

		
<h1>Comment juger de la qualité d'une base de données ?</h1> <p>Pr. L. Rachid Salmi</p> <p><i>Institut de Santé Publique, d'Épidémiologie et de Développement (ISPED) INSERM U-1219 Bordeaux Population Health (IETO) Université de Bordeaux et CHU de Bordeaux</i></p>		
19 juin 2019		CMFT, Paris, 1

<h2>Structure de la présentation</h2>		
<p><b>Éléments nécessaires pour décrire une base de données ou un système d'information</b></p> <p><b>Évaluer la pertinence et la qualité d'une base de données ou un système d'information</b></p> <p><b>Sources</b></p> <ul style="list-style-type: none"><li>✓ Département des maladies chroniques et des traumatismes, Institut de veille sanitaire. La surveillance épidémiologique des accidents de la circulation. Propositions et détermination des priorités. Saint-Maurice (France): Institut de veille sanitaire; 2005.</li><li>✓ Centers for Disease Control. Updated guidelines for evaluating public health surveillance systems. MMWR. 2001;50(RR13):1-35.</li><li>✓ World Health Organization. Data systems. A road safety manual for decision-makers and practitioners. Geneva, Switzerland: World Health Organization; 2010.</li></ul>		
19 juin 2019		CMFT, Paris, 2

## Pertinence de la réflexion



### **INVS, 2005**

- ✓ Connaissance biaisée de l'accidentalité routière
- ✓ Focus sur décès sans considérer la gravité
- ✓ Focus sur l'usager et comportements
- ✓ Contribution respective des différents facteurs

### **OMS, 2010**

- ✓ Terminologie partagée
- ✓ Améliorer l'exhaustivité
- ✓ Fournir des données détaillées (type, lieu, temps)
- ✓ Résultats disponibles opportunément

19 juin 2019

CMFT, Paris, 3

## Éléments nécessaires à la description de la base



- Quel est l'objectif ?**
- Quels sont les phénomènes mesurés ?**
- Quelles sont les sources d'information ?**
- Quelle est la période d'observation ?**
- Quelle est la population d'étude ?**
- Quelles sont les modalités de recueil ?**
- Quels sont les modes de transmission ?**
- Comment les données sont-elles analysées ?**
- Comment les résultats sont-ils diffusés ?**

19 juin 2019

CMFT, Paris, 4

## Quel est l'objectif ?

In summary, the transport sector needs data in order to:

In summary, the health sector requires data in order to:

- estimate the magnitude of fatal and non-fatal road traffic injuries;
- identify the risk factors involved so that health promotion programmes can target them;
- evaluate the effectiveness of injury management and treatment;
- ascertain current trends and the impact of prevention programmes;
- plan effectively for trauma care and rehabilitation services;
- plan and advocate for appropriate policies and legislation.
- plan and advocate for appropriate policies and legislation.

## Quels sont les phénomènes mesurés ? OMS, 2010

TABLE 1. Characteristics available for linking of road traffic injury cases in police and insurance data sets, Dominican Republic, 2010

Characteristic	Insurance (n = 5 176)		Police (n = 5 047)	
	No.	%	No.	%
Used in linking				
Collision date/hour	5 176	100	5 047	100
Collision type	5 176	100	4 889	97
Name	5 176	100	4 910	97
National identity number	5 176	100	1 513	30
Age	5 176	100	4 352	86
Gender	5 176	100	5 047	100
Injury outcome	5 176	100	5 047	100
Road user type	5 176	100	4 602	91
Road type	5 176	100	4 200	83
State	5 176	100	5 034	99
Vehicle type	5 176	100	4 864	96
Not used in linking				
Citizenship	5 176	100	5 047	100
Vehicle brand	5 176	100	2 819	56
Vehicle model	5 176	100	1 171	23
Date in medical services	5 176	100	0	0
Date out of medical services	5 176	100	0	0

## Quelles sont les sources d'information ?



**Services de police et de gendarmerie**  
**Justice**  
**Autres services de l'état**  
**Professionnels et établissements de santé**  
**Assurances**  
**Autres acteurs**

19 juin 2019

CMFT, Paris, 7

## Quelle est la population ?



**Notion de population à risque**  
**Individus, place, temps**  
**Numérateur seul ou dénominateur**  
**Représentativité**

- ✓ Exhaustivité
- ✓ Échantillonnage statistique
- ✓ Échantillonnage opportuniste

19 juin 2019

CMFT, Paris, 8

## Quelles modalités de recueil ? Quelles modes de transmission ?



### Systèmes passifs

- ✓ Informations issues spontanément de la périphérie sont recueillies par une structure

### Systèmes actifs

- ✓ Informations sollicitées ou recueillies auprès de la périphérie par la structure de coordination

### Avantages et inconvénients

- ✓ Capacité à détecter les cas : actifs > passifs
- ✓ Coûts : passifs < actifs
- ✓ Rapidité et rythme de la transmission

