Psytel

Ingénierie de l'information

On the subject of representativity in the IDB

Marc Nectoux (Psytel - F) and Dr Bertrand Thélot (Institut de veille sanitaire – F) / November 2007 / v3

This document revisits certain themes discussed in the previous document « 20 ideas and truths about the IDB » (December 2006), yet it also delves deeper into the notion of representativity in the IDB (Injury DataBase) system to conclude with simple recommendations and a repeated invitation to change our point of view on the IDB.

The question of representativity in the IDB system is as old as the system itself. Its critics have often used this argument, « IDB is not representative! », to deny any worth to the system, and a number of the participants have sought hard to create representativity in the data, but it has always escaped them. Others stated the IDB must become representative or be terminated, even though, it is clearly difficult to achieve. Yet the question of representativity in the IDB is not so complex, nor so crucial, when seen from the right angle.

1. The fundamentals of the system

Let us remember a few fundamentals of the system :

- In the first version the data collection system was called the European Home and Leisure Accident Surveillance System (EHLASS). Although it was not a surveillance system its official objective (Decision 86/138/CEE – 22/04/1986) was to "collect data on accidents related to consumer products in order to assist in the prevention of accidents, improve the safety of consumer produts and provide information and education to consumers for safe use of products ". The information was to be obtained from "services providing care to those injured". The project was for accidents taking place in the home or in the immediate environnment, as well as accidents during leisure, sports and school.

- The characteristics of the system were the following : an information system in the accident and emergency departments of hospitals for HLAs (Home and Leisure Accidents), oriented to the protection of consumers. It was thereafter reoriented, in its intent but not in the methodology, towards public health prevention. In this system twenty variables are collected concerning medical information and the circumstances of the accident, as well as products potentially causing the accidents. The methodological characteristics of the system have not fundamentally changes in the past 20 years, except for the recent extension to include all types of injuries (intentional and non-intentional, as well as workplace and traffic).

➔ The IDB system is, in essence and above all, a unique **bank of European cases**, with all the advantages and limitations which this entails.

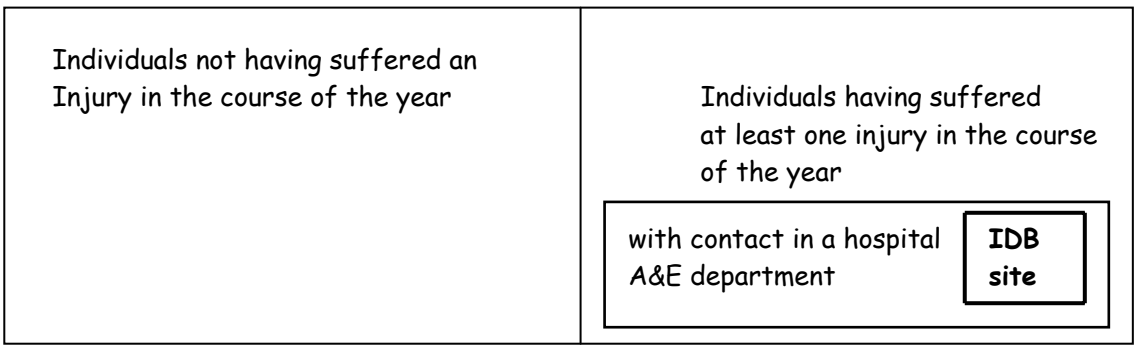
IDB is not : IDB is not a surveillance system, nor a register (as it is not an exhaustive data collection of all injuries but sample data), nor an alert system (there is no alert mechanism in the system). And it is not a strictly epidemiological data base : the data collection system is not representative regarding the chosen methodology, which will be discussed further below, and in its information on product categories it is for the protection of consumers.

IDB is : **IDB is a bank of European cases on large groups of injuries** (originally HLAs), collecting medical information and the accident circumstances, as well as the products potentially causing the accidents. This is the IDB and it is useful information !

