network of four machines. To conform with the "Loi Informatiques et Libertés" (Data Protection Act), this network is totally separated from the organisations computer network. Only the data processing and the LEAT[8] staff assigned the task of the maintenance and analysis of the register have access to this information. Access is obviously protected by pass words. After input, the forms are kept in bomb and fireproof cupboards located on our premises. The register conforms with, (and is recognised as such), the different regulations and laws which are likely to be applied, in particular the "Data Protection Act". As for injury information for the immediate deaths, they are collected by the Forensic Institute, which participates in the establishment of the register.

In cases of a verbal request or one made by telephone, a written consent may not be expected and answers are given with general agreement of the principles of data protection.

It can be pointed out that in case of no existing national rules for data protection the European ones are recommended for practise for an European STAIRS project.

It can, however, not be ignored that the legal regulations for the protection of the private sphere of patients are still relatively new and that the interpretation of these regulations is still partly unknown or difficult or even diverging. Often it will have to be decided between individual rights of the patient and the importance of the pursued research objective for common benefit. There still exists the principle "All has to be done to protect

[8] Laboratoire d'Épidémiologie Appliquée à la sécurité des Transports

General Form for Data Protection Agreement used by the researcher

Within the framework of the accident research project, carried out by ...., accidents will be documented in order to fund possibilities for the prevention of accidents and for reduction of serious accident consequences.

All personal data and the information of injuries and injury pattern are exclusively serving scientific purposes and will not be revealed to the police, judicial authorities, insurance companies or such institutions, without special consent. The accident research team warrants that all personal data will be cancelled after completion of the task.

For the purpose of accident research, I consent to the scientific evaluation of my personal data, collected in hospital and by the police and agree that photographs of my injuries are taken exclusively for the purpose of the task of the investigation Accident Research Unit at Medical University Hanover.

Case No.:

Vehicle No.:

Personal reference No.:

____
(Signature of person involved in accident)

4.6.4.5 Amendments to Legislation to facilitate safety research

EU and much national legislation seeks to protect the rights of the individual from unreasonable intrusion and it is likely to be very effective. However one consequence is that the opportunities for health and safety research become more restricted. The difficulty of access to individual records means that the numbers of cases available for an analysis may be insufficient for a decisive result. Including a requirement to seek permission, when casualties may be seriously injured or dead, can be insensitive and can mean that a longer period of research is needed before a clear result is achieved. While further data is being collected other road users are being injured and are unable to take advantage of new countermeasures.

It is recommended that legal exemptions should be available for research improving public health and safety, Publicly approved studies, which handle data according to an agreed code of practise, should be able to use this data for studies without seeking permission from the casualties or their relatives. These legal exemptions would be incorporated within both EU and national legislation.

5 Work Package 2 – Linkage with National Accident Databases

5.1 Table of Contents

5.2 Introduction

Workpackage 2 is about the linking of in-depth accident databases to Regional[9] or National accident databases. The objectives of the work package have evolved during the life of the STAIRS project. Consideration of consortia member experiences and methodologies with in-depth databases suggest that it is perhaps best to consider this workpackage as a feasibility study rather than a definitive guide. Experiences, to-date, suggest that extrapolating from In-depth accident data to Regional data should first be investigated before extending this to extrapolations to National and/or pan-European accident databases.

Consequently, the analysis of the work completed in WP2, (data collection process, statistical requirements, data availability etc.) leads the STAIRS consortium to propose a methodological and theoretical approach initially between Regional accident data and existing In-depth accident data. The underlying methodology could then be extended to National databases and eventually to pan-European accident databases.

This methodological approach being suggested deals with the global problems of linkage to Regional, National or Europeans areas. For the purposes of this document the methodological approach is only applied to the 1st step of the linkage (on the Regional area). For National and European levels, statistical requirements and sampling considerations are proposed without examples.

This is for two main reasons, summarised as follows :

1. It is preferable to test and validate the method on practical cases which can be checked with simple examples dealing with known data, and in accordance with the statistical requirements presented in the methodological framework. This is only feasible to a Regional level when one takes into account the actual knowledge and the data available.

   Further developments of this method on National and European levels could be undertaken later on, after this validation. It seems reasonable to the Commission and STAIRS consortium to focus the technical application of WP2 on the linkage only at Regional level in order to validate the proposed methodological approach. The achievement of this outcome constitutes the first milestone leading to linkage procedures.

2. The second reason concerns the standardisation of the In-depth accident data collection. It seems that one of the main goals of the Commission concerning STAIRS is to promote the exchange of the In-depth accident data between the Member States who are involved in the In-depth accident analysis, and such an operational standardisation is the key means needed for an extended linkage to National and European levels.

   Furthermore, this standardisation of In-depth accident data is fundamental because the Commission will be co-ordinating several actions already launched dealing with the two main topics : primary and secondary safety using In-depth accident data and involving several actors (data providers and users such as Research Institutes, Car manufacturers, Governmental organisations, Insurance companies, etc.).

In summary, the objectives of STAIRS WP2 are :

1. Because of technical, pragmatic and strategic reasons, to focus only on Regional areas for the application of the methodological approach, (standardisation results need to be validated by a real exchange of In-depth data between Member States before undertaking a pan-European linkage).

2. To propose National and European level statistical requirements and sampling consideration.

[9] *Regional in the context of this paper means any definable area of the country, which might be a police area, or an administrative area where In-depth accident data was being collected. It does not necessarily correspond with a 'Region' as understood within France, for example.*

5.3 Methodological approach

The sampling strategy affects what can be estimated from the In-depth database. Different strategies result in different constraints and assumptions on the estimation procedures.

From a statistical point of view the usage of methods of statistical inference require a probability sample. A probability sample exists if every unit of a total population (e.g. an accident within the population of interest) has a known non-zero probability of being selected. This means that every accident should, in theory, be selected at random from a known population of accidents and that every accident has the **same** finite probability of being selected from the lowest level sampling group being used. There may be just one sampling group (all accidents), or the population of accidents may be split into groups defined by a stratification variable. For some research questions it can be useful to stratify, within every strata the sampling of accidents has to be at random, and the achieved sample should not have any known bias.

In order to fulfil these preconditions in the context of accident investigations, the application of a clearly defined sampling design is necessary. For this reason mainly in social sciences very complex methods of sampling techniques were developed within the last decades to solve the problem of getting valid results from (e.g. population) samples (e.g. Stenger 1985, Schnell/Hill/Esser 1993). The sample size plays an important role for a linkage with Regional or National Accident database. The lower the sample size the more imprecise the results. When the aim is to draw some conclusions from random samples it has to be taken into consideration that the sample population is distributed over the whole nation and is not limited in time. Thus, the selection of investigation areas has to be oriented on aspects of representativity. Therefore relevant criteria for the area selection have to be examined and on the basis of these results the regions can be selected.

However, if the investigation is limited to only one area or the selection was oriented on practical more than statistical aspects, a generalisation is only possible with regard to features that are not influenced by characteristics of the investigation area (for example the accident structure, the topography and so on). Nevertheless it is possible to get (analytical) research results, such as medical-technical and biomechanical correlation, that are valid Nationally. This applies, for example, to statements concerning the probability of surviving an accident in dependence of collision speed or the age of the drivers.

Suppose that the variable of interest is only within the In-depth database and an estimate of its' distribution is required at National level. The In-depth distribution will be weighted by one (or more) variables which will scale the distribution to estimate the Regional and then the National distribution. The weighting variable(s) must be in common between the In-depth and Regional/National databases. Further, the weighting variable will reduce the sampling biases which may be in the In-depth database.

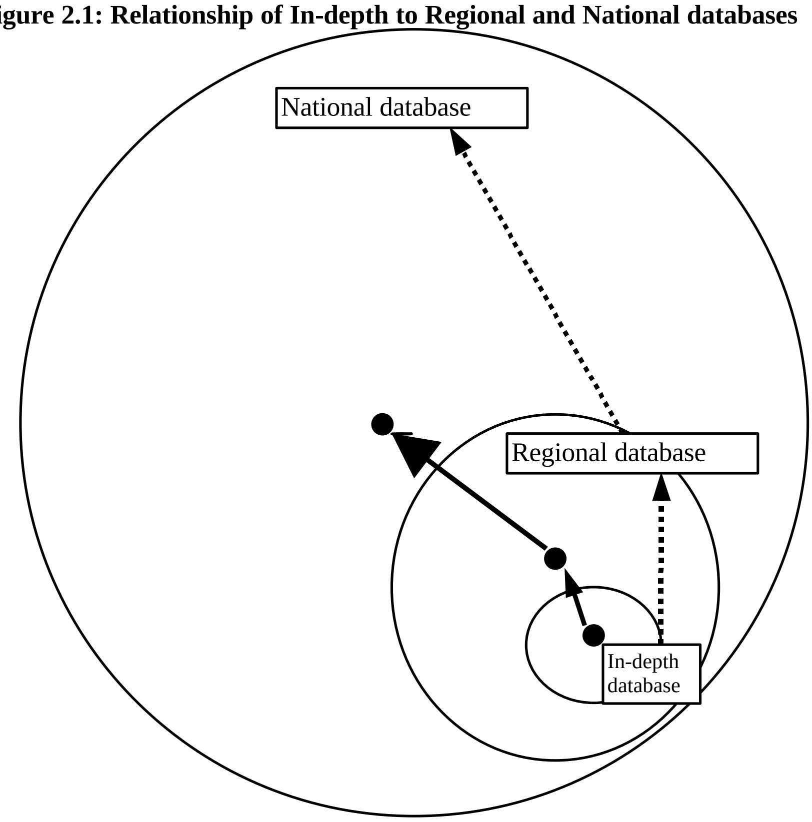
The following assumptions are necessary:

- the In-depth database accidents are included within the Regional/National database
- the Regional database may be a biased sample of the National database
- the In-depth database may be a biased sample of the Regional database

Note that we ignore the difference between National database and the real world (problem of the exhaustiveness of National Databases)

This is shown diagrammatically in the following figure 2.1,

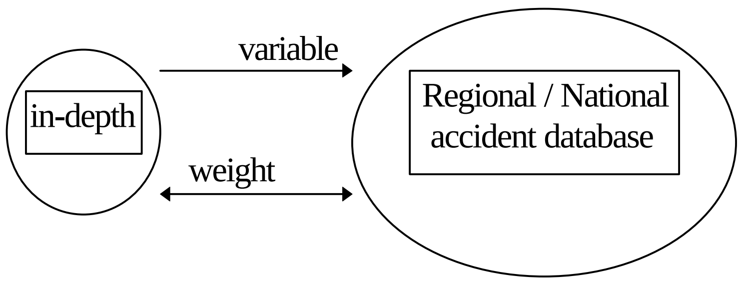
Figure 2.1: Relationship of In-depth to Regional and National databases



The question is then:

- is there any weight (w_i) to reduce the bias from the In-depth to Regional database
- is there any weight (w_j) to reduce bias from the Regional to the National database
  is $w_i = w_j$

Figure 2.2: Linking In-depth to Regional/National databases



We expect w_i & w_j to be functionally related to the variable of interest. The problem is to identify those weighting variables which are required for the variable of interest and to demonstrate that their use reduces bias.

