PSYTEL

Ingénierie de l'information

20 ideas and truths about the IDB

(Injury Data Base)

Marc Nectoux - December 2006 – v4.2

The context :

- After the many evolutions of the IDB (Injury Data Base) system, feeding a long and chaotic history since 1986 (a system successively named EHLASS, HLA, ISS, IDB and then IDB-AI for All Injuries), we now would like to give a global report on its functioning, at a time when some actors question its usefulness.

- Luckily and by choice, we have been associated in France and at a European level with the history of this system since its creation, and, more particularly, to the management of the French component of the IDB system, now called EPAC. Thus we know from the inside its strengths and weaknesses and, twice already, the Commission placed us in charge as project leader, of studies about the European system :

     – "Assessment of the European system functioning for the surveillance of the home and leisure accidents (report August 1997 – contract n°AO-2600/96/000262);

     – "Establishment and development of Data Mining and help to decision tools on home and leisure accidents, as part of the injuries prevention Program". (Report October 2000 – contract n°VS/1999/5238-99CVF3-301).

- It is as observer, user and actor of the IDB system that we express today 20 ideas and suggestions and what we believe to be "truths".

1. The three approaches :

- Let's remember the three different possible approaches for an information system about injuries, approaches that we frequently mentioned in our previous works:

- The goals are the calculations of global incidence rates, the representative aggregated results and the building of national epidemiological data : *it is the "macro-accidentological" approach.*

- The goal is the collection of information allowing spotting risk populations, situations, behaviours or products related with precise lesions, as well as the characterisation of accidents' classes bound to distinctive scenarios : *it is the "meso-accidentological" approach.*

- The goal is to spot very rare accidents, precise dangerous products (a given type, a given brand) : *it is the "micro-accidentological" approach.*

- The three approaches have their own qualities and faults, but clearly answer different objectives, relatively non-compatible, and use different methodologies and specific analysing tools.

2. The effective methodologies :

- According to each approach a corresponding methodology exists as well as analysing tools which result in different efficiency levels :

Methodology ➡️ Efficiency perimeter according to the approach	Data collection by representative survey	Data collection in the hospital emergency units	Data collection by alert system
Macro-accidentological	**High**	Low	Void
Meso-accidentological	Low	**High**	Void
Micro-accidentological	Very low	Very low	**High**

- Regarding the "macro" approach, the most appropriate methodology will be the collection of survey data : it is the approach which meets the first objective of a representative collection about a limited number of injuries. The main purpose is the calculation of an incidence rate; the secondary purpose is the descriptive epidemiological approach and product. It is the typical macro-accidentalogical vision where the statistical quality of the tool is essential. The number of collected accidents will be relatively low and will not allow analysing to different kinds of accidents.

- Regarding the "micro" approach, we prefer researching for direct information about selective reported cases, like in the European alert system such as RAPEX or the debuting system EASI (European Alert System on Injuries) that we established.

- For the "meso" approach, the collection in the hospital emergency units is perfect since it allows collecting continuously a great number of injuries that integrates validated medical data. The representative character of the data is more secondary, since it is the quantitative aspect that counts and allows a descriptive approach of a great number of accident types.

3. What is representativity for ?

- The criterion of representativity must be assessed regarding to the objective of the information system :

    - if the first objective is the calculation of incidences, global morbidity statistics and the building of completely reliable indicators in public health, the criterion of representativity is major.

    - if the objective is to carry out pragmatic epidemiological studies based on a great number of cases and leading to concrete preventive measures as far as injuries are concerned, the criterion of representativity is less necessary. The IDB system is based on this attitude since it exists.

- *Consequently, the representativity demand shall not paralyse the IDB.* Let's clarify the situation: if some Member States provide representation through a methodology, we would be delighted. However, if they do not provide any methodology, or one that is incomplete, it is not valid to completely reject the system and maintain that it is void.