2. Which representativity are we discussing ?

2.1. Non general representativity :

- Let us remember a simple fact. The IDB system is not capable of representing the entire spectrum of injuries («general representativity») for the simple reason that the system is fundamentally a collection of cases from hospital A&E department services. A case is based on an individual from a participating Member State having suffered, yes or no, an injury in the course of the year :



- Therefore the IDB system can only represent injuries which have had contact with hospital A&E departments.
- « General representativity » could only be achieved by *surveys limited in time and based on a representative sample of the population.* Even with this method important methodological biases may exist due to memory recall, use of telephone surveys, the inability to take into account deaths, etc.

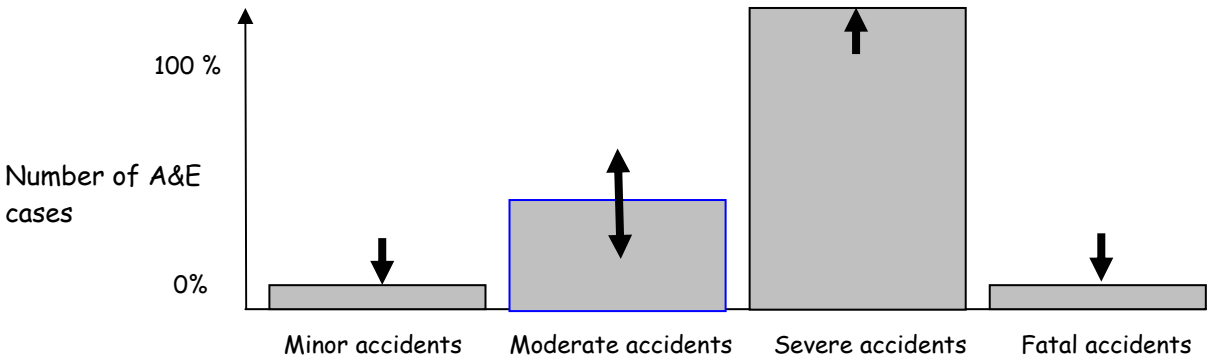
2.2 Non representativity in terms of severity :

- Amidst the scope of injuries which exist it is necessary to keep in mind that only injuries of a certain severity level are collected in a system based on hospital A&E departments :

- fatal accidents «on the spot» (eg. fall from a ladder, drowning, etc.) do not transit thru the hospital A&E departments; they are directly sent to medical examiners. Therefore the most severe accidents are not included in the IDB, meaning it does not provide a clear or precise profile of fatal injuries !

- on the other end of the spectrum, minor injuries (eg. small cuts, minor burns, etc.) also do not transit thru hospital A&E departments, but are seen by general practitioners, pharmacists, nurses or by the families or victims themselves. Therefore these are also not included in the IDB system.

- within the breadth of accidents of «moderate severity» contact with a hospital depends on a variety of factors : national healthcare structure (many or few general practitioners, free or not, etc.), geographical proximity, attractiveness and reputation of the hospital. These factors vary from one country to another, from one hospital to another, such that even for these specific accidents the variability may be high. Using a severity scale which is subjective but relatively practical, we are able to distribute the number of A&E cases in the following manner



➔ In summary, even if we had the representativity in terms of hospital recruitment, we would not have it in terms of severity.

2.3. Specificity of certain accident types collected :

- The choice of the Béthune Hospital (hospital of medium size near the city of Lille in the north of France) led us to have in the French IDB a certain number of «ski on artificial surface» accidents, due to the presence of ski slopes near Béthune. In this case we are not referring to « representativity » but rather to very specific types of accidents, atypical in general, which are present in the system due to the geographical proximity to the collection site. In statistical terms these « aberrations » are always possible and should be avoided for reasons of representativity.

- Let us take another example which is not so unique. If we wanted information on underwater diving accidents it would be imperative to have a hospital collection site near a location which allows for this sport. A random recruitment of hospitals would most certainly not allow for collection of these accidents. That is the reason why the majority of Member States practice the choice of *reasoned* collection sites.