The following points arise:

- the need to identify key variables for weighting purposes
- to try and identify types of variables which would use the same set(s) of weighting variables
- to use proxy weighting variables when necessary, e.g. δv is not a linking variable so one could use speed limit as a proxy
- to accept that there may not be suitable weighting variables and/or data in the In-depth database to provide the required estimate
- that it is essential to estimate the confidence interval on any estimate
- to realise that weighting from a small In-depth database to a Regional or National estimate may be imprecise and so have a large associated confidence interval

The above mentioned statistical requirements are the measure on which the realised sampling has to be evaluated for a linkage to Regional or National accident databases. To generalise the results of in-depth investigations to the National level the following selection steps are one way in which the statistical requirements could result in representativity:

(1) selection of the investigation areas
(2) selection of time intervals
(3) selection of accidents within the time intervals

There should be some check against known data at Regional level by which to check the adequacy of the in-depth sample, for example if the severity of injury by position seated in a car relative to the site of impact was the estimate being derived, then a suitable check might be the number of side impacts versus non-side impacts by severity. If the estimated number were consistent with the actual number then some confidence could be placed on the disagragated figures relating to seat position and side impact. Similar checks should also be made when moving from Regional to National level estimates.

The population of accidents for In-depth databases for at-the-scene investigations are not known a priori and hence must be selected opportunistically, however the sampling schemes applied ensure that there is no built in bias. Even though the strict statistical requirements are not being met, some linking to Regional or National databases may be possible. However, it is essential that the constraints of the data and the underlying assumptions are stated if Regional or National estimates are produced.

Methodologically, the estimation approach should first be to determine if the sample of in-depth accidents are representative of the Regional population from which they are sampled. This will check on the adequacy of the sampling method and the ability of any linking variables to provide adequate estimates at a Regional level. If the Regional level estimation was satisfactory then National estimates might be made provided that there were sufficiently strong sampling grounds. If the National level estimates also check out satisfactorily then the data may be suitable for pan-European purposes.

The reduction of bias by use of appropriate weighting variables has been discussed, weighting is required because the in-depth database is probably a biased sub-set of the Regional database and the Regional database probably a biased sub-sample from the National database. The selection of appropriate weighting variables will then reduce the bias. It cannot however, eliminate the bias caused because accident data were never included. This will be the situation if we are interested in a population of crashes which include, for example non-injury accidents in the UK since these would not be reported to the police. It is therefore important that the population of interest is not only defined but contained within the available data.

There is almost certainly some under-reporting of accidents where the damage and injury may be minor. The importance of this depends upon the answers being sought, and in some instances it may be desirable to conduct special surveys to try and determine the level of under-reporting. It has been suggested that conducting hospital surveys of patients reported as being injured in a road traffic accident is one way in which under-reporting can be investigated. This area of investigation is not of direct interest to the STAIRS consortia, but the consortia are very aware of the difficulties of under-reporting (FERSI 1996, IRTAD 1994, Koornstra 1998, Laumon 1997) and of the bias in in-depth and National databases which may result.

It is important to remember when applying the suggested methodology that the estimation process must make sense, in other words the population of interest must be available within the database and the data must be interpreted in an appropriate way. In order to achieve this it is vital to have a good working knowledge of the data, the data collection process, the data limitations and the statistical methodology.

It is also important to realise that some analyses will be interested in relationships between sets of variables, for example the relationship between footwell intrusion in frontal impacts by δv. This relationship may not be affected by weighting and so the results are applicable at an in-depth, a Regional, a National or a pan-European level. This is in contrast to analyses where the required answer is a number or a proportion, for example the number or proportion of casualties in particular circumstances, these estimates should be weighted to properly reflect the underlying population of interest.

The suggested methodology discussed above, is not unique, but it is the approach favoured by the STAIRS consortia. Other methods which have been used, for example a) the idea of weighting every observation such that when summed they automatically provide Regional or National estimates, and b) of meta-analysis.

a) The observation weights for this approach would be determined from the maximum number of 'linking' variables possible. The consortia's concern about this method is that different sets of weights may be necessary depending upon the population of interest and that obtaining good estimates of the weights may be problematic, there is also the problem in estimating confidence intervals and it is likely that using more weights than necessary will result in wider confidence intervals, (since the variability between weights will be reflected in the estimated standard error term).

b) Meta analysis is a technique for combining the results of different studies rather than the datasets. It is a quantitative method for literature review, useful in constructing statistical summaries and comparisons of studies. It can be a useful technique when the data are sufficiently similar that combining results can be justified. It is not considered further in this report (for more details, see Hedges, 1985 or Pettiti, 1994)

5.4 Supporting documents

The documents identified in the following table have been produced by WP2 members during the first twelve months of the project. They support the work in this work package.

Table 3.1. Supporting Documents

Title	**Country**	**Reference**
Common variables between CCIS & Stats19	UK	2.1.WP.FI.ATT.BS.006
Working document task 2.1: French approach	France	2.1.WP.1.ATT.BL.011
Linkage between the In-depth accident data and National Accident Database - methodological approach	France	2.1.WP.1.ATT.JL.016
National estimates from in-depth accident databases - UK	UK	2.1.WP.1.1.ATT.BS.022
Linkage between the In-depth accident data of Hanover and the National Accident Database	Germany	2.1.WP.FI.TT.IK.026

5.5 Descriptions and discussions of National approaches and their evolution

5.5.1 France

Four areas throughout France are concerned: One around Salon de Provence, in the south of France, investigated by INRETS, one in the "département du Rhône", essentially Lyon by another INRETS team, one around Amiens and the last around Evreux, in the North of France, managed by French car manufacturers (through CEESAR association). These areas were partly chosen because of the location within France, and partly because of the location of the different investigation teams. Hence, these four areas are not supposed to be a representative sample of France, although they are quite different from one another.

For the definition of the accident population and the data selection design, the principle is the same everywhere: to get a sample of accidents occurring inside the investigation areas. As the practical procedures are different from one area to another, we will only present, for illustration purposes, the Lyon investigation area.

5.5.1.1 Accident of interest

All road traffic accidents with at least one injured person are recorded, including those motorcyclists without an antagonistic vehicle.

5.5.1.2 Accident population

All accidents with probable injured people which are reported to the police inside the investigation area. It is principally made up of urban accidents occurring in the area under the authority of the Lyon city council.

5.5.1.3 Areas

The investigation area is the part of the built-up area which is controlled by the police (in France, three different professional organisations can deal with road accidents). This area is geographically defined, with a range of about 10 km, in order to be on the site in time to collect necessary data.

5.5.1.4 Data selection modus

There is no selection, notably according to accident severity. All accidents occurring during the investigation time are studied.

5.5.1.5 Data selection design

The data selection design involves an empirical sampling. Strictly speaking, this is not a so-called random design, as the probability of registering an accident is zero outside the time period of observation, and close to one otherwise.
This empirical sampling includes two steps: time interval selection and accident selection:

- time interval selection:
  - firstly one a priori selection of dates is made according to an equal distribution of the seven days of the week,
  - secondly a selection of hours is done by means of dividing the day in four equal periods (0-6h, 6-12h, 12-18h, 18-24h).
- accident selection:
  In theory, all accidents are eligible.
  In case of another reported accident whilst the team is already at the scene of an accident, the time of investigation is reduced to the recording of information which is likely to be erased very quickly, in order to arrive at the scene of the "new" accident as rapidly as possible. The team returns to the scene of the former accident as soon as the second accident has been dealt with.
  Therefore the selection of accidents inside an investigation time does not exist and in practice, this situation, which leads to conflict of interests, happens rarely as the average number of accidents observed within the investigation period is close to one.

5.5.1.6 Principle of accident selection (see above)

5.5.1.7 Practical aspects with examples (based all data from 4 areas)

To be able to obtain good estimates of statistical parameters, the ideal design would be a random sample. The further away we get from the ideal design, the more difficult it is to obtain interesting estimates. Basically, this concerns estimates of means or proportions, and of their respective variances. Central measures are mainly concerned with biases and variance estimations are affected by size due to sampling fraction, and proximity to zero.

5.5.1.7.1 Distribution according to occupant position and location of impact

As a first example, we'll try to obtain the distribution of the severity according to the position of the person in the car and the location of the impact, with a specific attention to side-impacts.
Hence, this implies selecting accidents with at least one casualty occurring in a car or car derivative. Unlike the English example (because our numbers are very much lower), we chose not to select according to the age of vehicles.
Following these criteria for 1996, 2604 crashes involving 3716 casualties have been recorded by the police inside the four investigated areas.
The number of In-depth documented accidents is 170, with 231 casualties.

With the empirical sampling design, we know that there is no built in bias. However, as the sampling design is not random, we can suspect that for some dimensions, distributions are quite different.
Generally speaking, the choice of relevant weighting factors is closely dependant on the question to be solved, but we can as a priority examine some main factors, which are connected with many parameters.
As possible biases, we'll consider on the one hand, hour of day and day of week, and on the other hand, the road class, and size of the built-up area.
Because we are interested in the distribution of the severity according to place in the vehicle and location of the impact, it is also important to look at the distribution of the severity.

The distributions of these variables are compared, between the In-depth studies and the local database, using a Chi-squared test (taking the Regional as the reference). To satisfy the independence assumption, the statistical unit is the accident.

Table 4.1.1. Severity of casualty - accidents for Region/In-depth

Database	percentage in database			total count
	fatal	serious	slight	
Regional	0.03	0.21	0.76	2604
In-depth	0.08	0.15	0.77	170

The Chi-squared test (χ^2=14.5, with 2 degrees of freedom) shows that the distributions are different. Let us note that the proportion of fatal accidents is higher in the In-depth database. However the proportion of serious crashes is lower in the In-depth database.

Concerning the time of accident and the month of accident, there is no significant difference between the two distributions.

Concerning the road class and size of built-up area, the In-depth database distributions are different from that of local databases. For the purpose of this example, we choose to use road class because this variable is better defined than the other.

Table 4.1.2. Road type - accidents for Region/In-depth

Database	percentage in database				total count
	motorway	main road	secondary road	other	
Regional	0.03	0.15	0.14	0.68	2604
In-depth	0.03	0.24	0.37	0.36	170

The two distributions are different (χ^2=105, with 3 degrees of freedom). The proportion of accidents in built-up areas is much lower in the In-depth database.

5.5.1.7.2 Estimates of number of casualties by seating position

As the numbers are quite low in the In-depth database, the motorway, main and secondary roads have been brought together to form two distinct categories.
Calculations are made on 208 casualties without missing values for the variables of interest.