- For the moment efforts are made to introduce the calculation of incidence rates into the IDB, but as long as we are not able to provide statistically reliable confidence intervals, the methodology is questionable. When the calculation is possible, like in France for instance, the associated confidence intervals are very wide.

- We suggest articulating together the present hospital collection and the leading of "representative IDB surveys" allowing for "adjustment" of the results of the hospital collection (Project IPP 2002 "Harmonized survey" n° 2002/IPP/203701).

- **Until now, everything happened as if we had the wrong IDB goal (the calculation of national incidence rates – the "macro" approach) for which the methodology used (the collection through the hospital emergency unit) is not appropriate; we are hesitant to use the IDB data collected according to the appropriate methodology➔ the analysis of the major accidents' classes – the "meso" approach.**

4. The extension of the IDB system, an opportunity and a danger :
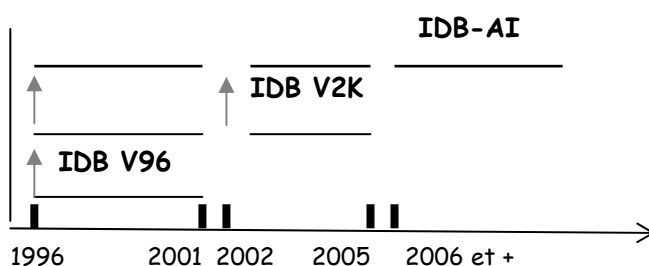
- The actors of the IDB HLA (Home and Leisure Accidents) had not finished to solve certain system problems (eg. coexistence of two collection methodologies, two encoding systems, non-uniformly checked data, inaccessible European database, etc.) yet the Commission urged these very actors to extend the system collection field. The new system should also take into account all the traumatisms, not only home and leisure accidents, but also traffic accidents, work place accidents and intentional injuries (suicides and attempts, as well as violence). It is the present IDB-AI (All Injuries) system.
- The extension is an opportunity because it spreads the system's field of competence and makes it more complete, but it is also a risk because the extension makes the collection heavier for the hospital staff, more expensive and scientifically more complex, because it mixes the collection of data about non-intentional and intentional injuries (see point n°17). It is important to leave the Member States (MS) free to choose the spread of their collection field according to their background and own means.
- Moreover, we must be aware that the differences between the States are and will remain strong in terms of importance given to the questions of prevention in the different fields, of used methodology, as well as cultural and ethical approach.

5. About the different positions of the Member States regarding the IDB :

- We must be aware that the MS have very different positions and interests regarding the IDB system. According to some MS, it is just a by-product of the pre-existing national system. Indeed, according to these States, IDB does not have any epidemiological interest; they almost never use the IDB data and barely know it. According to some other States, IDB is the only systematic collection system in that field and has therefore a great importance.
- One can legitimately give priority to the opinion of the States that actually use the IDB system, without ignoring the useful experience of more advanced States in the collection of information. But we cannot count on the former to use and promote the IDB data, because they prefer using the data of their own national system that they consider to be of better quality.

6. The necessity of a unique and controlled IDB base :

- Right now we potentially have three Injury database formats : the IDB V96 base (files 1986-2001), the IDB V2K base (files 2002-2005) and the future IDB-AI base (files 2006 and +). It is imperative and urgent to work on the unification of the bases in order to offer a unique, clean and checked database.
- That is our contribution : we checked and transcoded all the files IDB V96 from 1996 to 2001 into the IDB V2K format and we submitted a transcoding table from the IDB V2K coding to the IDB-AI coding. This is necessary but insufficient to rationally use the IDB.
- We must then upload into a unique database the controlled national files (validity control, but also coherence and likelihood controls – see point n°7) with an extra variable indicating the original coding (see point n°8) while keeping in mind that the newest data are the most homogenous.