19 juin 2019

CMFT, Paris, 9

## Comment est faite l'analyse ? Quelle diffusion des résultats ?



### Analyse

- ✓ Utilisation standardisée d'outils statistiques
  - Calculs de l'incidence ou de la prévalence
  - Comparaisons observés / attendus
  - Analyses chronologiques / spatio-temporelles
- ✓ Définition d'un rythme d'analyse
- ✓ Définition de critères d'interprétation

### Cibles prioritaires de la diffusion

- ✓ Fournisseurs de données
- ✓ Décideurs

19 juin 2019

CMFT, Paris, 10

## Évaluation d'un système d'information



**Comment juger s'il remplit son rôle ?**

**Critères explicites de jugement**

- ✓ Degré d'atteinte des objectifs
- ✓ Utilité et coût du système d'information
- ✓ Applicabilité du système d'information
- ✓ Qualité des données obtenues
- ✓ Opportunité de l'information

**SMART**

**Specific  
Measurable  
Applicable  
Realistic**

**Référentiels**

- ✓ CDC / OMS

19 juin 2019

CMFT, Paris, 11

## Utilité et coût du système



**Le système est utile si**

- ✓ Détecte les tendances séculaires et les clusters
- ✓ Stimule la recherche épidémiologique
- ✓ Identifie les facteurs de risque
- ✓ Permet l'évaluation des actions de prévention
- ✓ Conduit à l'amélioration des pratiques

**Le coût du système**

- ✓ Coûts directs : personnel, autres ressources
- ✓ Coûts indirects générés par la détection des cas ou des clusters

19 juin 2019

CMFT, Paris, 12

## Applicabilité du système



### Le système est-il simple ?

- ✓ Figure décrivant les lignes de communication de l'information

### Le système est-il flexible ?

- ✓ Facilement adaptable à un autre problème ?
- ✓ Peut-il être transporté dans une autre population ?

### Le système est-il acceptable ?

- ✓ Pour les participants / acteurs du systèmes
- ✓ Pour le public
- ✓ Pour les pouvoirs publics

19 juin 2019

CMFT, Paris, 13

## Qualité des données obtenues (1) Exactitude (1)



### Sensibilité du système (Se)

- ✓ capacité à identifier les sujets avec le phénomène
- ✓ rapport du nombre de sujets qui sont détectés par le système et qui ont le phénomène (vrais positifs) sur le nombre total de sujets qui ont le phénomène dans la population surveillée

### Spécificité du système (Sp)

- ✓ capacité du système à identifier les sujets qui n'ont pas le phénomène
- ✓ rapport du nombre de sujets sans le phénomène et qui ne sont pas identifiés (vrais négatifs) sur le nombre total de sujets qui n'ont pas le phénomène

19 juin 2019

CMFT, Paris, 14

## Qualité des données obtenues (2) Exactitude (2)



### Valeur prédictive d'une détection (VPP)

- ✓ proportion des détection qui correspond à des vrais phénomènes
- ✓ rapport du nombre de vrais positifs sur le nombre total de positifs
- ✓ avantages
  - correspond à l'interprétation des résultats
  - seules données disponibles
  - ne nécessite pas de croisement avec d'autres sources de données, comme pour le calcul de la sensibilité

19 juin 2019

CMFT, Paris, 15

## Qualité des données obtenues (3) Exactitude (3)



Estimation of the number of traffic-injured persons in la Réunion Island in June 1994 by the capture–recapture method

Recorded by RIS*	Recorded by police data files		
	Yes	No	Total
Yes	34	103	137
No	43	NA	
Total	87		346 <sup>†</sup>

\* Réunion Injury Study.\*

**Aptel I, Salmi LR, Masson F, et al.  
Accid Anal Prev. 1999;31(1):101-8.**

<sup>†</sup> Estimated by capture–recapture method (Appendix A).

19 juin 2019

CMFT, Paris, 16

## Qualité des données obtenues (4) Autres aspects



### Fiabilité du recueil des données

- ✓ Capacité du système à donner des résultats reproductibles
- ✓ Capacité du système à différencier les variations du phénomène observé des autres sources de variation (variabilité inter-observateur, variabilité intra-observateur, mesures répétées)
- ✓ Standardisation des définitions et du recueil

### Exactitude et fiabilité des variables mesures

### Disponibilité de dénominateurs

### Stabilité de la qualité des données dans le temps

19 juin 2019

CMFT, Paris, 17

## Opportunité de l'information



**L'information produite par un système d'information doit être fournie en temps voulu aux personnes qui en ont besoin**

### Personnes cibles

- ✓ Personnes qui contribuent au bon fonctionnement
- ✓ Les preneurs de décision
- ✓ La communauté professionnelle concernée

### Conditions

- ✓ Utilité de l'information
- ✓ Délais courts / réactivité du système

19 juin 2019

CMFT, Paris, 18

## Application de ces critères



### Calendrier de l'évaluation

- ✓ Au moment de la conception
- ✓ Au moment des renouvellements du financement
- ✓ A intervalles réguliers

### Évaluateurs

- ✓ Responsables scientifiques du système
- ✓ Financeurs
- ✓ Équipe extérieure d'évaluation