Table 4.1.3 Proportion of casualties by seating position

Casualties by severity and speed limit	Proportion in in-depth severity of casualty			Estimated numbers in Regional database		
	fatal	serious	slight	fatal	serious	slight
motorway, major and secondary roads						
side and site of impact	0.36	0.00	0.02	**27**	**0**	**25**
side not site	0.14	0.07	0.07	**11**	**17**	**76**
not side of impact	0.07	0.14	0.09	**5**	**34**	**101**
not side impact	0.43	0.79	0.81	**32**	**186**	**885**
total count	14	14	86	75	237	1087
others						
side and site of impact	0.00	0.14	0.08	**0**	**62**	**152**
side not site	0.00	0.14	0.08	**0**	**62**	**152**
not side of impact	0.00	0.14	0.16	**0**	**62**	**304**
not side impact	1.00	0.57	0.67	**15**	**249**	**1258**
total count	1	7	86	15	436	1865

Numbers in Bold are calculated by multiplying the proportions in the In-depth data by the Regional marginal counts of the Regional data.

The table 4.1.4 shows the total numbers of casualties by seating position and location of impact, after taking into account the road type.

Table 4.1.4. Estimates at Regional level - using severity and road type linking

estimate and (se)	fatal	serious	slight	total
side and site of impact	27 (6)	62 (18)	177 (52)	266
side not site	11 (8)	79 (24)	228 (50)	318
not side of impact	5 (9)	96 (22)	405 (47)	506
not side impact	47 (5)	435 (10)	2143 (15)	2625
total	90	673	2952	3715

The Variance estimations are calculated according to annex A. As these calculations are based on the normal assumption, we can see that this hypothesis is hardly verified for small counts.
Otherwise, we obtain 10% of fatalities when side and site of impact, and 24% of serious injuries, as we have 3% and 25% for side not site, and 2% and 17% for not side impact. These proportions are not surprising as we expected that the proportion of fatalities would be highest when the car occupant is on the side and on the site of the impact.

The table 4.1.5 shows the estimations of Regional numbers calculated from the distribution on the In-depth level, but without taking into account the road type. We can see that the estimates are quite different from the previous ones, which was expected as the χ^2 was significant for the road class distribution.

Table 4.1.5. Estimates at Regional level - using severity linking (but not road type)

estimate and (se)	fatal	serious	slight	total
side and site of impact	30	32	162	224
side not site	12	64	216	292
not side of impact	6	96	378	480
not side impact	42	481	2196	2719
total	90	673	2952	3715

5.5.1.8 Evolution

Data collection at Lyon was discontinued on 1st January 1998. The investigation area of Salon de Provence has been extended from this date. This will allow the collection of about 250 accidents in each of the three remaining areas. Another objective of this change was to enable the concentration the Lyon team on secondary safety studies which are based on the registry of road accident casualties in the “département du Rhône”.

5.5.2 Germany

5.5.2.1 Accidents of interest

The aim of the German in-depth investigation is to build up an in-depth accident database that can be used for analyses of nearly all categories of traffic accidents with at least slightly injured persons. So it is intended to have a database that allows analyses of accidents with cars or lorries as well as accidents with vulnerable road users such as bicyclists or pedestrians. Therefore the project is neither limited to specific research questions nor to definite days.
The database should contribute to give answers to the following two questions:

- analytical questions as, for instance, biomechanical associations
- descriptive questions as, for instance, percentages of specific types of accidents

The database is used to examine traffic accidents in a much more extensive way than it would be possible with the National accident database.

5.5.2.2 Accident population (of database)

The basis for the sample are all police-reported traffic accidents with at least slightly injured persons. As described above this procedure includes accidents with cars as well as with injured pedestrians or bicyclists. There is no instruction to prefer a special type of accidents. This results in a database that includes nearly all types of accidents even though some of them are very rare.

5.5.2.3 Areas

The German in-depth investigation has been carried out in urban and rural regions of Hanover (21 communities and counties). The regions cover about 2300 km^2 with nearly 1.3M. inhabitants. The decision in favour of Hanover was based on accident and traffic structural aspects as well as on reasons of availability and accessibility of detailed hospital information and optimal co-operation with the police. In order to realise an on-scene investigation it is indispensable to have a narrow co-operation with accident hospitals (to get data of the injured persons), the police, the fire brigade and so on. It is essential to have some privileges such as flashing blue light to be immediately at the scene of the accident.
At the moment the investigation is limited to the regions of Hanover. Generally, the decision in favour of the investigation area is essential for questions concerning the linkage to National accident databases.

5.5.2.4 Data selection modus

The data selection modus is on scene. This means that for the most part of the information will be collected at the accident location immediately after the accident has happened. But some details will also be ascertained in a retrospective way (in hospitals, garages etc.).
The decision in favour of the data selection in an on-scene modus was based on the fact that some types of accident information can only be ascertained in an adequate manner when being collected right at the accident location immediately after the event. This refers particular to characteristics of the road users' relevant behaviour as well as their present physical and psychological condition. Furthermore the application of this method is necessary to examine skidmarks, the state of the crashed vehicles, the conditions of the environment (weather conditions etc.) etc. All these characteristics are helpful to determine, for instance, the speed collision or to evaluate injuries.
In addition to the described advantages, the decision in favour of the on scene selection modus has some consequences for the complete accident selection procedure. A special accident selection design was developed in order to be able to meet the requirements of a random sample (Deming 19xx, Hautzinger 19xx).

5.5.2.5 Data selection design

Due to the decision in favour of on-scene investigations and due to financial and organisational reasons the investigation team is not able to examine all accidents of interest in the area. So the data selection would be divided into two steps.

(1) time interval selection:
During 12 hours a day (0-6 and 12-18 or 6-12 and 18-24) accidents will be investigated on scene. There is a weekly change of the two investigation time intervals.

(2) accident selection:
The principle is to register the data of the last reported accident after the re-establishment of the team's readiness for action. This selection step will be described in detail later on.

The purpose of this design is to guarantee that the selection process will not be influenced by subjective evaluations of the investigation team or police officers and thus to prevent systematic biases.

5.5.2.6 Principle of accident selection

Within the selected time intervals the accident selection is oriented on the following principle: It is planned to drive to the first accident in the investigation time interval. After that the team has either to continue the investigation with the last accident registered by the team co-ordinator or it has to wait until it is informed about the next accident. The aim of this instruction is to avoid influences of the team or other persons on the accident selection process.

5.5.2.7 Practical aspects with examples

Test of systematic biases
Results of random samples are only really meaningful when it is known to what extent the sample reflects the accident population in the investigation area. For this reason it is necessary to examine the sample for systematic biases and to calculate weighting factors for every investigation year.

If using the above described method, systematic biases can appear on both selection steps: on the step of time selection or on the step of accident selection. In order to investigate the occurrence of such biases, distributions of the in-depth data have to be compared with the accident statistic of the police for the region. The following analyses rest on data collected 1995.

- *Systematic biases caused by time selection modus*

Data gathering takes place 12 hours a day (2 six-hour intervals). If there are no systematic biases resulting from the selection of the time intervals, it has to be expected that - summarised over a year - half of all accidents in the region of Hanover registered by the police will happen in the time intervals of data gathering.

Table 4.2.1 Accidents registered 1995 by the police in the region of Hanover

	Injury severity						Total	
	Slightly injured		Seriously injured		Fatally injured			
	n	%	n	%	n	%	n	%
During investigation time	2198	48.5	379	47.4	31	49.2	2608	48.3
Not during investigation time	2338	51.5	421	52.6	32	50.8	2791	51.7
Total	4536	100.0	800	100.0	63	100.0	5399	100.0

Table 4.2.1 shows that in 1995 2608 (48.3%) of all 5399 accidents with injured persons (region of Hanover) happened during the data selection intervals. This corresponds to the expectation of 50% for non-systematic bias. A distinction for injury severity shows the same results. So it can be summarised that there will be no biases caused by the selection of the time intervals.

- *Systematic biases caused by accident selection modus*

In contrary to the time selection modus deviations can be shown for the second level (accident selection modus). Such deviations can be a result of the applied data collection procedure.

(a) Time related characteristics

The first characteristic under examination is the time of accident (table 4.2.2).

Table 4.2.2 Distribution of accident time in 1995

Time of accident	Region		In-depth	
	n	%	n	%
0 - 6	305	5.7	59	6.7
6 - 12	1548	28.7	273	30.9
12 - 18	2389	44.2	368	41.7
18 - 24	1157	21.4	183	20.7
Total	5399	100.0	883	100.0

Table 4.2.2 demonstrates that there are visible differences between the in-depth data and the regional accident statistic. The period of time between 0-12 a.m. is over-represented. In contrast, accidents between 12 and 24 are underrepresented.

(b) Injury severity

One of the most important characteristics of the in-depth data is the injury severity. A comparison of the in-depth data to regional accident statistics shows significant differences (table 4.2.3).

Table 4.2.3 Distribution of the injury severity in 1995

Injury severity	Region		In-depth	
	n	%	n	%
Slightly injured	4536	84.0	553	62.6
Seriously injured	800	14.8	299	33.9
Fatally injured	63	1.2	31	3.5
Total	5399	100.0	883	100.0

Table 4.2.3 demonstrates significant deviations between regional accident statistics and the in-depth data: While in the regional area 84% of the accidents are accidents with slightly injured persons, the percentage in the in-depth data is with 62.6% on a significantly lower level. In contrast to this the accidents in the in-depth data with seriously injured or killed persons were over-represented. Thus the in-depth data have been weighted by injury severity in order to get representative data for the region of Hanover.

(c) Accident location

The third important characteristic to be examined is the accident location. A consideration of the distributions shows similar differences between the regional statistic and the in-depth data (table 4.2.4).

Table 4.2.4 Distribution of accident location in 1995

Accident location	Region		In-depth	
	n	%	n	%
Inside urban area	4109	76.1	592	67.0
Outside urban area	1290	23.9	291	33.0
Total	5399	100.0	883	100.0

In 1995 three quarters (76.1%) of all accidents registered by the police occurred inside urban area. In contrast to this the in-depth data contain only 67.0%. This means that accidents which happened inside urban area were underrepresented, and this result should also be taken into consideration for the weighting procedure.

- *Reasons for the systematic biases*

As shown above, systematic biases of the in-depth data do not occur in the context of the time selection modus but in context of accident selection modus.

The most important reason lies in the information practice of the investigation team. In 1995 the investigation team was informed by the police about only one third of all accidents which happened during investigation time. Non-reported accidents are mainly accidents with slightly injured persons. Obviously the police inform the investigation team mainly about severe accidents.

Furthermore during the night shift, when only few accidents happen, nearly all known accidents can be documented. During this time accidents are characterised by at least seriously injured persons. This fact is another reason for the overrepresentation of accidents with seriously injured or killed persons. In the case of the shifts during rush hour, it is vice versa: During this time most of the accidents happen so that the documentation rate is lowest. Moreover this time is characterised by accidents with slightly injured persons.