7. The different types of control to make on the IDB base :

- The national teams are in charge of a minimum *validity control* on their own files. During the controls, we only check that the codes belong to the valid code list of the coding system. It is an automatic mono variable control. But it is also necessary to make *coherence controls*, and most important *likelihood controls*.
- A few coherence controls are possible with the IDB data (for instance: If "Treatment "= Hospitalisation, then the "Hospitalisation period" > 0 day). They are by definition multi variable. But it is above all necessary to introduce *likelihood controls*: For example, if all the codes of the variable "Activity" are in the Sport section in an annual file, there is obviously a mistake, despite the validity of each individual code. They are multi recorded controls based on the idea of the results' likelihood in view of the acquired experience. They could be done by comparison with data files of the previous years in order to detect the abnormally important gaps in the repartitions of the variables' modalities.
- We think that we also should perform the controls at the level of the European database before loading the cases by making automatic corrections without destroying any records.

8. Two quality indicators for each IDB cases :

- The easiest way is to associate with each case of the base (an accident) two quality indicators : one that takes into account *the quality of the case answer* when passing the different types of control, the other takes into account its original coding system (V86, V96, V2K, a national system, etc.) and the number of transcodings it has been through.
- We can then choose to work with the whole data or with data of a better quality according to the need of the study.

9. Focus on the exploitation of the IDB as it is :

- Many people emphasise the poor quality of the IDB base, its lack of representativeness, saying what would or will have to be done, when we should be promoting the contents of the database as it is now.
- It is surprising and regrettable that, among the many projects within the "Injury Prevention Programme" that the Commission finances, none of them ever was mainly about the exploitation of the IDB base through concrete epidemiological studies promoting the great qualities of the database while acknowledging the flaws.
- For 20 years European and national institutions have been spent a lot of energy, money, time and talent in the collection of data, their control and storage in databases and very few, too few, in the exploitation of the European database. In our opinion the IDB is "normally imperfect" if we agree to use it with the appropriate approach. Nonetheless, it is imperative to attach to the data a short and detailed description of the collection methods and the validation process used in each State, in order to establish the shortcomings without putting a strong strain on the results.

10. The current Public Access tool :

- The Public Access tool is a technically well designed tool and easy to use despite a few technical faults due to its youthfulness; but it is too restrictive to be really useful. We are on the wrong path when emphasising the calculation of incidence rates instead of the accessibility to the anonymous cases. We still prefer the "macro" point of view without seeing that the real efficiency of the IDB lies in the "meso" approach. A safety policy too strict, added to the inappropriate exploitation of data and a certain lack of quality due to insufficient controls, imply that the IDB Public Access data might not

be used enough, in our opinion, like during the unhappy experience of the HIEMS data. It is a pity since the IDB has many great advantages for an effective epidemiological exploitation.

