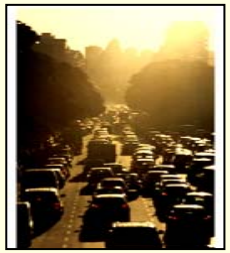
Botanical Garden of the
University of Valencia (Spain)
14-15/June/07



Urban Accident Analysis Systems

Project co-financed by the European Commission,
Directorate-General Transport and Energy
(TREN-03-ST-S07.30828)

QUALITY AND REPRESENTATIVITY OF THE TRAFFIC ACCIDENT DATA IN URBAN AREAS: STATE OF THE ART (Deliverable I).

Elena López de Cózar

State of the art

- Collection of traffic accidents: general aspects
  - Definitions and procedures
  - International and national databases
- Accident data quality
  - Under-reporting
  - Under-recording
  - Errors and biases
- Urban accident analysis systems in Europe : some practical cases
- Conclusions and recommendations for the optimal operation of the traffic accident data collection, management and analysis systems

Accident data quality

- Definition:
  - To what extent are all the accidents registered? → UNDER-REPORTING
  - Are the accident records complete?
  - Is the registered data precise or reliable?
  - Level of accessibility and delays of the data to be able to be used by the users (Pfefer, Raub y Lucke, 1998)

→ QUALITY

Accident data quality

- Under-reporting
  - It makes reference to those accidents or traffic victims that for different reasons are not registered. It is related to the degree in which the records truthfully represent the real number of traffic accidents and victims.
- Under-recording or missing data
  - This happens when certain information fields are not recorded in a certain proportion of accidents, either for the impossibility to collect the data, because they create special difficulties for its completion, or for other circumstances difficult to specify.
- Errors and biases
  - *Errors:* incorrect data, that do not show any tendency, and are distributed approximately at random.
  - *Biases:* systematic incorrect or erroneous data registered showing a tendency towards a certain value or rank of values

Under-reporting

- The daily experience
- Scientific studies :
  - Linkage methods
  - Capture-recapture methods