- *Calculation of the weighting factors*

The above considerations demonstrate the necessity of weighting the in-depth data in order to be representative for the region of Hanover. For methodical reasons the following characteristics will form the basis for the weighting procedure:

(a)	time of accident	(0 - 6, 6 - 12, 12 - 18, 18 - 24)
(b)	injury severity	(slightly injured, seriously injured, killed)
(c)	location of accident	(inside urban area, outside urban area)

The weighting factors were calculated by using the following formula:

$$w_{ijk} = \frac{N_{ijk} / N}{n_{ijk} / n} = \frac{N_{ijk}}{n_{ijk}} \cdot \frac{n}{N}$$

w_{ijk} : weighting factor
i : time of accident
j : injury severity
k : location of accident
N : number of all regional accidents
N_{ijk} : number of all regional accidents for ijk
n : number of all documented accidents
n_{ijk} : number of all documented accidents for ijk

Normally accidents with slightly injured persons are underrepresented and are weighted by factors greater than 1. In contrast, accidents with seriously or fatally injured persons are over-represented and have to be weighted by factors less than 1. At this point it has to be emphasised that the weighting factors have to be renewed for every investigation year.

- *Goodness of adjustment through weighting the data*

The usage of the weighting factors has consequences for the analyses of the data: It leads to more adjusted distributions of the variables which are included in the weights. The aim of weighting, however, also is to get a better correspondence of other variables between the in-depth data and the regional statistic. The following measure (similar to a chi-square statistic) will be used to assess the goodness of adjustment through weighting:

$$C = n \sum_{m=1}^{M} \frac{(p_m - \theta_m)^2}{\theta_m}$$

p_m : relative frequency of value m of a variable in the in-depth data
θ_m : relative frequency of value m of a variable in the regional statistic
M : total number of values of a variable
m : specific value of a variable

In this context C is more considered as a descriptive measure for the assessment than a test statistic. The closer this measure to zero the better the correspondence between the sample and the total population (the region).

The following example illustrates the usage of this measure. Table 4.2.5 shows a comparison of analyses with weighted and unweighted data concerning the accident day (weekdays vs. weekend).

Table 4.2.5 Distribution of accident day in 1995

Days		Region	In-depth	
			Weighted	Unweighted
		%	%	%
Mo		16.1	14.2	13.7
Tu + We		31.1	30.2	30.0
Th + Fr		31.4	30.6	30.6
Sa + Su		21.4	25.0	25.7
Total	%	100.0	100.0	100.0
	n	5399	883	883
C			7.66	11.04

A first view shows that the percentages of the weighted data are closer to the regional statistics than those of the unweighted data. For example, the percentage of accidents which happened during the weekend (Sa - Su) decreases from 25.7% (unweighted) to 25.0% (weighted). In accordance to that the C value decreases from 11.04 (unweighted data) to 7.66 (weighted data). So the table demonstrates that the chosen weighting procedure leads to a visible improvement of the distributions.

- *How good is the sample design?*

Finally the goodness of the sample procedure has to be considered. To get an assessment of the selected sample design, the concept of replicated samples will be applied. Therefore the total sample will be divided into k=3 independent sub-samples (replications). All accidents which occur during the first, the fourth, the seventh day and so on are the first sub-sample. In the same manner the second, the fifth, the eighth day and so on are the second sub-sample. Finally the remaining days are the third sub-sample. By using this procedure the three sub-samples contain nearly the same sample size. Then for every sub-sample the weighting factors will be calculated. On the basis of the separated weighted sub-samples, the distribution of an accident variable will be analysed, e.g. the injury severity MAIS. Table 4.2.6 contains the distributions of the MAIS for the three sub-samples.

Table 4.2.6 Comparison between the distributions of the three replications concerning MAIS (1995 - weighted data)

MAIS		Replications			Total mean
		1	2	3	
0		1.8	2.1	2.0	2.0
1		83.8	80.5	82.5	82.2
2		10.4	13.7	10.1	11.4
3-4		3.6	2.9	3.7	3.4
5-6		0.4	0.8	1.8	1.0
Total	%	100.0	100.0	100.0	100.0
	n	394	347	406	1147

On the basis of the percentage of MAIS 1 the calculation of the confidence intervals can be demonstrated.
In this case, there are three estimates of MAIS 1:

$Y_1 = 83.3 \qquad Y_2 = 80.5 \qquad Y_3 = 82.5$

In order to calculate the total mean of the three estimates, the following formula has to be used

$$\overline{Y} = \frac{1}{k}\sum_{k=1}^{K} Y_k$$

$$= \frac{1}{3}\cdot\left(83.8 + 80.5 + 82.5\right) = 82.2$$

$\overline{Y}$: total mean of the three estimates
Y_k : estimates for every replication
k : number of replications

The variance of $\overline{Y}$ will be estimated using the formula

$$\operatorname{var}(\overline{Y}) = \frac{1}{k(k-1)} \sum_{k=1}^{K} \left(Y_k - \overline{Y}\right)^2$$

$$= \frac{1}{3 \cdot 2} \cdot \left((83.8 - 82.2)^2 + (80.5 - 82.2)^2 + (82.5 - 82.2)^2 \right) = 0.92$$

The standard deviation of the total mean is $\sqrt{0.92} = 0.96$. By using a significance level of 5% the confidence interval is

$82.2 \pm 196 \cdot 0.96$.

So the unknown percentage of persons injured with MAIS 1 is between 80.3% and 84.1.
In the same way the remaining confidence intervals can be calculated; the results are contained in table 4.2.7.

Table 4.2.7 Estimates, standard deviations and confidence intervals (1995 - weighted data)

MAIS	estimates (in %)	standard deviation	confidence interval
0	2.0	0.09	1.8 to 2.2
1	82.2	0.96	80.3 to 84.1
2	11.4	1.15	9.1 to 13.7
3-4	3.4	0.62	2.2 to 4.6
5-6	1.0	1.02	(-1.0) to 3.0
Total %	100.0		

Table 4.2.7 shows that the confidence intervals have only a small range. Therefore the three replications produce nearly the same estimates.
So it is demonstrated that the chosen sample procedure is an appropriate way for gathering the in-depth data.

5.5.2.8 Evolution

The German in-depth investigation of the Accident Research Unit Hanover has been carried out since 1973 on behalf of the Federal Highway Research Institute (BASt). Due to the objective of a generalisation of the results for the region of Hanover a statistical investigation plan has been developed and has been in progress since 1985. Currently it is planned to extend the investigation for probably two further regions within the next years. The aim is to get a representative sample for Germany and to have more accidents for analysing.

5.5.3 United Kingdom

5.5.3.1 Accident of interest

Within the UK in-depth accident database - CCIS (Co-operative Crash Injury Study), a record is only created for accidents reported to the Police in which at least one car or car derivative (hereafter car is also taken to include a car derivative) was involved which was aged less than 7 years old and had to be towed away. The Police would usually only be informed about accidents where there was at least one injury and hence the CCIS database will usually only contain injury accidents.

5.5.3.2 Accident population (of database)

The accident population includes accidents are were reported via the Police to either one of the two University based CCIS collection centres or to one of the five vehicle inspectorate teams who collect the tow-a-way vehicle information. In practice this includes accident data from 9 Police areas. Only accidents involving cars where at least one car was less than 7 years old are of interest. (Earlier years of the CCIS data also included some accidents where there was damage only, i.e. no injuries but still a tow-a-way).

The UK National database - Stats19, depends upon there being an accident where the Police attended and where there was at least one casualty.

5.5.3.3 Areas

The areas for in-depth accident investigation were partly chosen because of the location within the UK of the University based investigation centres, and partly because of the location of the Vehicle Inspectorate teams who are contributing. The Universities are based in Loughborough and in Birmingham, and are able to examine a far higher proportion of eligible accidents than the VI teams. This can be clearly seen in the following table where the University of Loughborough team examine nearly 1/5 of all accidents and the University of Birmingham about 1/9 accidents.

Table 4.3.1 Accidents in 1995 CCIS and UK National database where the accident met the CCIS definition for inclusion

Police Area code	Area	Stats19	CCIS counts & (%)
06	Greater Manchester	3164	73 (2.3%)
20	West Midlands	2291	258 (11.3%)
21	Staffordshire	1422	50 (3.5%)
22	West Mercia	1198	70 (5.8%)
23	Warwickshire	649	21 (3.2%)
31	Nottinghamshire	1023	189 (18.5%)
33	Leicestershire	862	176 (20.4%)
44	Hampshire	1718	98 (5.7%)
52	Avon & Somerset	1080	71 (6.6%)
-	CCIS area	13407	1006 (7.5%)
-	all of the UK	56708	1006 (1.8%)

The sampling methodologies for each of the areas are similar but due to resource differences then the VI teams are not able to investigate so many accidents. The University of Loughborough team investigate nearly twice the number of accidents as the University of Birmingham team, hence it is perhaps most useful to consider three areas, representing the VI teams, the Loughborough and the Birmingham teams.

Table 4.3.2 Accidents in 1995 CCIS and Regions

Police Area code	Area	Stats19	CCIS counts & (%)
31 & 33	Loughborough	1885	365 (19.4)
20	Birmingham	2291	258 (11.3%)

others	VI teams	9231	483 (5.2%)

5.5.3.4 Data selection modus

Nearly 100% of fatal accidents are followed-up, but a lower proportion of accidents involving lesser injury are studied. Fatal accidents represent about 30% of the database, the remainder being accidents where there was a severe or slight injury. Investigation teams examine the vehicle(s) to where they were towed-a-way.

5.5.3.5 Data selection design

Selection of accidents to investigate is made using a stratification procedure. Accidents are stratified by severity of worst casualty, most fatal accidents are selected for investigation, as many serious as possible and those slight accidents which can be inspected are randomly selected according to the available resources. A consequence of this is that slight injury accidents are under represented within the database compared to fatal or serious accidents. There is no systematic bias in the selection process.

5.5.3.6 Practical aspects with examples

In the UK the sampling method applied is a stratified design, with severity of the worst casualty being the stratification variable. In this way a higher proportion of fatal and severe casualty accidents are included relative to slight injury accidents. This results in a more efficient sampling strategy for comparing accident severity’s, but it does mean that any results for the whole sample must be weighted to reflect the actual proportions of fatal/severe/slight accidents in the population.

The example will first examine the representativeness at Regional level, investigate which linking variable(s) to use for weighting purposes and to determine the adequacy of the weighting to produce Regional estimates. The VI areas and University of Loughborough area will be considered separately, these have roughly 5% or 20% sampling proportions of accidents respectively.

As an example let us suppose that an estimate is required at Regional level of the **number** of occupants with different injury severity’s related to seating position in a car (or car derivative), in a side-impact. The main objective is not the actual estimate derived, nor the associated precision, not even the linking variables used but to demonstrate a linking methodology.

The number of occupants at Regional level are estimated using weighting variables. These estimates can be compared with the number of side impacts v non-side impacts are known at Regional level. The standard errors for all estimates have been calculated using the approach as described in the attached annex.

A number of variables have been identified and can be compared using the Chi-squared statistic. The variables in common between the in-depth accident database and the Regional databases are:

- severity of casualty - according to the Police estimate,
- speed limit of the road on which the accident happened,
- road class,
- time of the day - by each hour and grouped into 6 hour periods,
- month of the year.