2.4. Randomisation for the recruitment of hospitals impossible to achieve :

Three types of arguments exist as to why the randomisation for the recruitment of hospitals is impossible to achieve in the IDB system regarding representativity :

- Those who have had contact with the administrators of the national IDB systems (NDAs : national data administrators) who are in charge of hospital recruitment for the IDB know that the choice and recruitment of teams is a difficult and absorbing task. Above all it is necessary that the site teams are highly involved in the system. Thus it is necessary to motivate each unit at the same time, the team at the hospital (those who sign the contracts), the medical A&E team (those who authorise and supervise the data collection) and the service administrators (those in charge of data collection). If one of these teams lacks motivation the entire structure is in danger. Therefore each level of personnel at a collection site must voluntarily and willingly take part in a public health action for the long term, in addition to caring for victims which is the main priority. This task is not simple, especially in the context of an A&E department where staff is often overburdened with work. Therefore it is imperative that the sites are really « voluntary » !

- If we wanted to apply randomisation to the entire A&E departments of a Member State, with a confidence interval of sufficient precision, it would be necessary to recruit a very large number of hospitals in those Member States of high population. Let us take the example of France which has approximately 580 A&E departments. If we were working with a sufficiently high level of statistical precision, they are certainly several tens of hospital which it would be necessary to include in the system. This greatly exceeds the financial and management capabilities of a Member State. Naturally the case is different for Malta or Cyprus.

- Finally, if we wish to have a diversified recruitment of *a priori* all major categories of accidents and population types, one would make a reasoned choice which would not necessarily produce the same effect as randomisation. This reasoned choice allows for diverse recruitment in terms of geographic criteria (south/north, mountainous/sea, etc.), types of population (large urban centres, rural zones) and the size of hospitals (large and small hospitals). This reasoned choice is applied in France for the recruitment of 12 hospitals in the system.

➔ Randomisation of hospitals is impossible to achieve and actually not desirable. Thus the IDB will never be representative in the pure statistical sense !

2.5. «Partial» representativity :

- With the system of incidence calculations based on the use of « catchment area », meaning the zones surrounding the hospital, we are able at best to obtain incidence rates for large classes of accidents or population groups, using the national hospital data collection systems such as hospital discharge registers. These hospital incidence rates are based on a number of hypotheses. The transition from incidence rates from each hospital being used to devise a national incidence rate remains problematic as it relies on the assumption that the results obtained from the hospital are representative of those of the entire country.

- The confidence intervals that we can technically apply only represent a part of the uncertainty related to the estimations of incidence rates, with the rest remaining non quantifiable . The calculation of national incidence rates relies on the application of the old and respected << rule of threes >> then to a sophisticated statistical method. As Alfred Sauvy said many years ago : « the preciseness of the decimal often masks the incertitude of the number >>.

- Certain Member States maintain that their national IDB system is representative as a national data collection system. Nevertheless the European system is not representative, even partially, as it is clearly not representative in many Member States, including France.

➔ In summary, the European IDB system is not and will not be representative in any way using the hospital A&E methodology which was not made for this purpose!

3. Representativity is not a central concept of the IDB system

- In this chapter we would like to show that representativity is not a central concept of the current IDB system. IDB provides very useful information on non-representative data when taking into account its levels of efficiency :

3.1. The concept of « levels of efficiency » :

We define this concept according to two axes :

Axis 1 – Accident types (general versus specific) :
- Let us distinguish between «general» accidents which are *a priori* rather frequent and spread throughout a country (eg. accidents related to sports such as football, etc.) and geographic specific accidents or seasonal (eg. ski accidents, underwater diving, etc.).

Axis 2 – Type of approach :

- Let us recall what was elaborated in the document «20 ideas and truths about the IDB». There are three different approaches for an injury information system :

- the *« macro-accidentological »* approach is related to calculations of incidence rates globally, aggregated representative results and building national epidemiological data systems; ensures representativity but does not provide richness of detail for analysing different accident types, nor does it require a great number of cases.

- the *« meso-accidentological »* approach allows for identifying populations, situations, behaviours or products at risk with regards to precise injuries, as well as defining classes of accident scenarios; hospital A&E department data collection is an example which yields a great number of injury cases, integrated into validated medical data, yielding descriptive data on a wide variety of injuries. The quantitative aspect is primary and representativity secondary.