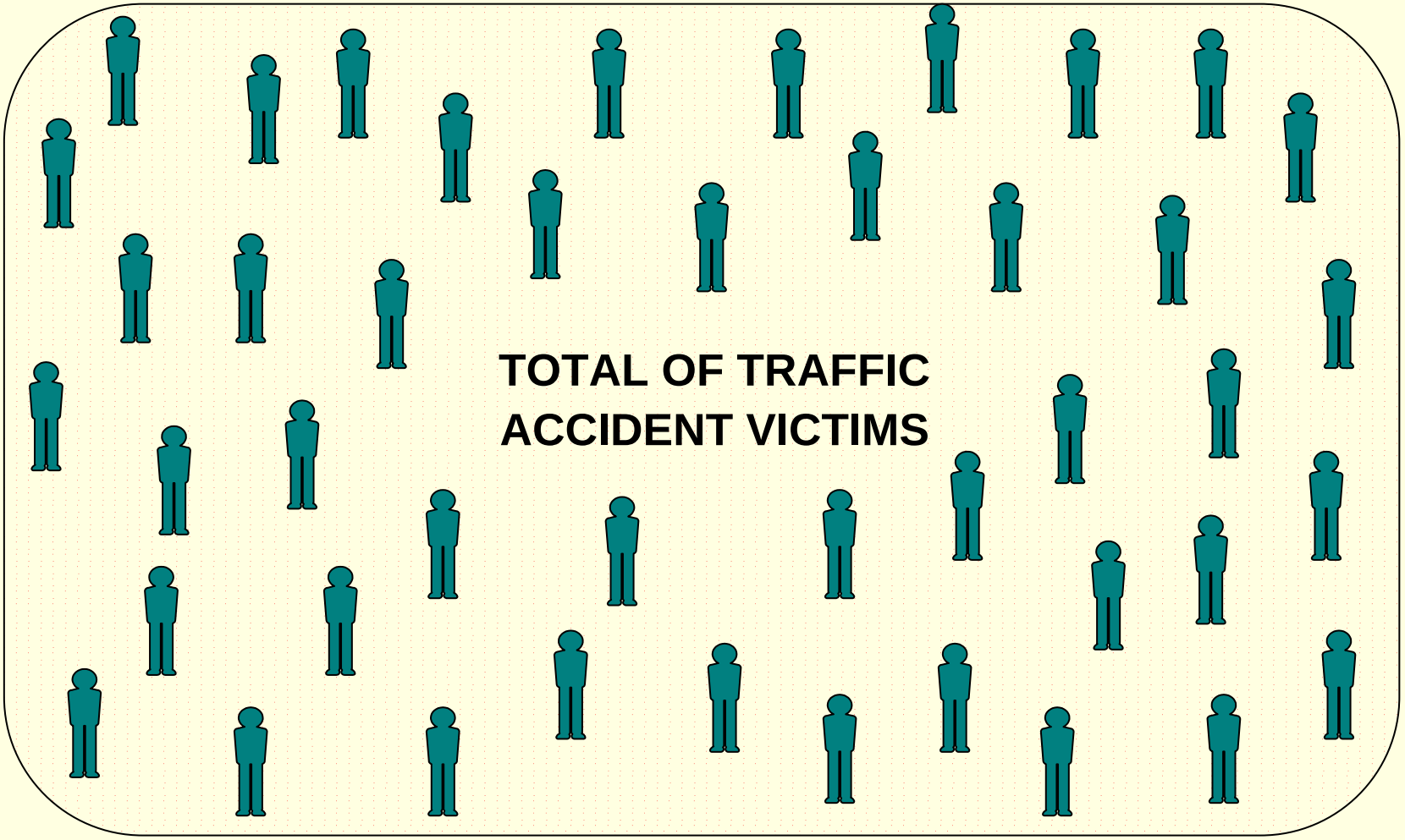
Under-reporting: daily experience

- Task conflicts in accident situation. Emergency task: traffic control, risk signposting, attention to the victims ,...
- The users do not require police attention (verbal or insuring agreements): uninjured implied persons that detect injuries later, slight and material damage only accidents, drivers without license, without insurance or under the effect of alcohol-drugs, ...
- Restriction or ignorance of the criteria about what has to be collected as traffic accident
  - Accident severity: slight and material damage only accidents
  - User type
  - Special traffic “accidents”: suicides, natural deaths, murders,..
  - Vehicles in motion: falls within the vehicle, when getting on or getting off the stopped vehicle, vehicles without driver,...
- Lack of personal resources to attend all the accidents
- Lack of motivation (overload)
- Police competences towards traffic (geographic areas of action)
- Political decisions