The distributions of severity and speed limit are compared (for this example), using a Chi-squared test. The distributions and conclusions are given below, (the sample size totals vary due to missing values):

Table 4.3.3. Accidents by severity of worst casualty - counts (%s) and estimated values - Loughborough area

accident severity	fatal	serious	slight	total
Regional - counts	12 (0.6%)	138 (7.2%)	1775 (92.2%)	1925 (100%)
In-depth - observed	12 (3.9%)	77 (25.0%)	219 (71.1%)	308 (100%)
In-depth - expected	1.9	22.1	284.0	308

Table 4.3.4. Accidents by severity of worst casualty - counts (%s) and estimated values - VI areas

accident severity	fatal	serious	slight	total
Regional - counts	44 (0.5%)	551 (5.9%)	8768 (93.6%)	9363 (100%)
In-depth - observed	16 (4.8%)	82 (24.8%)	233 (70.4%)	331 (100%)
In-depth - expected	1.6	19.5	310.0	331

The Chi-squared statistics (χ^2=204.4 for Loughborough area and χ^2=353.9 for the VI areas, both on 2 degrees of freedom), confirmed that the in-depth database distribution of casualty severity was **different** from that of the Regional databases and hence the variable should be used for weighting purposes.

Table 4.3.5. Accidents (%) by speed limit (includes fatal/serious/slight)

	speed limit in mph					
Database	<=30	40	50	60	70	total count
Loughborough - local	47.2%	9.7%	2.1%	30.3%	10.2%	1925
In-depth	27.6%	10.4%	2.7%	50.5%	8.8%	297
VIs - local	51.8%	8.8%	2.3%	25.4%	11.7%	9363
In-depth	25.9%	6.7%	5.5%	47.3%	14.6%	328

The Chi-squared statistics confirmed that the in-depth database distributions of speed limit was **different** from that of the Regional databases and hence should also be used for weighting purposes. (χ^2= 64.9 and 123.1 respectively, for Loughborough and VIs on 4 degrees of freedom).

The in-depth database distribution of accidents is clearly quite different from the Regional database distribution across the linking variables analysed. This is not too surprising because the sampling methodology for the in-depth database focuses more on the fatal and serious injury accidents. Further analysis of these linking variables within accident severity was then conducted.

Table 4.3.6 . Accidents (%) by speed limit and most severe injury – Loughborough

	Regional			In-depth counts		
Speed	fatal	serious	slight	fatal	serious	slight
≤ 40 mph	8.3%	41.3%	58.4%	8.3%	25.0%	44.8%
50/60 mph	75.0%	45.7%	31.7%	75.0%	68.4%	46.2%
70 mph	16.6%	13.0%	10.0%	16.6%	6.6%	9.0%
total	12	138	1775	12	76	210
Chi-squared within accident severity level				0.0	16.0	20.8

Table 4.3.7 . Accidents (%) by speed limit and most severe injury – VIs

Speed	Regional			In-depth counts		
	fatal	serious	slight	fatal	serious	slight
≤ 40 mph	20.4%	31.4%	62.7%	40.0%	18.5%	37.1%
50/60 mph	61.4%	53.5%	25.9%	53.3%	65.4%	48.3%
70 mph	18.2%	15.1%	11.4%	6.7%	16.1%	14.6%
total	44	551	8768	15	81	232
Chi-squared within accident severity level				4.05	6.5	71.1

The distribution of fatal accidents is very similar (i.e. not statistically significantly different), between the Regional and in-depth data, which is to be expected given the sampling methodology. However the serious and slight accidents are distributed differently between Regional and in-depth data for both areas.

The analysis of these linking variables shows that they should be used for weighting purposes when estimating from In-depth to Regional statistics. It is very clear that the severity of occupant injury is an important variable, this is because it was known a priori that the sampling methods for the in-depth database were biased towards getting fatal and severe injury accidents.

For the purpose of this example only these linking variables have been used and in a two-way combination of speed limit by severity. The data from the in-depth database for the proportion of casualties in each seating position by casualty severity and speed limit is given in table 4.3.8.

Table 4.3.8 Proportion of casualties by seating position

Casualties by severity and speed limit	Proportion in Loughborough severity of casualty			Proportion in VIs severity of casualty		
	fatal	serious	slight	fatal	serious	slight
≤40 mph						
side and site of impact	0.00	0.16	0.12	0.43	0.04	0.07
side not site	0.00	0.15	0.06	0.00	0.08	0.04
not side of impact	0.00	0.18	0.21	0.00	0.19	0.10
not side impact	1.00	0.52	0.61	0.57	0.69	0.79
total count	2	62	164	7	26	144
50/60 mph						
side and site of impact	0.09	0.06	0.07	0.04	0.06	0.07
side not site	0.04	0.02	0.02	0.04	0.05	0.06
not side of impact	0.13	0.14	0.17	0.17	0.15	0.10
not side impact	0.74	0.78	0.74	0.74	0.74	0.77
total count	23	121	167	23	101	189
70 mph						
side and site of impact	0.17	0.25	0.06	0.25	0.10	0.08
side not site	0.00	0.05	0.17	0.00	0.14	0.05
not side of impact	0.00	0.20	0.28	0.25	0.19	0.17
not side impact	0.83	0.50	0.50	0.50	0.57	0.70
total count	6	20	36	4	21	60

Four seating positions were defined, albeit one was not a side impact. The side of impact and if the seating position was at the site of impact is the first combination, the next being for the side of impact but not the site of impact, the third was if the impact was not on the side the occupant was seated and the final code was for non-side impacts.

The estimated number of casualties in the Regional databases by seating position are not known, hence the need to derive figures from the in-depth data. However, it is known if damage was done to the side of the vehicle and hence it is possible to derive a variable which can be used to give a reasonable check on the accuracy of the estimated values using the linking variables as discussed. This variable will not be exactly equivalent to that being estimated but provides a useful check.

The number of casualties by seating position and the associated standard error of the estimate are given in Table 4.3.9a and Table 4.3.9b.

Table 4.3.9a. Estimates for Loughborough area - using severity and speed limit linking

estimate and (se)	fatal	serious	slight	total
side and site of impact	3.4 (0.7)	38.7 (5.2)	223.9 (35.9)	266
side not site	1.0 (0.2)	23.2 (4.1)	145.8 (29.5)	170
not side of impact	3.1 (0.3)	50.3 (5.3)	498.8 (49.2)	552
not side impact	26.4 (0.8)	195.8 (6.8)	1533.4 (57.9)	1756
total	34	308	2402	2744

Table 4.3.9b. Estimates for VI areas - using severity and speed limit linking

estimate and (se)	fatal	serious	slight	total
side and site of impact	14.3 (4.9)	62.9 (19.9)	834.5 (169.1)	912
side not site	3.1 (2.5)	7.2 (23.3)	570.2 (136.5)	650
not side of impact	17.1 (5.9)	182.9 (34.0)	1234.6 (199.6)	1435
not side impact	70.5 (7.1)	766.0 (40.8)	9128.7 (272.0)	9965
total	105	1089	11768	12962

The estimated numbers of side impacts and non-side impacts are compared to the side damaged (yes/no) variable derived from actual Regional data. These are given in Table 4.3.10a and Table 4.3.10b. This shows the estimated number of side impacts by casualty severity level, and indicates quite close agreement.

Table 4.3.10a Estimated and actual numbers of side impact casualties - using severity and speed limit linking - Loughborough local area

National		fatal	serious	slight	total
estimated (se)	side impact	7.6 (0.8)	112.2 (8.5)	868.6 (67.7)	988
	not-side	26.4 (0.8)	195.8 (6.8)	1533.4 (57.9)	1756
actual	side impact	10	110	778	898
	not-side	24	198	1624	1846

Table 4.3.10b Estimated and actual numbers of side impact casualties - using severity and speed limit linking - VI local areas

National		fatal	serious	slight	total
estimated (se)	side impact	34.5 (8.0)	323.0 (45.8)	2639.3 (295.1)	2997
	not-side	70.5 (7.1)	766.0 (40.8)	9128.7 (272.0)	9965
actual	side impact	26	295	3122	3443
	not-side	79	794	8646	9519

The application of a linking methodology to link an in-depth accident database with a Regional database has been illustrated via a lengthy example. The methodology is not difficult and is represented algebraically in annex A. It relies on finding variables in common between the in-depth database and the Regional database and to use these variables to weight in-depth database variables to a Regional level. The conclusion reached for the example is that most estimates are within 2 standard errors of the actual values and so the estimation procedures seem to work fairly well. However, the precision of the estimates, as shown by the size of the standard error, is better with the higher sampling rate from the Loughborough area.

5.5.3.7 Evolution

The UK in-depth database has evolved over time. This has been a process of trying out slightly different accident selection approaches or changing the depth of information gathered. For example, some damage only accidents have been included and the benefits in subsequent analysis were investigated. This idea was dropped, in-part because of the resource implications but also because the prime objective within the UK in-depth accident database is to investigate the damage done.

The depth of information gathered does affect the number of investigations that can be made. With less information then more cases can be investigated but are more cases with less information better than fewer cases with less complete data? The current consideration is that it is of greater value to have more in-depth information at the expense of more cases.

5.5.4 Summary

The questions to be answered using in-depth database data also need to be considered carefully. If the in-depth data are not representative of the National database then this may be a result of the sampling methodology. Hence only questions relating to the sampled sub-set of accidents can be answered.

5.5.4.1 Limits - France

The French In-depth database is built from four quite different areas. The Lyon area is essentially urban, contrary to the three others and the sampling fractions are quite different between the four areas, with a quite low one for Lyon. This partly explains the bias is high on the road type.

The limits of this approach are quite clear, as the sampling design is not random, biases exist. The first point is to identify these potential biases. The second point is to take them into account if they can lead to confusion. Of course, the number of factors that can be taken into account cannot be high. In practice, if the number of observations is low, the number of adjusting factors must be low too. In our data, for the asked question (just cars in 1996), it is not possible to adjust more than 2 factors.
Another problem with a low number of observations, is that 0 cells are not uncommon. This is specially true with factors with many values or with unbalanced proportions for the different values (as it is the case between side and not-side impacts).
Finally, variance estimates are highly dependant of the rate N^2/n. In our example, the size n of the In-depth database is very low compared with the size N of Regional database. Therefore the confidence intervals are quite large. In an extreme case, when one estimate in one cell is close to zero, the normal assumption which has been used becomes questionable.

5.5.4.2 Limits - Germany

For reasons of the typical regional structure and the topographic situation of the Hanover region it has to be investigated to what extent a linkage can be drawn between the in-depth data and the German National accident statistic. As shown above, a representativity for the region of Hanover seems to be given. Therefore the generalisation of the in-depth data seems to be unproblematic for questions concerning medical-technical and biomechanical correlation.
But a generalisation to National level will be much more problematic if the specific situation of the region of Hanover is important for the research question. For example, to determine the speed on motorways in Germany and the resulting injury severity in dependence of the collision speed, a transference is not possible for reasons of the typical situation in the area of Hanover. The general distribution of the speed on motorways is influenced by many factors (volume of traffic, speed limits, areas of roadworks etc.) and there seem to be important differences between the region of Hanover and Germany in total. Furthermore some accident types (e.g. accidents in mountains) do not occur for e.g. topographic reasons. Thus the in-depth data are not easily used for a generalisation to National level.