11. Development and use of specific analyzing tools :

- During the previous studies we developed and used analysing tools specific to the IDB, such as the SSRD (Synthetic Score of Relative Dangerousness) that allows organising the relative dangerousness (in the IDB endogenous sense) of products in hierarchy, or the scenario method or even the Automatized Alert System (AAS) method. We recently spread the use of the SSRD in order to also allow organising in hierarchy the dangerousness of sports, mechanisms, activities, etc.
- The Dutch team of the CSI (Consumer Safety Institute) developed on its side a score that has the same goals – the Priority Score (PS) – but based on a hierarchy of experts' opinions and not characteristics of descriptive statistics coming from the whole studied cases.
- It is, in our opinion, a direction that the other network teams should take with us. The existence of 6 million data deserves that we question which specific tools are needed for extracting the most useful prevention information (i.e. data mining tools).
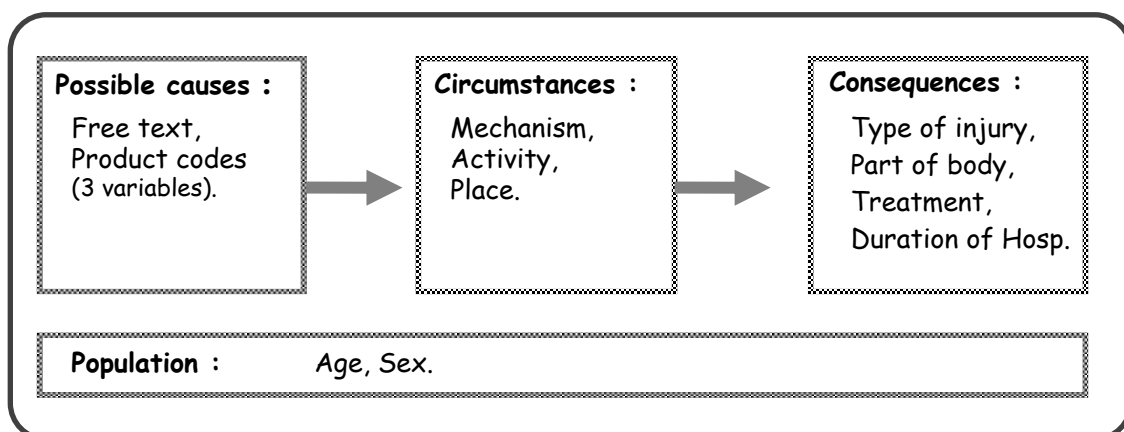
12. The importance of free text :

- In our opinion the free text is an important variable for two reasons at least. It allows :
    - Selecting the cases when, for example, there is no specific product code or in addition to this one. In France, more than two thirds of the specific studies are conducted from a selection of characters in the free text (f.e: accidents due to "Champagne corks") ;
    - Precisely analysing the selected cases in order to have a better knowledge of accident circumstances and/or details about the type of products involved in the accidents.
- Until now the information in the free text was scarcely used at a European level. We could think about developing a multilingual program of keywords of the free text's selection and translation for an automated exploitation.

13. The establishment of injuries' scenarios :

- The main variables of the IDB system can be dispatched according to their role in the chronology of the injury's sequence :

Injuries' scenarios within IDB

Possible causes :	**Circumstances :**	**Consequences :**
Free text, Product codes (3 variables).	Mechanism, Activity, Place.	Type of injury, Part of body, Treatment, Duration of Hosp.

Population :   Age, Sex.

- From this detailed description, it is imperative to spot injury scenarios by crossing three variables Mechanism X Activity X Place, by determining together the populations, the medical implications (as severity indicator) and the possible causes (products, consumers' behaviours, etc, by also using the free text). We already depicted this easy statistic methodology, without loss of information, in previous documents. The descriptions of these scenarios can be used in the public health sector as well as in consumer protection.

14. Public health and consumers safety :

- It seems essential to us to keep both axes of the IDB: the public health axis, since the IDB data allow the epidemiological analysis of major injury categories, and the consumer protection axis, since the IDB is the only European database that also contains information about the products possibly involved in the accidents.
- It seems to us that the IDB data are still underused in the approach "analysis of risk products". Even if this base cannot be seen as an alert system allowing identifying new dangerous products, it contains general information that is really relevant in terms of safety of product classes and consumer behaviours.
- We proposed to the DG SANCO B3 (Products and Services Safety) to set up some "*Flash IDB studies*" allowing a quick and simple answer to questions that the Commission departments have in terms of products safety or characterisation of injuries types.

15. A procedure of quick answers to questions of the Commission :

- The IDB data have existed for 20 years and for 20 years they are not being used at the European level. It is no longer about proposing another study lasting many months with results that will be carefully put away. *It is about setting up a procedure of quick answers* -that uses the current IDB data- to questions (captured) according to the entry: are horse riding accidents serious ? (severity entry); what are the most dangerous do-it-yourself tools ? (dangerousness of products entry); are a lot of children burnt by the water in the bath ? (quantity aspect entry), what are the most frequent injuries related to skiing ? (medical aspect entry); what population is victim of carbon monixide poisonings ? (population entry); what happens during a barbeque accident ? (scenario entry); etc.
- It is not about carrying out long epidemiological studies with a high scientific level; it is, for a small group of competent experts to quickly answer (less than 48 hours) and provide simple responses to easy questions that any kind of prevention actors could ask, notably and in priority to the Commission.