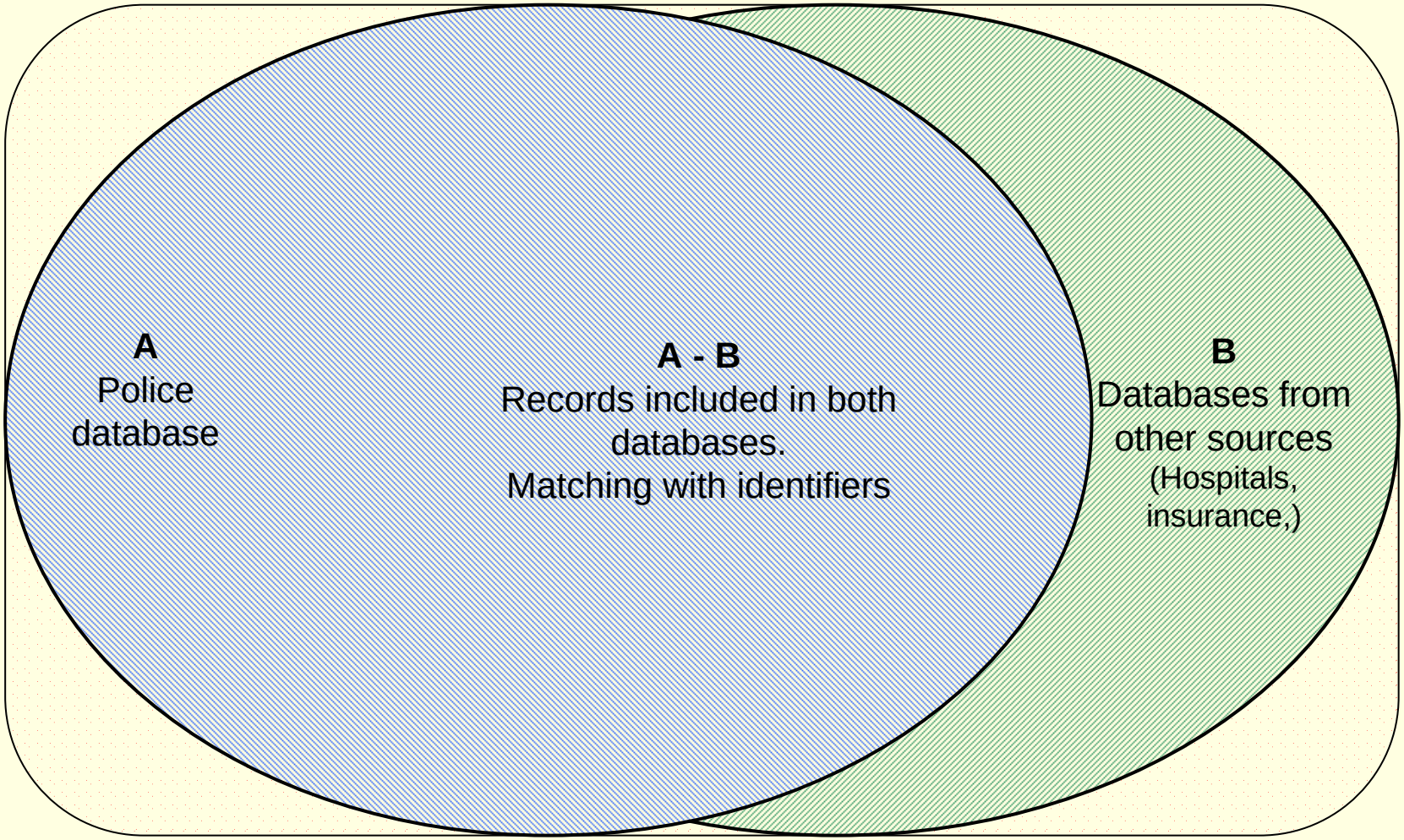
Under-reporting: Scientific studies

- Unit of measurement : traffic accident victims
- Contrast sources:
  - Police data vs. other sources (health, insuring, private insurance for workers,...)
  - Surveys
- Methods
  - From aggregated data: to contrast the total data
  - From individual data (disaggregated):
    - Linkage
    - Capture - Recapture