5.5.4.3 Limits - UK

The UK in-depth database has a built in bias because it only contains data from accidents where there was at least one car which was less than 7 years old and had sufficient damage to result in being towed away from the scene of the accident. The UK National database also has biases because it only relates to those accidents where there is a casualty, hence damage only accidents are not included. There will be damage only vehicles on the National database because they were involved in an accident situation where one of the other vehicles did contain a casualty. These biases can probably not be fully adjusted for because one sector of the population is under represented, and hence there may not be as much information in the in-depth database for some accident types.

5.5.4.4 General

Sampling strategies vary from participating country to participating country, the detail of the strategies is reported more fully elsewhere. However, the sampling strategy affects what can be estimated from the In-depth database. Different strategies result in different constraints and assumptions on the estimation procedures, see Table 4.4.1.

None of the sampling strategies are right or wrong, they were designed for different purposes. Knowledge of the three sampling strategies provides a useful basis for further discussions.

Sample size also plays an important role for a linkage with Regional or National Accident database. The lower the sample size the more imprecise the results.
The following points are worth noting:

- accidents in In-depth databases are selected due to the sampling methods employed,
- a broad based accident database may be advantageous in being able to answer a broad set of questions,
- a stratified sample may have some advantages for answering certain types of question,
- the ideal design depends on the question being asked and the estimate required,
- the ideal design for one estimate will in general be different for another,
- the sampling designs in use for In-depth databases are not ideal and if they are to be used for linking to Regional or National databases, then the assumptions and limitations must be clearly stated.

Table 4.4.1. Sampling Strategies

	France	Germany	UK
Basis	All accidents with at least slightly injured persons	All accidents registered by the police with at least slightly injured persons	All accidents to car or car derivatives reported via the police and where at least one vehicle was towed-a-way which was ≤7 years old
Region(s)	Four areas throughout France	Urban and rural region of Hanover	Seven Police areas throughout England
Data collection	On Scene	On scene	Investigation teams examine the vehicle(s) at where they were towed-a-way
Data selection design	Different sampling strategies in the four areas, but all Empirical with time intervals	Random sample design (two-step design: time interval selection, accident selection)	Stratified by severity of worst casualty, accidents selected at random with a higher proportion of fatal and severe accidents
Principle of accident selection	Investigation team drives to the first accident in the time interval. After that the team has to wait until it is informed about the next accident or it conducts an investigation of the last accident which occurred during the previous one.	The investigation team drives to the first accident in the investigation interval. After that the team either continues the investigation with the last accident registered by the team co-ordinator or it has to wait until it is informed about the next accident.	Investigation team is informed of the accident via the police and attend selected accidents at the site to where the vehicle was towed at a later date.
Number of documented accidents	219	883	1006
Number of accidents in the Regional area during investigation time	n/a	2608	13407
Number of accidents in the Regional areas	3694	5399	13407
Ratio of all documented to Regional & National accidents	1:17 1:573	1:6 1:439	1:13 1:56
Number of National accidents which met inclusion criteria	125406	388003	56708

* The data for France are from 1996 and for Germany and the UK from 1995

5.6 Linkages possible today

Linkage between In-depth and Regional accident databases has been demonstrated in the previous section for France, Germany and the UK using the methodology as described. In all three countries it is necessary to check the adequacy of the linking and weighting process, this has been illustrated in the examples in the previous section.

Linkages beyond Regional levels may theoretically be possible. However, this has not yet been done and it would first be necessary to examine the differences between the Regional and National structure taking into account the following factors:

- accident structure,
- accident-risk structure,
- traffic infrastructure / environment,
- vehicle structure with regard to motor manufacturer and car categories,
- geographical, climatic and topographical consideration,
- structure of settlements.

On the basis of these analyses it should be possible to identify the constraints and to develop an appropriate linkage for each Country. Such analyses will also be required if a linkage to a pan-European level is to be achieved. This is further complicated by the fact that different countries use different coding schemes. The CARE+ project is providing a mechanism to facilitate the comparison of accident statistics between member states using data from the member states disaggregated data. It would be helpful for STAIRS to have a mechanism for data exchange across Europe using a common coding scheme, this would provide a framework for possible linkages.

Nevertheless, it is still feasible to investigate relationships between accidents and potential causation variables, the nature of such correlations are likely to be generalisable beyond the in-depth sample on which they are based. For example, the correlation between δv and vehicle body deformation is likely to be constant regardless of the sample analysed.

With a country such as France where the weather and road conditions vary considerable by latitude, then it may not appropriate to use all the in-depth Regional data for some analyses. If, as an example an estimate is say related to icy conditions and we require a pan-European estimate then it would be sensible only to include those Regions from available National data where icy conditions were not rare. It is then acceptable to use selected Regions, i.e. we are taking into account climatic and topographical factors. This general principle can be applied at a pan-European level.

5.7 Guidelines for Pan European In depth Accident and Injury data base

An In-depth database may not be the most efficient way of providing a specific answer, but if an answer is needed now and not in 'n' years time, then an In-depth database may contain the required data. It must therefore, have sufficient data in depth and breadth to be able to answer most likely questions.

It is not difficult to design an in-depth accident investigation database once the objectives of having the data have been fully defined, i.e. once the questions to be answered and the hypothesis to be tested etc. are known. However, the reality is that questions will be asked over time by researchers, politicians, motor manufacturers or by public opinion. It is therefore not possible to define the best database for all purposes so a compromise must be made, and the guidelines must reflect that from a pan-European perspective.

The following questions should be asked:

- what is the objective of having the in-depth accident database?
- what level of question are likely to be asked - National/Regional/Pan-European?
- what depth and types of data are likely to be required?
- what sampling strategy should be used - opportunistic/stratified/randomisation?
- what mode of investigation is likely to be best - at scene/follow-up?
- what comparisons with National/Regional databases can be made for linking?
- should unique accident reference numbers be included for direct in-depth to Regional/National linking purposes?
- which variables are more reliable?
- how can a common pan-European coding of variables be achieved - via the CARE project?
- what constraints will be built-in, in terms of the questions that can be answered?
- what Regions (across Europe ?) could be used?

Decisions for an in-depth accident investigation

The decision whether to have an at-the-scene investigation or a follow-up investigation, or perhaps a combination of both, is first a question of what will be the aim of the in-depth investigation. The main goal of the implementation of an in-depth investigation is the gathering of detailed information about road crashes: vehicle information, medical data, environmental details, psychological aspects and so on. The evaluation of injury prevention measures as well as accident prevention measures is dependent on the data collection form.

Normally it can be said that an on the scene investigation is more expensive than a follow-up study but this depends on organisational possibilities (e.g. the definition of the accident population, notification by emergency services etc.), when implementing an in-depth investigation.

If an at-the-scene investigation is to be made then a different set of requirements apply as compared to a follow-up investigation. The following table 6.1, compares some of the basic requirements/consequences of two possible approaches.

Table 6.1 Comparison of at-the-scene v follow-up approaches

requirement/consequence	**at-the-scene**	**follow-up**
requires immediate notification by Police	yes	not so essential
a team of investigators on stand-by for pre-determined periods	yes	no
sampling method is opportunistic, in so far as it is not possible to know when they will happen	yes	no
sampling can be stratified and randomised more fully	possible	yes
accident causality can be investigated fully	yes	limited
information about all road users involved (including passengers) available	yes	limited
information about environmental aspects available	yes	limited
information about driver behaviour (psychological aspects) available	yes	no

Another decision relates to the population of interest. Is it better to focus on just car or car derivatives which are not very old and so the data may influence the vehicle design standards, or is it better to try and cover all types of accident in order to be able to answer questions about all road users and their accidents? This is a difficult question to answer and depends upon resources available as well as organisational possibilities and the underlying objective of having an in-depth database, is it for motor manufacturers or to provide answers to Political questions?

An in-depth database which contained all types of road user accident information would be able to answer a broader range of questions. Provided that it had sufficient numbers then these questions could be answered with a certain precision. If resources were very limited and there is only a limited interest in, for instance, severe vehicle accidents then it might be preferable to select certain types of accidents and provide a greater precision but for a limited type of estimate.

5.8 Summary

In summary the following steps should be considered when designing a within-country or a pan-European in-depth accident database:

- can we identify the types of questions/estimates required?
- what is the population of interest?
- what is the required precision on any estimate?
- are resources limited - what are the realistic limits?
- what is the optimal sampling approach?
- what is the best data collection approach?
- what will be the constraints on the answers that can be provided?
- what is the acceptable delay in obtaining an answer?
- Practical restriction

Specific issues to be considered from a Pan-European perspective
If estimates are to be made on a Pan-European basis then certain conditions will have to be met. The first is the acceptability of the National estimates, which in turn require the acceptability of Regional estimates. Only once these pre-requisites have been made can a Pan-European estimate be considered. These are the stages as described above in this deliverable, and where the linking from In-depth to Region has been illustrated via examples from France, Germany and the UK.

Any constraints which apply at a Regional or National level will also apply at a Pan-European level. Further, all the constraints from different Countries must be taken into account, hence if one Country only has data on new cars then a Pan-European estimate including that Country can only be based on new cars. This does not necessarily mean that different sampling schemes used in different Countries cannot be combined at a Pan-European level, but it does mean that some care should be exercised before combining such data. However, it does mean that the population of interest must be defined and should be compared to the population actually sampled in order to establish if the estimation required is possible.

Once details of all European In-depth databases are known it would be desirable to determine the population(s) in common, and hence the questions that can be estimated from data across Europe. For particular sub-studies, different sets of data from different countries may be applicable. For example, it may be necessary to take data from some Regions in some countries because of the question being answered, it would be more sensible to combine data from Regions with similar weather patterns if estimating a multi-National estimate which was about icy conditions.

Generally speaking, methodology is available to calculate estimations for parameters of interest resulting from complex sampling surveys, according to three important characteristics : clustering, stratification and sampling weights. As an example of this kind of design, the National Highway Traffic Safety Administration (NHTSA) in the US, has organised for a long time the National Automotive Sampling System (NASS), based on a multi-stage probability sample of which is selected the Crashworthiness System (CDS) for providing In-Depth data collection (for police-reported crashes involving at least one passenger car that was towed away from the scene due to crash damage). Note that lot of statistical papers deal with the subject of complex sampling design, most of them using a classical parametric approach (Kish, 1965, Skinner, 1989) and some of them using psuedo-replication techniques to obtain variance estimations (Shao, 1996).

The broader the available data the broader the range of questions that can be addressed. This suggests that, as a policy, it may be better to broaden the data and scope of in-depth databases - but not at the expense of sample sizes. The broadening of in-depth databases implies not just more accident data but data from more accidents. There is clearly a resource implication in this policy, because with fixed resources then more in-depth data generally implies fewer accidents and so a narrower population. The question must be asked about the payment of data collection either more in-depth or bigger population, if it is for a pan-European benefit then should the expense also be on a pan-European basis? More specifically, if pan-European research and estimation into accidents using in-depth data is required then an appropriate funding structure must be established.