- the *« micro-accidentological »* approach is to identify rare accidents or a very specific product which is potentially dangerous of a specific type or manufacturer; such as the European RAPEX system.

- For each of these approaches there is a corresponding methodology and different analysis tools. Thus for the «macro» approach the most appropriate methodology would be to gather data via a population based study for which the statistical quality of the methodology is imperative. The number of accidents actually collected would be relatively little in terms of persons interviewed and would not allow for detailed analysis of different types of accidents but the representativity would be assured. For the «micro» approach one would give preference to studies on direct case information, as in the European alert system. For the « meso » approach the methodology of data collection in hospital A&E departments is perfectly appropriate as it allows for a continuous collection of a large number of injuries, within integrated, validated medical data. The representativity factor of the data are more secondary as the quantitative aspect is of primary importance («notion of a bank of cases»), providing a descriptive approach of a large number of types of accidents.

- We define the levels of efficiency as a combination of Type of approach and Accident types, as follows in the IDB :

Levels of efficiency in the IDB

Approach ➔	Macro	Meso	Micro
General accidents	Efficiency +	Efficiency +++	Uneffective
Specific accidents	Uneffective	Efficiency ++	Uneffective

➔ The IDB system performs particularly well on information related to general accidents within a «meso approach» and less well when dealing with specific accidents on a micro approach.

3.2. Specific examples :

- Equestrian accident : these are *a priori* frequent enough and the sport practised relatively everywhere. In a « meso » approach we could use the IDB to have a good idea of the total number of these accidents (also an idea of the number at the national or European level), of their relative severity and the principal (major) accident scenarios involved ➔ Efficiency +++

- Underwater diving accidents : these are specific to certain geographical zones. In a « meso » approach we could have an idea of their total number (nationally and at the European level), but only have an idea of their relative severity and accident scenarios if we had the hospital data collection sites in the vicinity of the geographic zones ➔ Efficiency +

➔ One must accept that the IDB system provides us with a particular picture, a certain point of view which is non-representative on accidents, but this point of view is far from being uninteresting as it

combines a multitude of usable information for improving prevention, if used while keeping in mind the efficacy levels.

3.3. Reasoned approach :

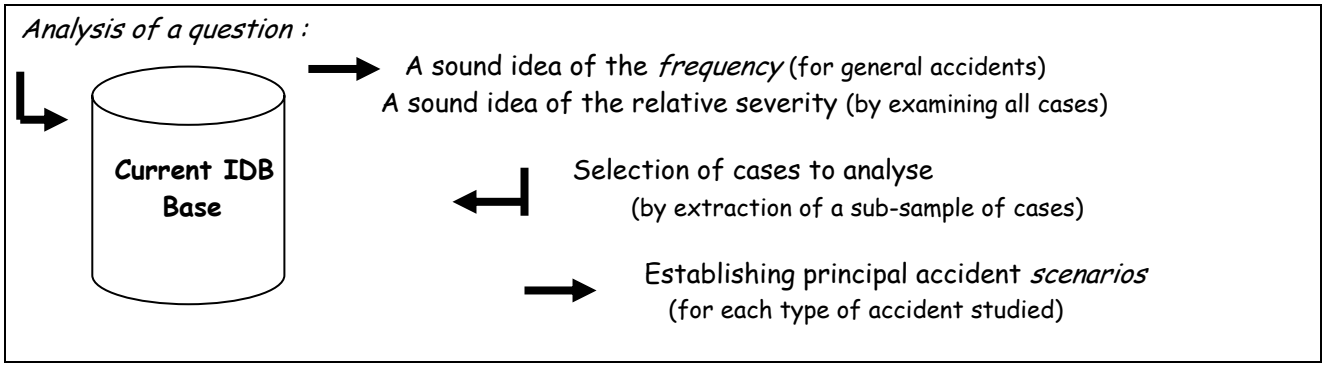
- Usage of the IDB requires a *reasoned approach* based on a strong knowledge of the system and its contextual information, based on three factors :

- accident frequency for general accidents: from rare (< 10 cases per 100 000), frequent (> 1000 cases per 100 000) ;

- relative severity using severity tools such as the Synthetic Score of Relative Dangerosity (SSRD in France, PS in the Netherlands);

- identifying principal accident scenarios using the combination of all of the IDB variables, including the free text (note : we have already developed a simple methodology for producing accident scenarios in the previous document (« 20 ideas and truths about the IDB ») based on the IDB variables. These scenario descriptions are useful for public health efforts in consumer safety).