16. The confidentiality of data :

- First of all, we can ask : why so many questions about the confidentiality of data that nobody uses ? More seriously, it seems essential to us to have access to single-record data; the cases are naturally anonymised.
- Simple and well-known means exist to guarantee that the confidentiality is respected (no hospital code, no birth date, age grouped by 5 years, no date nor hour of the accident, etc.). We could profitably be inspired by the rules of Eurostat to access mortality data.
- The confidentiality problem must not be another obstacle to the data exploitation. It seems relatively easy to respect the criteria of confidentiality and to allow an interesting exploitation of the database (loading into a spreadsheet document – as Excel - with a limited number of selected cases but containing the whole variables and not possible loading for a number of selected cases < 5).
- For the public access, some precautions regarding data use and interpretation of results must be strongly emphasized for a non-informed public.

17. A distinct methodology for the collection of intentional injuries :

- It is obvious to us that the extension of the collection to intentional injuries, among them gender violence, introduces a break of nature in the IDB collection. The issue is not only the adoption of new technical elements (extension of the coding and nomenclature system), but, for the hospitals that participate to the IDB collection, *to take into account the new problematic, like gender violence, in a global strategy.*
- The extension of the hospital collection to injuries due to violence implies a strong redefinition of the reception and the treatment of the victims (staff training, specific protocol of collection, a separate room in the unit, help and useful information to the victim, ethical involvement of the staff, etc.), otherwise the under-estimation of the frequency and gravity of that kind of injury could have a counterproductive effect; as well an ill-adapted collection of information about violence cases may endanger the victims as much as the collection team. *We developed these points of views in a study of the DAPHNE 2004 programme (DG JLS) "Study of the extension of the European ISS collection on the injuries to the violence towards adolescents and women".*

18. For a new approach philosophy of the IDB :

This new philosophy is articulated around three main and complementary points :

- **Point 1** : *Recognise the "competence" of the current IDB data.* No information system will ever get to an absolute scientific rigour. There are always approximations, grey zones of uncertainty. But, the lack in statistic accuracy must not prevent us, under the excuse of "scientific rigour", from finding pragmatic solutions, "useful" point of views. **The present Injury database is in our opinion, as we already said, "normally imperfect".** We must consider it as a set of useful data and ask the question "what can we learn from these data ?" and not always the (legitimate but not productive on the long term) questions : "what is not possible to do with the data ?", "how should the data be ?", "how can we improve the data quality in the future ?", etc. The IDB data as they exist have a great informative potential that we can use now.

- **Point 2** : *Use the IDB data for the purpose they have been collected.* We have 6 million data available coming from hospital emergency units waiting to be used at their right "meso-accidentological" level. Let's do at last pragmatic studies from these data in order to have a better understanding and prevention- at a European level- of, for example :
  - Horse riding accidents
  - DIY accidents
  - Football accidents
  - Accidents due to alcohol (state of intoxication), etc.
  - Quad accidents
  - Children's burns
  - Fires