5.9 References

Hedges L.V., Olkin I., Statistical methods for meta-analysis. San Diego, CAA: Academic Press, 1985

Kish L., Survey Sampling. New York: John Wiley & Sons, 1965

Deming, W. E.: Sample Design for Business Research. New York: Wiley 1960.

FERSI (Forum of European Road Safety Research Institutes). Road safety research and policy in Europe. Mission Paper, Netherlands, 1996

Hautzinger, H.: Statistische Methoden zur Auswertung der Erhebungen am Unfallort. Schriftenreihe Forschungsberichte der Bundesanstalt für Straßenwesen, Heft 220, Bergisch Gladbach 1990.

IRTAD, Under-reporting of road accidents recorded by the police, at the international level. Special Report OCDE-Road Transport Programme, Public roads administration, Norway, 1994

Koornstra M., The quantifying of road safety developments. Proceedings of the VTI konferens, Birmingham, 1996, Linköping, Sweden : VTI, 1998; 7A, 2 : 167-83

Laumon B., Martin J.L., Collet P.et al, A French road accident trauma registry : first results. Proceedings of the 41st annual conference of the Advancement of Automotive Medicine, Orlando, 1997, Des Plaines, IL. : AAAM, 1997 : 127-37

Pettiti DB., Meta-analysis, decision analysis, and cost-effectiveness analysis in medicine. New York: Oxford University Press, 1994

Schnell, R. / P. Hill / E. Esser: Methoden der empirischen Sozialforschung, 4., überarb. Auflage, München: Oldenbourg 1993.

Shao J., Resampling methods for sample surveys, Statistics, 1996, 27: 203-254

Skinner C.J., Holt D., Smith T.M., Analysis of complex surveys, New-York: John Wiley & Sons, 1989

Stenger, H.: Varianzschätzung bei komplexen Stichprobenerhebungen. In: Schriftenreihe der DVWG, Band B81, Aktuelle Probleme und neue Methoden der Verkehrsstatistik, S. 143-150, Bergisch Gladbach 1985.

5.10 Annex A

Calculation of estimates for a stratified sample with 95% confidence intervals

Definitions:
H strata
n_h sample size in h^{th} strata, h=1.H
N_h population of h^{th} strata
y_{hi} metric of interest for i^{th} sample in h^{th} strata, $i=1.n_h$
y the population mean to be estimated
f_h is the sampling faction
W_h is the proportion of the population in stratum h

where, $W_h = N_h / N$ and, $N = \Sigma_h(N_h)$

population mean, $y = \Sigma_h (W_h * \Sigma_i (y_{hi} / n_h))$

$= \Sigma_h ((N_h /N) * \Sigma_i (y_{hi} / n_h))$

and variance of the mean, $var(y) = \Sigma_h (W^2_h * var(y_h) / (n_h -1))$

If the metric of interest is a count of occurrences, then y_{hi} would be coded as 1 or 0 depending if the characteristic was observed for the i^{th} sample in h^{th} strata, or not. The proportion in the h^{th} strata is then given by:

$p_h = \Sigma_i (y_{hi} / n_h)$

with variance, $var(p_h) = ((1-f_h) * p_h * (1- p_h) / (n_h-1))$

The population estimate for the number of occurrences is then:

$No = \Sigma_h (N_h * p_h)$

with variance, $var(No) = \Sigma_h (N^2_h * (1-f_h) * p_h * (1- p_h) / (n_h-1))$

The 95% confidence interval is calculated by:

(population estimate) $\pm$ 1.96 * $\sqrt{}$(population estimate variance)

6 Work Package 3 – External Consultation

6.1 Table of contents

6.2 Introduction

The objective of this package was to ensure that outside groups that collect in-depth crash injury data were aware of the details and progress of the project. They were able to make comments, suggestions and provide feedback on the technical specification of the protocol.

The objectives and results of each task in Work Package one and two were disseminated by mail, e-mail, the world-wide-web, personal visits and academic papers. There was strong encouragement to industry, the academic community, the International Standards Organisation and other groups to participate. Comments were fed back to the relevant Work Packages for group discussion and if necessary, modifications, of the task and/or conclusions.

There was initial confusion as to the deliverable from the STAIRS group (many were expecting a large, working database to be produced), but when clarified all the interested external organisations were willing to add their thoughts and views on a whole range of issues.

6.3 Vehicles for the message.

6.3.1 Conference papers and posters:

Initially, the main vehicle for raising the profile of the project and for generating interest in it, was a series of academic papers and posters at various conferences. These were:

- RISC 1997 (Road Infrastructure and Safety Research Conference) - *The Standardisation of Accident and Injury Registration Systems (STAIRS).*
- FERSI 1997 (Forum of European Road Safety Research Institutes) - *The Standardisation of Accident and Injury Registration Systems (STAIRS). First results.*
- IRCOBI 1997 (International Research Council on the Biomechanics of Impact) - *The Standardisation of Accident and Injury Registration Systems (STAIRS). First results.*
- ESV 1998 (Enhanced Safety of Vehicles) - *An approach to the Standardisation of Accident and Injury Registration Systems (STAIRS) in Europe.*
- Finnish Road Safety Conference - Helsinki 1998 - *Standardisation of Accident and Injury Registration Systems (STAIRS) in Europe.*

Initially, the papers were used to bring about an awareness of the project and to generate interest in it. Later, they were used to disseminate the first results and to elicit a response from other experts in the field. Finally, the papers gave the results from the group concerning both in-depth data and the statistical methodology.

The feedback and interest shown in the presentations was positive, to such an extent that invitations were extended to members of the consortium to speak at other venues.

Copies of the above papers are given on the CD Rom.

6.3.2 Presentations:

One of the main presentations was to a meeting of the International Standards, Working group 7 - Road vehicles - Traffic accident analysis. After the presentation it was decided to try to adopt STAIRS as an ISO standard and this is currently being pursued by the group. Individual parts of the protocol have also been used within other areas of their remit, for example, the crash severity paper.

In the UK a presentation was given to the Co-operative Crash Injury Study (CCIS) consortium, which included representatives from some of the main vehicle manufacturers, research groups, and government. This was given at the formative stage of the in-depth protocol development when there was no available completed work on the statistical side to demonstrate how it would be applied. Even with this lack of a working protocol, the response was extremely positive.

A presentation was also given to the UK's Institute of Road Safety Officers at their yearly conference. The audience included professionals within the primary and secondary safety fields. A great deal of interest was shown in the areas concerning the vulnerable road user along with the ability to identify if alcohol was present. The ability to use the data to reflect UK local and national datasets was seen as helpful as was the capability to look at smaller sub-sets at a more detailed level.

A meeting was also held with other Directorate-General within the European Commission to explain the scope of the work and how it could help them to make data-led decisions concerning vehicle safety and legislation. The idea that was well received.

6.3.3 The Workshop:

One of the major tasks of the STAIRS project is to disseminate the results of each of the assigned work packages. It was decided that a Workshop, attended by all interested parties, would be the most efficient. The workshop would naturally be of interest to car manufacturers and research institutes who are involved in in-depth accident investigations. Representatives of the bodies in charge of national databases would also be interested.

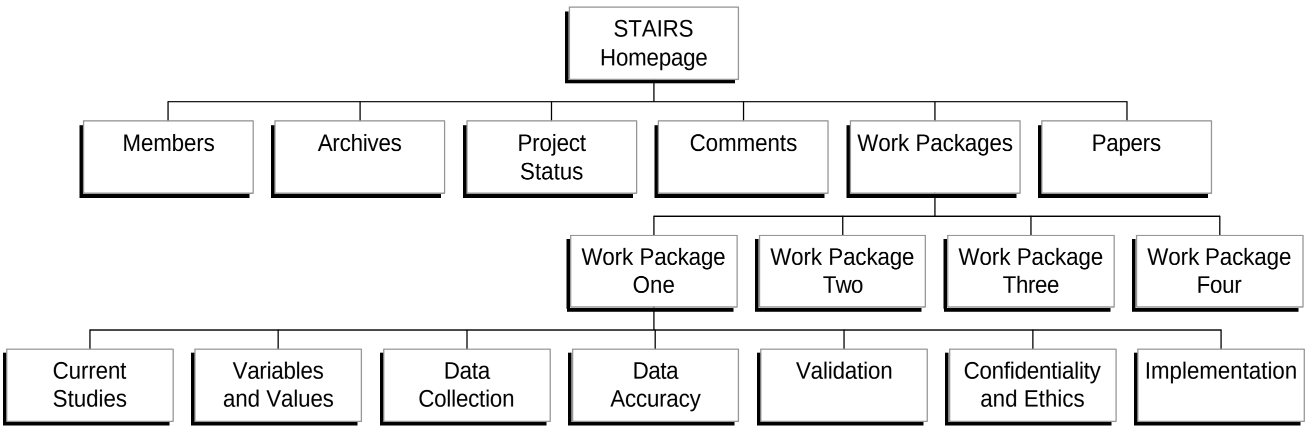
The idea of the workshop was to begin a two-way discussion between the interested parties and the STAIRS partners. It is hoped that from these discussions the protocols developed can better reflect the wants and needs of a wider target audience.

The outcome of the workshop was extremely positive. A good selection of delegates from the differing areas of research, government and manufacturing were present. The feedback was helpful, and areas identified for further work or clarification.

6.3.4 The Website (www.ice.co.uk/stairs):

One of the main vehicles used to disseminate the ideas and outcomes of STAIRS is the website. This has undergone a number of changes and involved constant updating. There are links within the site to other organisations involved in the work. Efforts are now being made to produce a mirror site at INRETS in France.

The site is organised as follows:



An example of one of the webpages:

6.4 Conclusions.

All available medium have been used to disseminate the ideas and outcomes of the work done by the STAIRS group. A wide variety of people from Governments organisations, private industry, and other in-depth collection groups have been actively encouraged to look at the outcomes and deliverables. Much interest has been generated and certain groups, such as VALT (Finnish Motor Insurers Centre) have said they intend to adopt the system in total or part, within their working practices. INRETS from France, are already using the pedestrian variables within their current study on the subject.

The website has the greatest potential for continuing to attract further interest, and will require constant updating to ensure that this interest is retained. An open forum could be included on the site in order to allow on-going discussions or ‘threads’ to develop, with the intention of further developing the protocol or receiving feedback concerning the use of either the in-depth collection process or the statistical element of the work.

7 Work Package 4 Workshop – Presentation of the results

7.1 Table of Contents

7.2 Executive Summary

One of the major tasks of the STAIRS project was to disseminate the results of each of the assigned work packages. It was decided that a Workshop, attended by all interested parties would be the most efficient. The workshop would naturally be of interest to car manufacturers and research institutes who are involved in in-depth accident investigations. Representatives of the bodies in charge of national databases would also be interested. Each partner from the consortium submitted a number of suitable candidates for this activity and after the group decided upon how many guests were to be invited, a final list was drawn up. Invitations were sent out from a central co-ordination point and follow-up phone calls were made a month later.

The idea of the workshop was to begin a two-way discussion between the interested parties and the STAIRS partners. It is hoped that from these discussions the protocols developed can better reflect the wants and needs of a wider target audience.