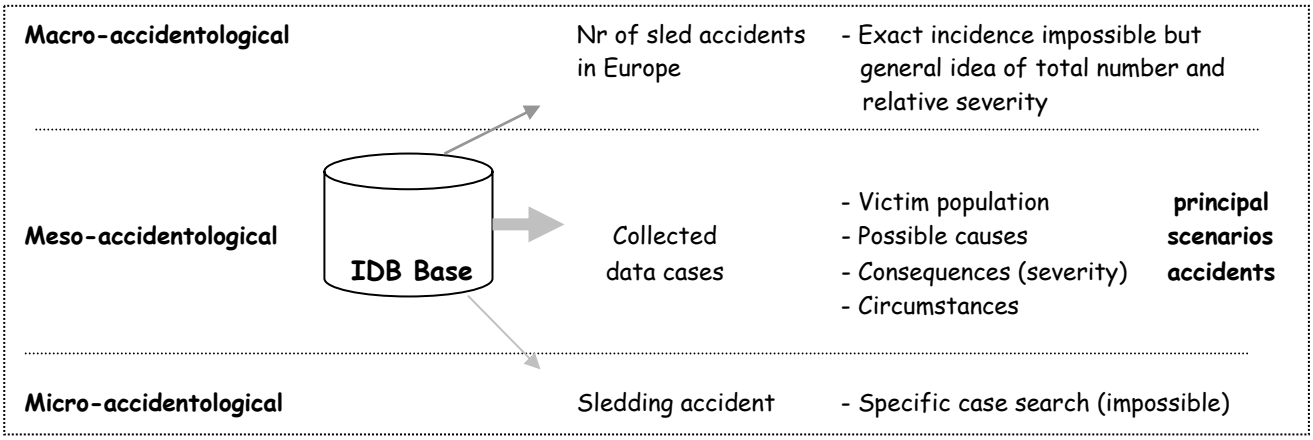
- Let us provide a graphical representation of this potential strategy :



Note : we have already developed a simple methodology to derive accident scenarios from the IDB variables (***population*** : age, sex - ***circumstances*** : mechanism, activity, location - ***consequences*** : injury, body part, type of treatment, length of stay - ***possible causes*** : products and free texts) in the previous document (<<20 ideas…>>). The descriptions of the scenarios can serve the public health sector as well as the consumer protection sector (approach would be <<class of product>>).

3.4. Let's look again at the example of sled accidents :

- If, for example, we would like to conduct a study on sledding accidents in Europe, the IDB would have to be used in principle as a bank of cases allowing us to determine the principal accident scenarios and to measure as far as possible the degree to which the sled as product was involved :



➔ IDB was not created to support high-levl scientific studies or theoretical or epidemiological studies. It was created to respond to a certain type of questions, inspired by or emanating from a practical and operational approach based on the information it contains; with the goal to provide concrete prevention measures related to injuries. It also allows for building a cartography of large classes of accidents.

4. Possible improvements

Certainly improvements are necessary and possible in the functioning of the IDB system. We propose improvements along each approach :

4.1. Improve the macro-accidentological approach :

- To answer the question of « general representativity » it is absolutely necessary to perform population studies in addition to the IDB, as a specific European IDB population study (same questions and nomenclatures as in the permanent data collection) or by developing an « IDB module » within the framework of the European Health Interview Survey by Eurostat.

- With this complementary methodology performed every 5 years for example the relatively rare types of accidents will not be included, but we will have a more exact knowledge of the incidence rates for large accident classes. This would be preferable to increasing the number of data collection sites as this would not significantly improve representativity.