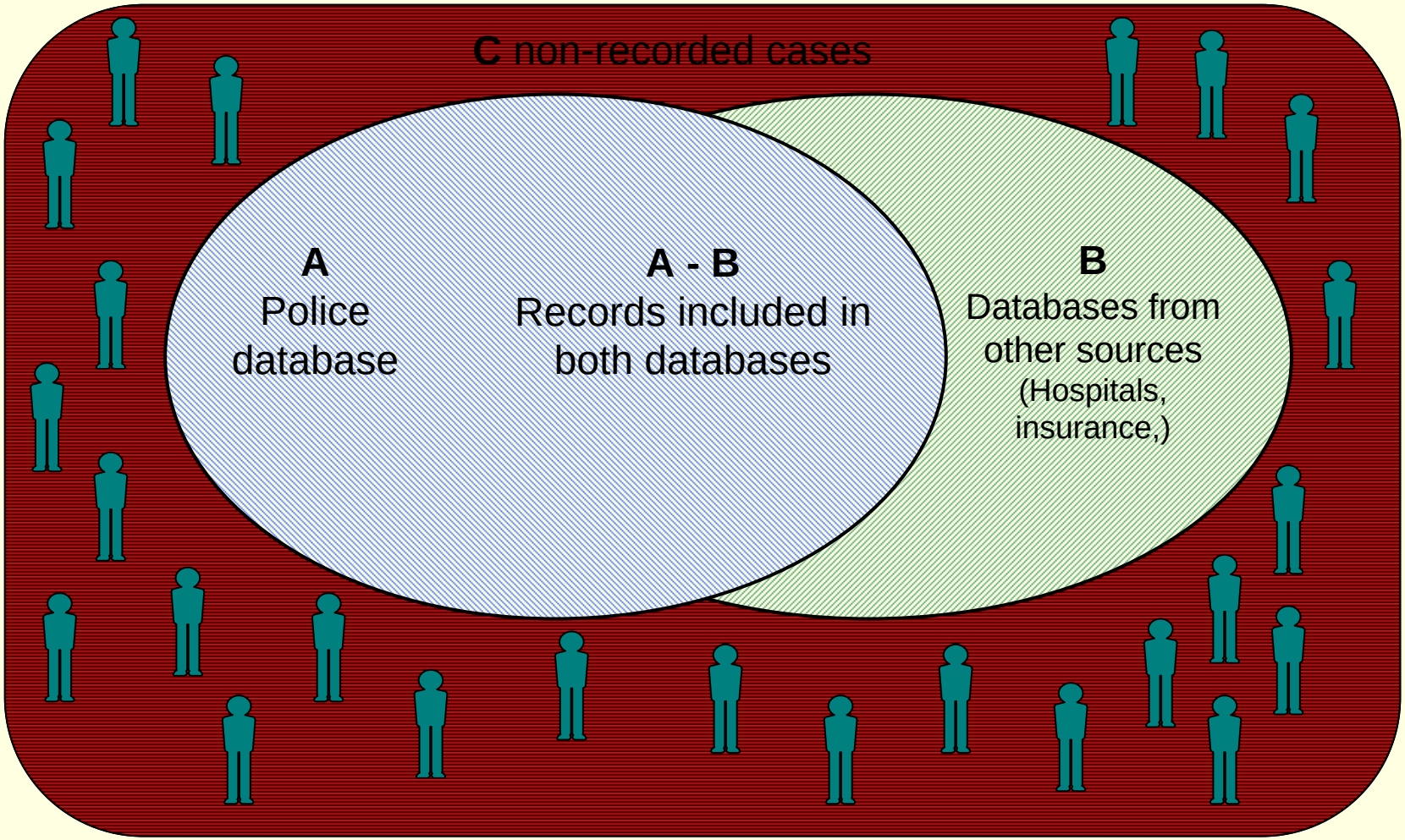
Under-reporting



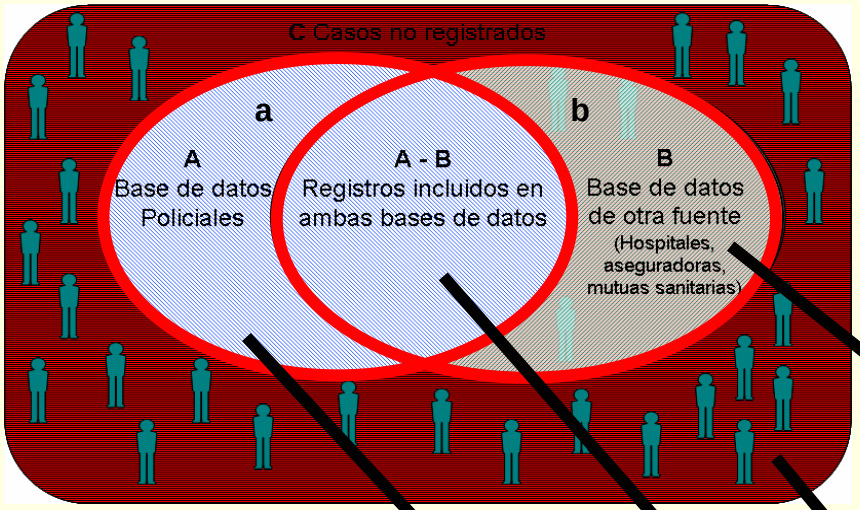
Under-reporting: linkage methods



Under-reporting: capture-recapture methods



Under-reporting: capture-recapture methods



b+C= Under-reporting level of the police records

a+C= Under-reporting level of the other source records

Police data

Other sources' data	SI	NO	
SI	AB	b= B-AB	B=b+AB
NO	a= A-AB	¿C?	**a+C**
	A=a+AB	**b+C**	TOTAL=A+B+C

Under-reporting studies: problems

- Studies on samples of accidents/victims (population estimates)
- Heterogeneity in the record systems (it makes difficult the comparison between data sources):
  - Different definitions of accident/victim
  - Different criterions from record and classification
  - Different information available (variables)
  - Influence of socioeconomic parameters in the record of victims (e.g. quality and efficiency of the health system)
- Estimation methodology (usually it is not described and in many cases it is not specified to what the percentage of representativity is referring)
- Variability in the results obtained in different studies: based on the used sample, compared data sources, ...

Under-reporting: Results

- As it increases the severity of the accident, it increases the probability that it is registered.
- Greater proportion of notification in non-urban zones that in the urban zones.
- The night accidents are notified more than the daylight ones.
- (Simpson, 1996) Higher proportions of:
  - Victims whose vehicles were seriously damaged.
  - Victims whose injuries were apparently detected in the same scene of the accident
  - Victims who were transported to the hospital by the emergency services
  - Victims attended in the hospital immediately after the accident
  - Accidents with more than one victim

Under-reporting: Results

- Groups with greater levels of under-reporting:
  - Cyclists
  - Slight victims
- Groups registered to a greater extent
  - Car occupants (drivers and passengers)
  - Pedestrians (greater severity)
  - Motorcyclists
  - Serious victims / fatalities (100%)
  - Young and elder groups
  - Vans, lorries and buses

Under-recording

- The daily experience
- Scientific studies:
  - Descriptive and exploratory data methods
  - Randomness of missing values

Under-recording: The daily experience

- Absence of the involved persons: casualty severity, hospital move or runaway
- Lack of technical and personal resources to record certain variables: blood alcohol/drugs level, psycho-physiological symptoms (sleepiness, fatigue...)
- Lack of training for the data collection
- Excess of requested information
- Lack of motivation
- Ignorance of the information
- Inadequate design of the record protocol: no alternative adjusts