The workshop would also allow the Commission to review the results and reflect on the further work to be carried out on the pan-European harmonisation of accident and injury databases.

In the morning, presentations were given on reviewing existing in-depth studies, values and variables to be collected, data quality, data collection methods and confidentiality and ethics. Discussion and feedback was encouraged at points throughout the morning, and areas for further review and consultation were identified.

In the afternoon there were presentations on the problems associated with developing a statistical linkage to national databases, and a series of examples from France, Germany and the UK were shown. Problems associated with the differing collection areas within each country were highlighted and the strategies used to overcome them described. Feedback given by the delegates on this topic ranged from adopting a completely new approach to weighting, to comments on the suitability of those who use the data.

The workshop was discussed by the group the following day, and it was agreed that the outcome was extremely positive. A good selection of delegates from the differing areas of research, government and manufacturing were present. The feedback was helpful, and areas identified for further work or clarification. Follow-up written comments are expected from the delegates.

Copies of the slides used in the presentations and a full, written description of the presentations can be found on the CD Rom. Only the points of discussion will be given here.

7.3 Objective:

The aim of the STAIRS project was to define the fundamental requirements for a pan-European, in-depth crash injury database.

In order to be useful, the recommendations of the project must be widely disseminated. As a complement to Work Package 3, the objective of this package was to inform interested parties about the complete results and conclusions.

7.4 Agenda and delegates:

The meeting took place from 09.30am on June 18th, in the Commission building on Rue Froissart, room 1b. The times and topics for discussion were as follows:

9.30 - Arrival
10.00 - Welcome (R. Bastiaans)
10.10 - Delegates introduction (P. Lejeune)
10.15 - Introduction to STAIRS (G. Vallet)
10.20 - Existing studies (G. Vallet)
10.35 - Variables and values (R. Ross)

11.05 – Discussion

11.50 - Data accuracy (R. Ross)
12.05 - Data collection (G. Vallet)
12.15 - Confidentiality and Ethics (P. Thomas)

12.25- Discussion

13.00 – Lunch

14.00 - Data linkage objectives (P. Lejeune)
14.05 - Methodological approach (B. Laumon)
14.20 - Examples and problems (J. L. Martin)
14.35 - Examples and problems (I. Koßmann)
14.50 - Examples and problems (B. Sexton)
15.05 - Pan-European dimension (B. Sexton)

15.20 – Discussion

16.15 - Conclusions and next steps (R. Bastiaans)

16.30 – Close

Delegates that attended the workshop were:

Mr. R. Bastiaans (European Commission)
Mr. R. Burberry (DETR), United Kingdom.
Mr V. Carrara (Elasis), Italy.
Mr. P. Fay (Ford), United Kingdom.
Mr. L. Hantula (VALT), Finland.
Mr. A. Hobbs (TRL), United Kingdom.
Mr. N. Jahn (ACEA), Belgium.
Mr. M. Koornstra (SWOV), Netherlands.
Mr. T. Lekander (Swedish National Road Administration), Sweden.
Mr. Y. Page (Ceesar), France.
Mr. A. Plasencia I Taradach (Institut Municipal de Salut Publica), Spain.
Mr. C. Thomas (LAB, PSA, Peugeot, Citroën/RENAULT), France.

7.5 Work Package One.

7.5.1 Discussion: Existing studies/Variables and Values

A discussion session followed. The first question came from Dr. Koornstra and concerned the fact that Delta-V had not been described, nor were there any variables on lights and lighting of the road. Mr. Vallet pointed out that each group described Delta-V in a different way, gave it a different meaning and used different methods of calculation. Due to these circumstances, it was decided that the differences were too great to be solved within the time-scale given, and further and wider research involving more partners would be necessary to obtain consensus. The work package partners felt this was a positive outcome and hoped the recommendation would be taken up by the Commission. Continuing, Mr. Vallet pointed out that STAIRS was focussed on secondary safety. Lights and the lighting of the road is considered to be more within a primary safety context. This was the reason that lighting variables were not included within the STAIRS protocol. He went on to say that as the system is modular and can be used in a variety of ways, there is no reason why these types of variables cannot be added by the individual collection system concerned.

Prof. Hobbs agreed with Mr. Vallet concerning the Delta-V problem, saying that it was difficult to define the behaviour of the structures involved. He also posed the question of whether they can be objectively assessed? He used an example of coding the contact width on a vehicle and the difficulties that arise due to induced damage.

He commented further on the problems he saw with the “D” point (an internal measurement point of a vehicle) and how to identify it, especially with the different design of seats currently on the market.

Another area Prof. Hobbs mentioned concerned the knee impact area variable. He felt that it should not be defined too rigidly as many areas outside of the zone described can cause injuries. The two final areas he commented upon were: the pedestrian impact areas seem over complex and need to be refined, and information on airbag algorithms would be helpful.

Other areas that the delegates would like to see expanded were seat positions and two-wheelers. Mr. Placensia stated that many of the problems with two-wheelers occurred in Southern Europe, and guidance in this area would be appreciated.

He also raised the topic of individual disability and disability scales, commenting that guidance in this area would have been helpful. He continued, explaining that IIS (Injury Impairment Scale) did not work very well. A proxy, such as the length of time before returning to work might be useful as a guide. Prof. Hobbs explained that a number of disability scales had been tried over the years, with varying degrees of success and saw this as another area that requires further research.

Mr. Fay questioned the section of the document that describes interior measurements within the vehicle, especially the measure from the seat back to the centre of the steering wheel. He pointed out that many vehicles now have adjustable steering columns and this would cause problems with the accuracy of the measurement. He suggested taking the measure to a fixed point as well as the column. Continuing, he stated that the intrusion matrix described worked well for cars, but had reservations about side facing seating or vehicles with three or more rows of seats. Finally, he pointed out that the terms used to describe body styles were out of date. Many new styles had been developed that would not fit into any of the categories given. He agreed to send an updated list to the group.

A problem with the coding of the principal direction of force was brought up from the delegates. Prof. Hobbs stated that due to possible rotation of the vehicle during the crash phase, this variable could not be very accurate. The group agreed, but pointed out that assumptions had been made as to the crash process. However, it was noted that these had not been identified within the document and it was agreed that an explanation of the assumptions will be given it the next draft.

7.5.2 Discussion: Data Accuracy/Data Collection/Confidentiality and Ethics

Mr. Placensia began the discussion by agreeing with the problems of data retrieval: some data is not always present, or it is not suitable for answering the question. Training in these tasks is critical, both for the quality of data and professional standing.
Prof. Hobbs mentioned that future legislative proposals would further restrict access to the kind of information we require, and it may be helpful if this was highlighted in our document. Other delegates put forward the problems experienced in their particular countries, and it was seen that there was a great deal of disparity between the availability of access to information.
Prof. Hobbs suggested that this may be an area that requires further work with a wider selection of organisations, to ensure that our important area of study is not excluded from the information it requires.
The morning session was then brought to a close by Mr. Lejeune.

7.6 Work Package Two

7.6.1 Discussion

Dr. Koornstra opened the discussion, stating that some very good work had been done. However, there are problems of underreporting that the group have not taken into consideration. He suggested a supplementary, hospital based approach so that all crashes are taken into consideration and this would better reflect reality. He would like the group to make recommendations on national data and its representation of reality.

A further suggestion was made concerning weighting. Dr. Koornstra suggested all linking variables are potentially of interest and could be combined into a single weighting variable. This would simplify the use of the data for researchers as no further weighting need be applied. The STAIRS group replied by raising issues they thought would cause problems using this technique. However, as the details of this methodology had not been studied a comprehensive reply could not be given. Dr. Koornstra promised to send a written explanation of his multivariate approach along with the meaning of individual weight estimations.

Mr. Page commented that he preferred the STAIRS system as it would help with the investigation of sub groups, such as motorcyclists. He continued, and said that the linkage proposals are very interesting and could establish a stake in future studies.

Prof. Hobbs then posed the questions “why are we doing this?” and “what do you do when you have a population estimate?” He explained that sometimes there are no data available, or it has been missed, and weighting does not give a reliable estimate. If the data has been weighted can you then carry out statistical tests on it? He would like the report to stress the limitations of any analysis or technique very clearly.

In the discussion that followed, it was re-iterated that it was important to consider sampling design when estimating, especially with respect to the standard errors and confidence intervals. It was also important that the population of interest needed to be clearly defined. This should then ensure that when data are combined at higher levels (National or pan-European), estimates would be compatible.

Prof. Hobbs made one last point. This concerned the aggregation of the databases. He put forward the view that it will be many years before the databases are, in any significant way, close to harmonisation. It would seem more appropriate to find a method of combining the analysis rather than the data. This may be another area that needs to be pointed out in the report.

The discussion was then brought to a close and Mr. Bastiaans rounded off the workshop with his conclusions.

7.7 Conclusions and next steps

Mr. Bastiaans stated that this morning he had already indicated the importance that the European Commission attach to the availability of European, harmonised road safety data to support road transport safety policy development. STAIRS is clearly aimed at such a harmonisation effort. He considered the fact that major European research institutes have found each other in the preparation and execution of the STAIRS project, as fortunate and hopefully giving. But it is nevertheless clear that there is still a long way to go before we actually have Europe-wide harmonised in-depth accident and injury data, databases or one database. There are barriers to overcome, for instance:

- the way in which data is collected or in fact not collected at all in different Member States,
- with respect to ethical questions, privacy and confidentiality of data,
- the costs involved in data acquisition, data analysis and information provision,
- the differences in importance that various people attach to road safety in general across the European Union,
- the willingness to harmonise our efforts, which will imply a change in "the way we do things around here".

STAIRS is only one of the steps in this process, but he considered it an important step. From this point of view, the acronym STAIRS as opposed to one STEP - suggests too much. But the fact that the delegates have taken the trouble to come and attend this workshop in Brussels made him hopeful. It means that the delegates are at the very least interested to see what is going on in the area. He hoped that the delegates hadn't come as a watchdog, but as a representative of an organisation willing to think about further co-operation. Many of them had actually provided the consortium with very good and valuable technical comments, and he was sure the consortium will take these comments on board in their further work and in their final report.

DG VII are currently defining future Research and Development activities to be carried out under the Fifth Framework Programme in the Key Action entitled "Sustainable Mobility and Intermodality". At this stage, he couldn't promise anything, because the Commission are still in an open phase of programme development and tentative definition. But he would be in favour of taking the framework developed in STAIRS further, put it to the test in validation, and assess its added value and usefulness, as well as its limitations to different user groups. This may also be a good opportunity to enlarge the co-operation, not only limited to research institutes, but also to other stakeholders such as road authorities, the police, car manufacturers, insurance companies, and consumer organisations. The co-operation should not be limited to European Union Member States, but possibly should also include Central and Eastern European countries, some of which actually have shown interest in this area of work. He would like to encourage the delegates to stay interested in the further progress of the STAIRS project, not only by visiting the STAIRS homepage on the World Wide Web, but preferably also by direct contacts with the consortium.