4.2 Improve the meso-accidentological approach :

We propose the following improvements :
- Establish and harmonise quality control procedures on-site : this procedure is applied in the French system whereby each year a site visit takes place at random for usually one day per month, to compare the A&E registration documents with the IDB files. In this way we control for coding quality and exhaustiveness of the data collection.

- Plausibility controls : in addition to the validity (mono-variable) and logic controls (multi-variable) we propose plausibility (likelihood) controls for annual data files, to compare the distribution frequency of each variable between year N and year N-1. These controls would enable us to identify important changes and incoherences in the data which the other controls do not show.

- Create an IDB base of excellence - *IDB++ :* we propose creating a sub-database of high quality data from Member States who uphold precise quality controls for studies that require high quality but not necessarily a high number of cases. The aim is for all countries to evolve in order to reach this level as soon as possible.

4.3. Improve the micro-accidentological approach :

- We propose establishing the European Alert System on Injuries (EASI) : a system for signalling when a rare, new, noteworthy accident occurs or when high increases are seen in an existing accident type and to search for comparable accidents in other Member States.
- This system is different to the Commission's RAPEX (Rapid Alert System for Non-food Products) : EASI does not only involve product alerts; it would alert on certain activities or behaviours that are potentially dangerous as well, without administrative implication.
- This could be done by automatic signalling using distribution frequencies of certain variables between two time periods in the IDB.

5. Conclusions

C1. The «general» representativity will only be improved by extending the current data collection system in the hospital A&E departments with a «European IDB population study» or a module of studies by EHLIS coordinated by Eurostat. Without this the representativity, even limited, will never be achieved with the IDB as it is not in the central concept of the system, which is above all a **European bank of cases.**

C2. Let us start by analysing the millions of cases which are in the IDB system and use the information they contain for large European studies, rather than discussing what the IDB could have or should have been! This database does not include all the information on all types of injuries and all types of products, but if we know how to interpret the findings well, we will see that it contains information which exists nowhere else. Thus let us try to **understand and value the real richness of the data that we have been collecting for years!**

C3. Certainly improvements are necessary and possible. We propose the following improvements based on the three approaches :

➔ concerning the "macro" approach : **articulation of IDB with an European population survey** for « general representativity » or a survey in the framework of the European Health Interview Survey (EHIS) coordinated by Eurostat.

➔ concerning the "meso" approach : improve **quality control** (homogenous quality controls on the IDB collection sites and introduce plausibility (likelihood) controls on national data files) + **improved IDB analyses** (eg. flash studies, large European studies based on the IDB, etc.) + create an **IDB++ base of excellence.**

➔ concerning the "micro" approach : develop **a notification system, European Alert System on Injuries (EASI).**

C4. More globally, we recommend changing our point of view on the IDB : currently we are trying to establish a ***specific IDB semantic*** to define operating concepts (eg. the *reasoned choice* of IDB sites, annual *plausibility controls* of the data, *levels of efficiency* by accident types, determining principal *accident scenarios*, *injury trends*, etc.). What we propose is to discuss the « IDB from the inside » starting from the foundation rather than from the outside which is an administrative/technocratic perspective.

C5. Analysis of IDB data requires a **reasoned *approach*** utilising the contextual information at the hands of data experts, based on three factors : frequency, severity and accident descriptions. Thus it is not a «push button » system which automatically provides results without due consideration or a proper perspective.

- To conclude, let us stop thinking that we can reach a (limited) «general representativity » using the current methodology which was clearly not conceived for such a purpose; and let us stop thinking that nothing is possible without this sacred representativity !
- Concerning the «general» accident class we are able to have a good idea of their number, severity and we are able to describe the principal accident scenarios for each of these classes, and this thanks to the IDB. This is very useful information. What are we waiting for to develop and support this practical and operational approach !

Contacts : *Psytel - Marc Nectoux*
Tél. : (33 1) 40 46 17 79 *nectoux@univ-paris5.fr*
InVS – Dr Bertrand Thélot
Tél. : (33 1) 41 79 68 75 *b.thelot@invs.sante.fr*