Under-recording: problems

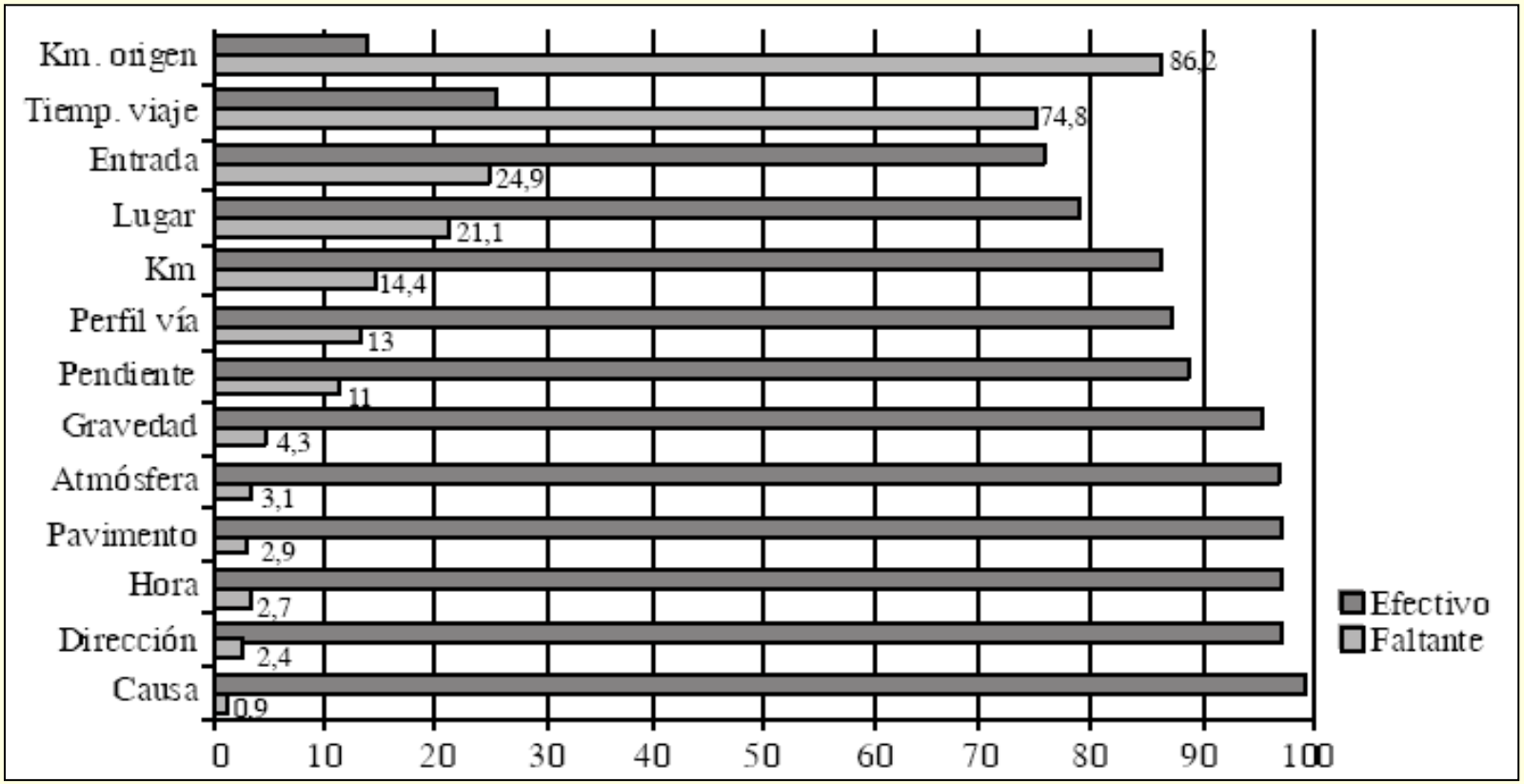
- Computational problems: certain technical statistics are based on matrix operations that require to have all the complete information .
- Problems in the results: the existence of certain patterns of missing values causes important problems hardly monitorable by the statistical techniques ➔ non-acceptable results.

Under-recording: Results

- As the injury is less serious the probability of Under-recording increases.
- Groups with greater levels of Under-recording:
  - Location of the accident: complexity of particular zones of the road network or lack of homogenous criterions (what is the exact location?)
  - Alcohol consumption: accidents with serious victims or fatalities, women, "non-responsible" drivers, pedestrians,...
  - Speed offence (e.g., in Spain 20% in motorway and 50% in urban zone)
  - Use of the safety belt (in Spain 7% in motorway and 39% in urban zone)
  - Vehicle characteristics: brand and model of lorries and heavy vehicles