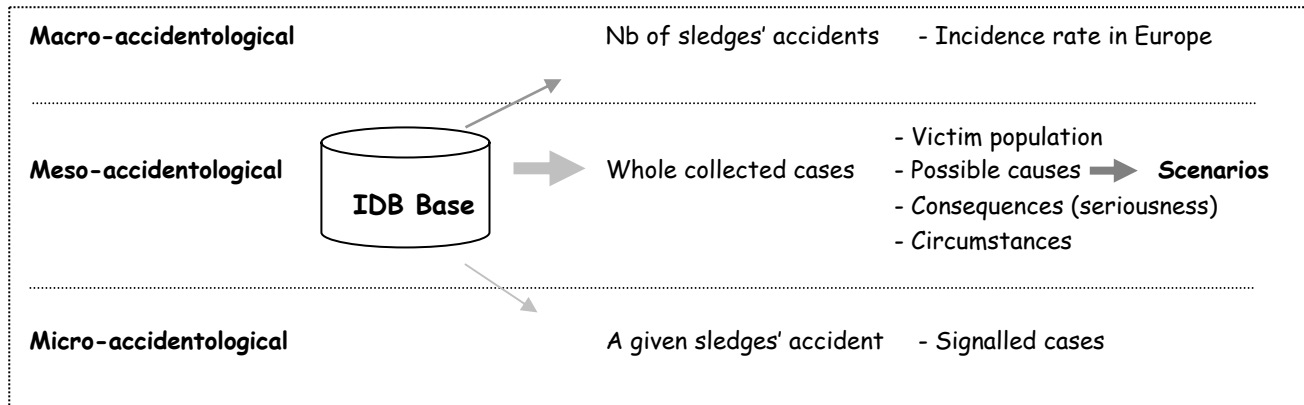*In most cases, the IDB base is the only European database containing information about the above topics.*

- **Point 3** : *Adopt a reactive logic of success. We cannot* be afraid to exploit the data. Our first goal is not to publish a detailed theoretical research work in scientific journals, *but to provide, particularly to the Commission, in a very reactive manner, practical and reliable information that do not exist anywhere else.* It does not prevent us from developing a whole strategy of data quality improvement or publishing scientifically oriented works.

➔ This is about building a new horizon to the Injury database by developing operating concepts (for example : "meso" approach) and adapted tools (for example : "scenario method") beyond a paralysing statistical hegemony.

19. Let's take an example, sled accidents :

- If, for example, we want to carry out a study about sled accidents the IDB base must be used in priority like a set of cases that allows determining scenarios of accidents and measuring the possible degree of implication of the product:

Example: sled accidents

Macro-accidentological		Nb of sledges' accidents	- Incidence rate in Europe
Meso-accidentological	IDB Base	Whole collected cases	- Victim population - Possible causes ➡ **Scenarios** - Consequences (seriousness) - Circumstances
Micro-accidentological		A given sledges' accident	- Signalled cases

20. About 10 easy actions to emphasize the IDB system :

At the end of these reflections, we would like to list 10 easy actions in order to promote the contents of the IDB in the previously depicted directions and definitively transform this quasi "data cemetery" into a living database :

1. Speed up the transcoding of the entire IDB files into a unique format (IDB-AI) and upload this with national controlled files : validity control, coherence control, but also likelihood control;

2. Make the base of anonymised cases more accessible in order to conduct many epidemiological studies on the data;

3. Develop specific and easy analysing tools emphasising the epidemiological content of the base (for example : the Synthetic Score of Relative Dangerousness, the Priority Score, the scenario method, etc. i.e. Data Mining tools);

4. Develop and use of a multilingual program of selection and translation of key-words in the free text for an automated exploitation;

5. Support projects inside the PHP centrally using the IDB data with adapted tools;

6. Better use the IDB data for the protection of consumers (for example : Flash IDB studies);

7. Set up a quick answer procedure, notably for the Commission, with the IDB;

8. Take into account the fact that the collection about intentional injuries, among them gender violence, is of another nature and consequently involves a distinct methodology.

9. Study the possibility to compliment the IDB hospital collection with representative IDB surveys, in order to solve the question of representativeness;

10. Favour the opinions of the actors who know and really use the IDB base and want to make it progress.

Contact : Marc Nectoux
Psytel – 33 rue du Commandant René Mouchotte
F-75014 Paris
Tel. : +33 1 40 46 17 79 / +33 6 07 62 45 84
e-mail : nectoux@univ-paris5.fr