Under-recording: Results

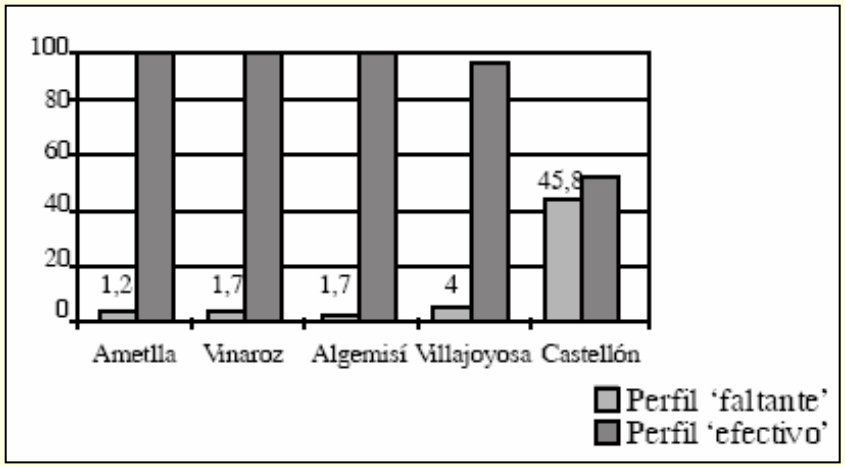
- Missing data (Ledesma, Sanmartín y Chisvert, 2000)



Proportion of missing data some variables in a Spanish motorway (AP-7)

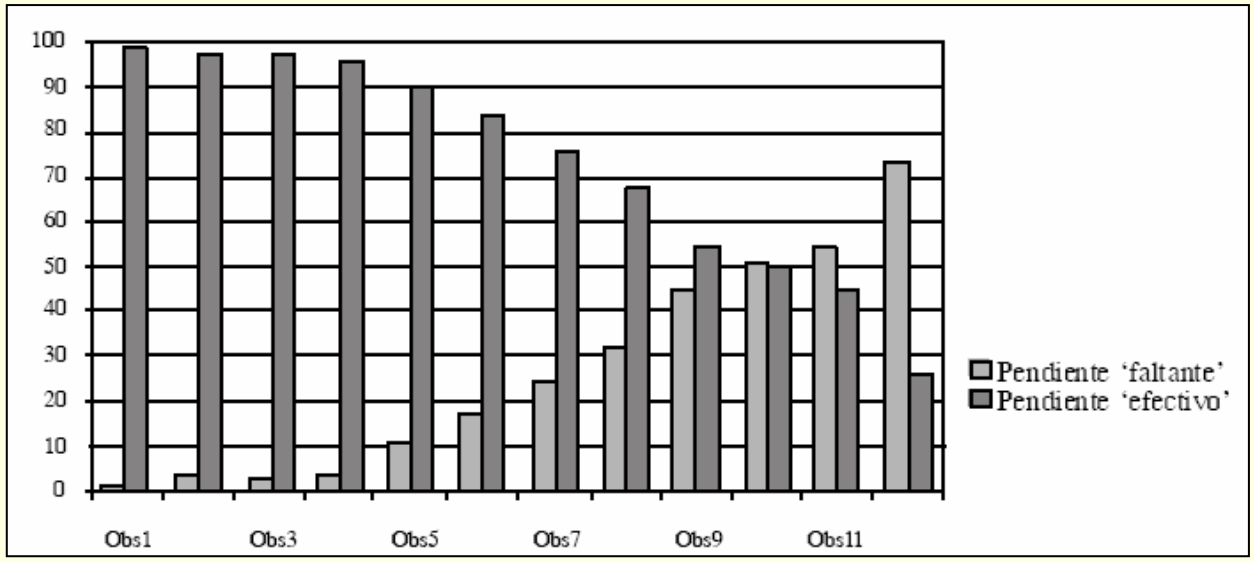
Under-recording: Results

- Missing data patterns (Ledesma, Sanmartín y Chisvert, 2000)



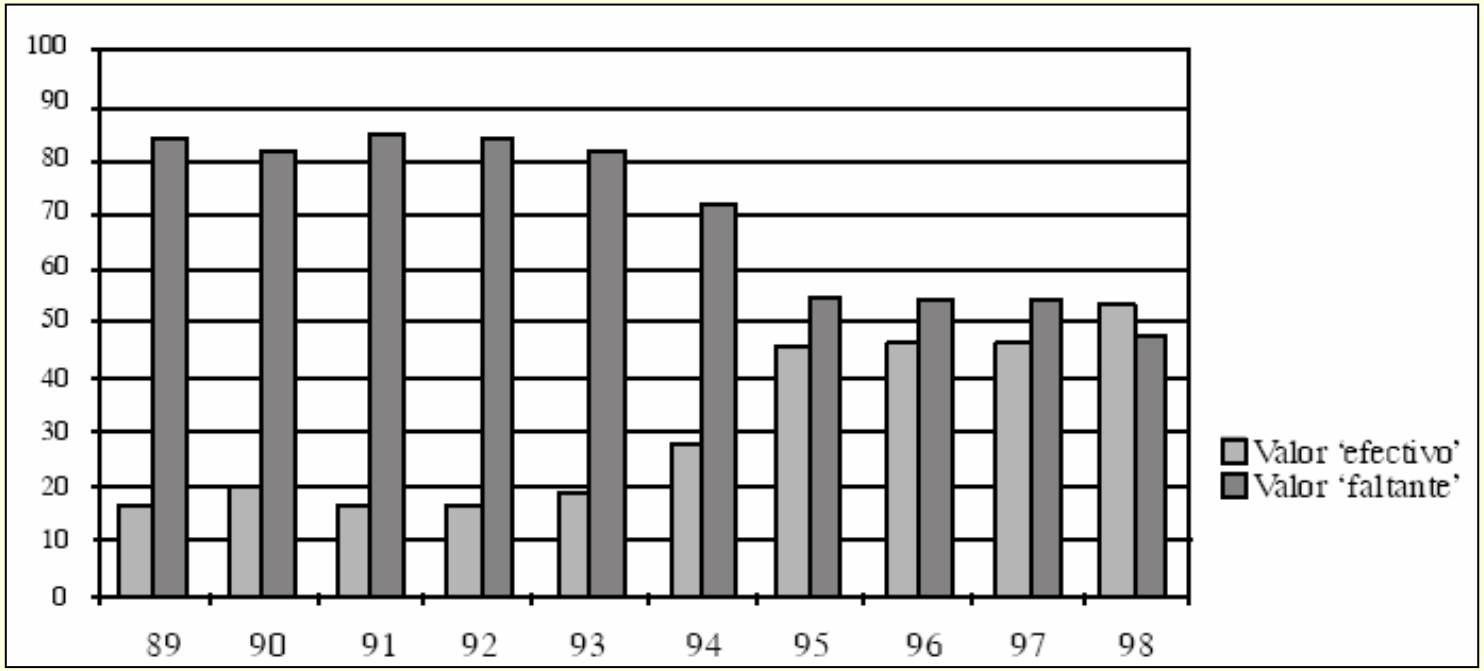
Proportion of missing data in the variable "Profile of the route", according to freeway sections, in a Spanish motorway (AP-7)

Proportion of missing data in the variable "Pending of the route", for different observers, in a Spanish motorway (AP-7)



Under-recording: Results

- Missing data patterns (Ledesma, Sanmartín y Chisvert, 2000)



Proportion of missing data in the variable "Time of trip", through the years, in a Spanish motorway (AP-7)

Errors and biases

- The daily experience
- Scientific studies:
  - Data fidelity
    - Are they correct? ➔ Linkage methods
    - Are they correctly transcribed? ➔ Double entry or revision methods by several operators
  - Consistency and coherence
    - Logical and deterministic tests:

      To detect impossible data (e.g. Sunday-working day; single-vehicle accident-two vehicles involved; 30th of February)
    - Empiric or probabilistic tests:

      To detect improbable data (e.g. heavy rain and dry road; driver- 10 years old)

    ➔ Exploratory techniques and control filters: consistency, out of range values or extreme values

Errors and biases: The daily experience

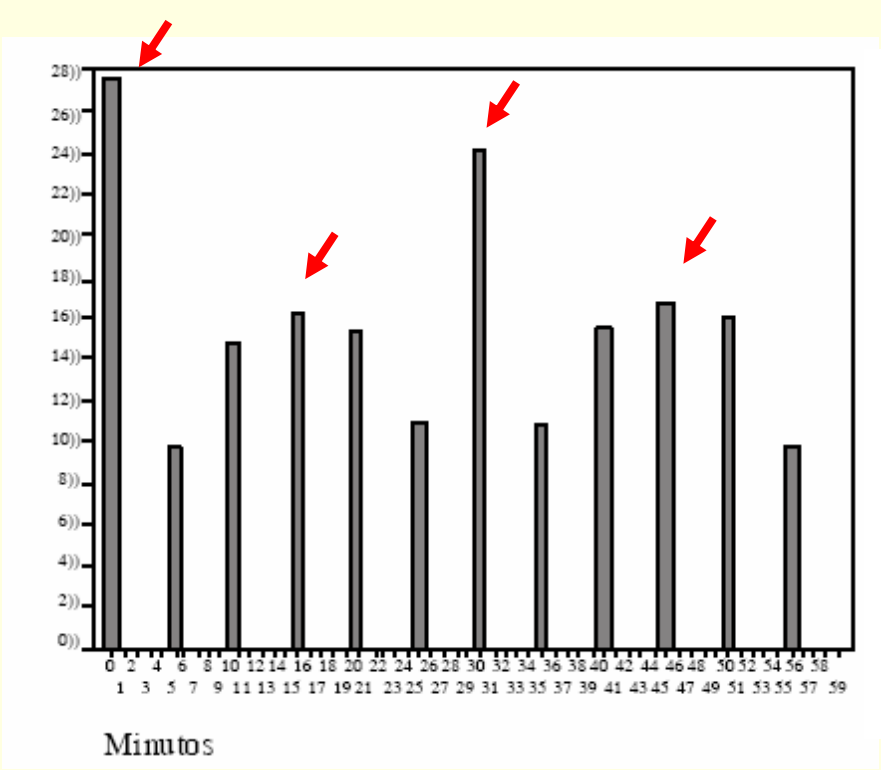
- Classification errors: Severity, accident type, injuries, cause of the accident...
- Lack of training regarding the content: severity criterions
- Rounding errors
- Subjective assessment of the information
- Erroneous judgements based on the experience or on prejudices
- Lack of motivation
- Errors in the information given by the involved persons and witnesses (intentional or not)
- Errors of manual/computer transcription
- Errors in the information processing (transformations, screenings, etc...)

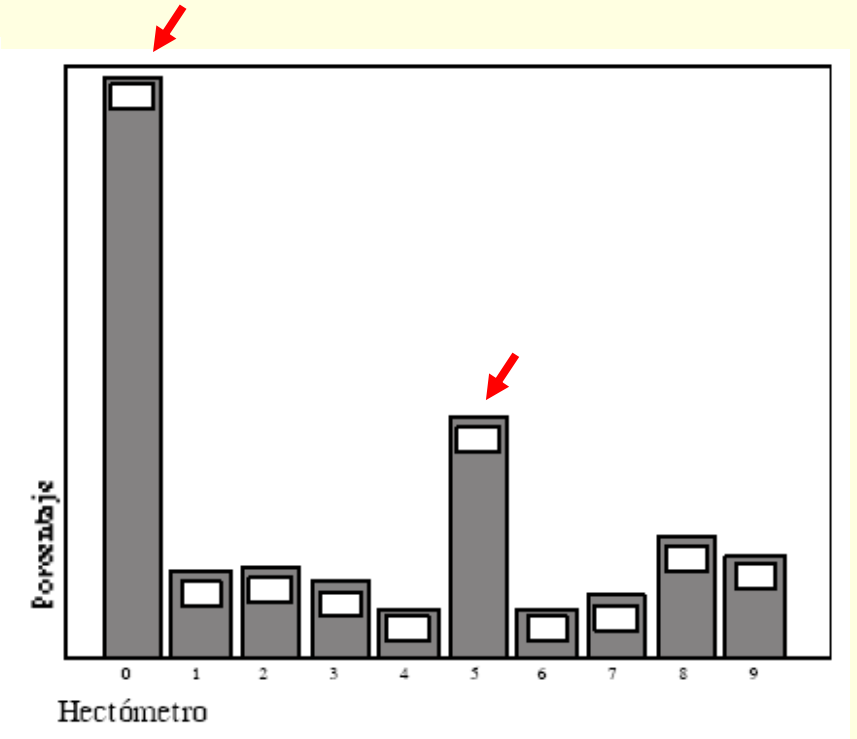
Errors y biases: results

- Severity
  - As the injury is less serious, the probability of classification error increases
  - Under-evaluation and over-evaluation of the severity
  - Biases in particular groups:
    - The pedestrians and motorcyclists have a greater possibility of being classified as serious casualties than car occupants (vulnerable groups).
    - Single accident users (running off the road, salidas de vía, strikes against obstacles,...)
    - Over-estimation of the severity as the age increases (the elderly are the more vulnerable)
- Vague location of the accident (GIS systems and geographical references)
- Alcohol consumption: delay in the sample collection (↓ level of alcohol), intravenous treatments (↓ level of alcohol), loss of blood (↑ level of alcohol) and states of shock (the rate of alcohol elimination varies)
- Age: rounding errors

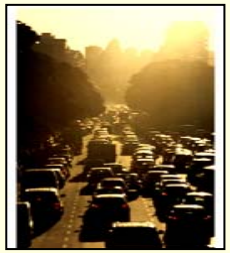
Errors and biases: results

- Rounding biases (Ledesma, Sanmartín y Chisvert, 2000)





Botanical Garden of the
University of Valencia (Spain)
14-15/June/07



Urban Accident Analysis Systems

Project co-financed by the European Commission,
Directorate-General Transport and Energy
(TREN-03-ST-S07.30828)

QUALITY AND REPRESENTATIVITY OF THE TRAFFIC ACCIDENT DATA IN URBAN AREAS: STATE OF THE ART (Deliverable I).

Elena López de Cózar