

Modélisation Statistique et Biomathématiques en Médecine

Jean Gaudart

jean.gaudart@univ-amu.fr

Faculté de Médecine
Aix-Marseille Université

1. Introduction à la modélisation
 - 1.1 Qu'est-ce qu'un modèle?
 - 1.2 Démarche diagnostique et prise de décision en médecine
 - 1.3 Critères de causalité

2. La modélisation mathématique en médecine
 - 2.1 Modèles compartimentaux
 - 2.2 Epidémiologie mathématique

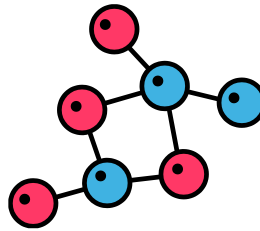
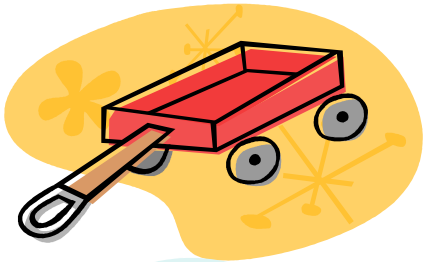
3. La modélisation statistique en médecine
 - 3.1 Introduction
 - 3.2 Analyse multivariée

I. Introduction à la modélisation

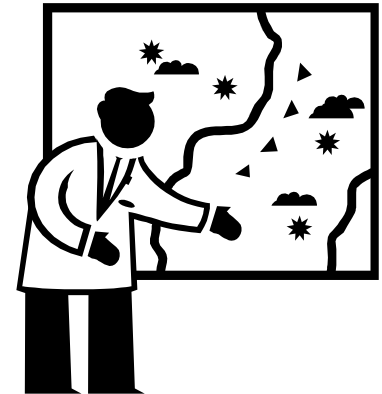
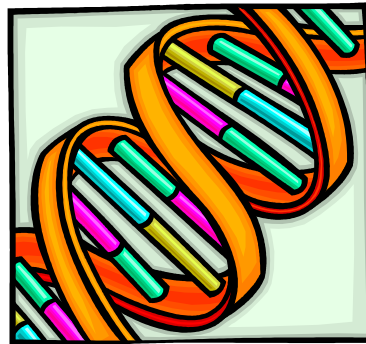
I.1 Qu'est-ce qu'un modèle?

- Représentation simplifiée de la réalité
- Dont l'objectif est de comprendre les causes et les mécanismes à l'origine d'un phénomène
 - ⇒ Décrire (résumer)
 - ⇒ Expliquer (comprendre)
 - ⇒ Prédire (simuler)

- Exemples



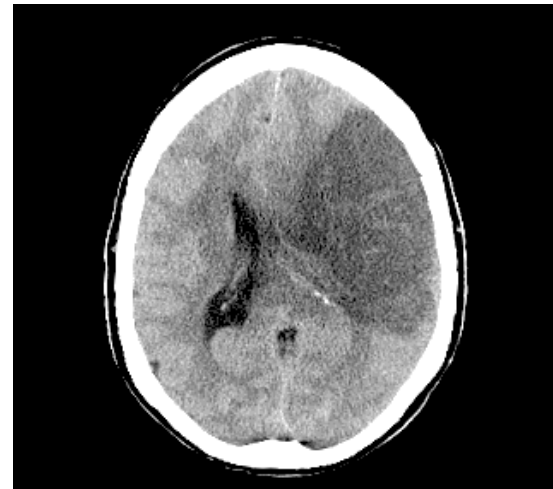
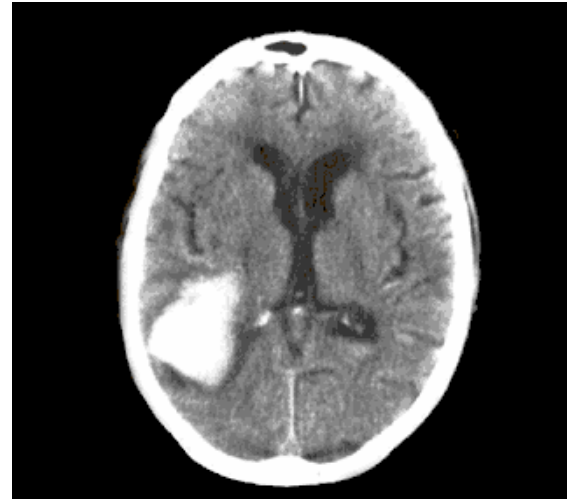
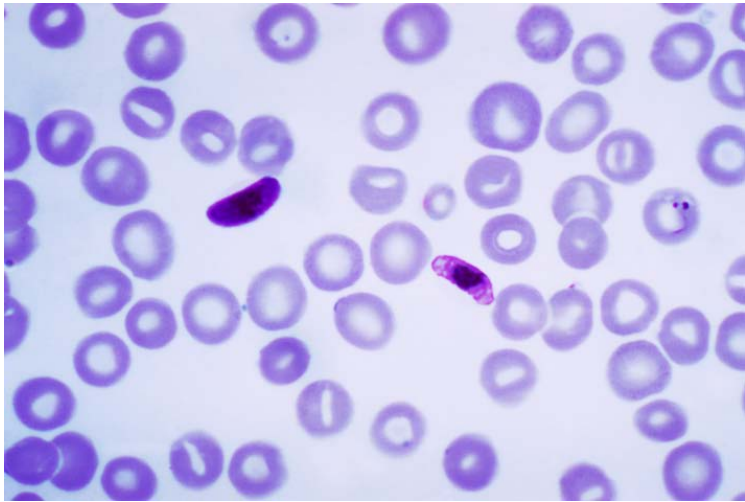
$$E=mc^2$$



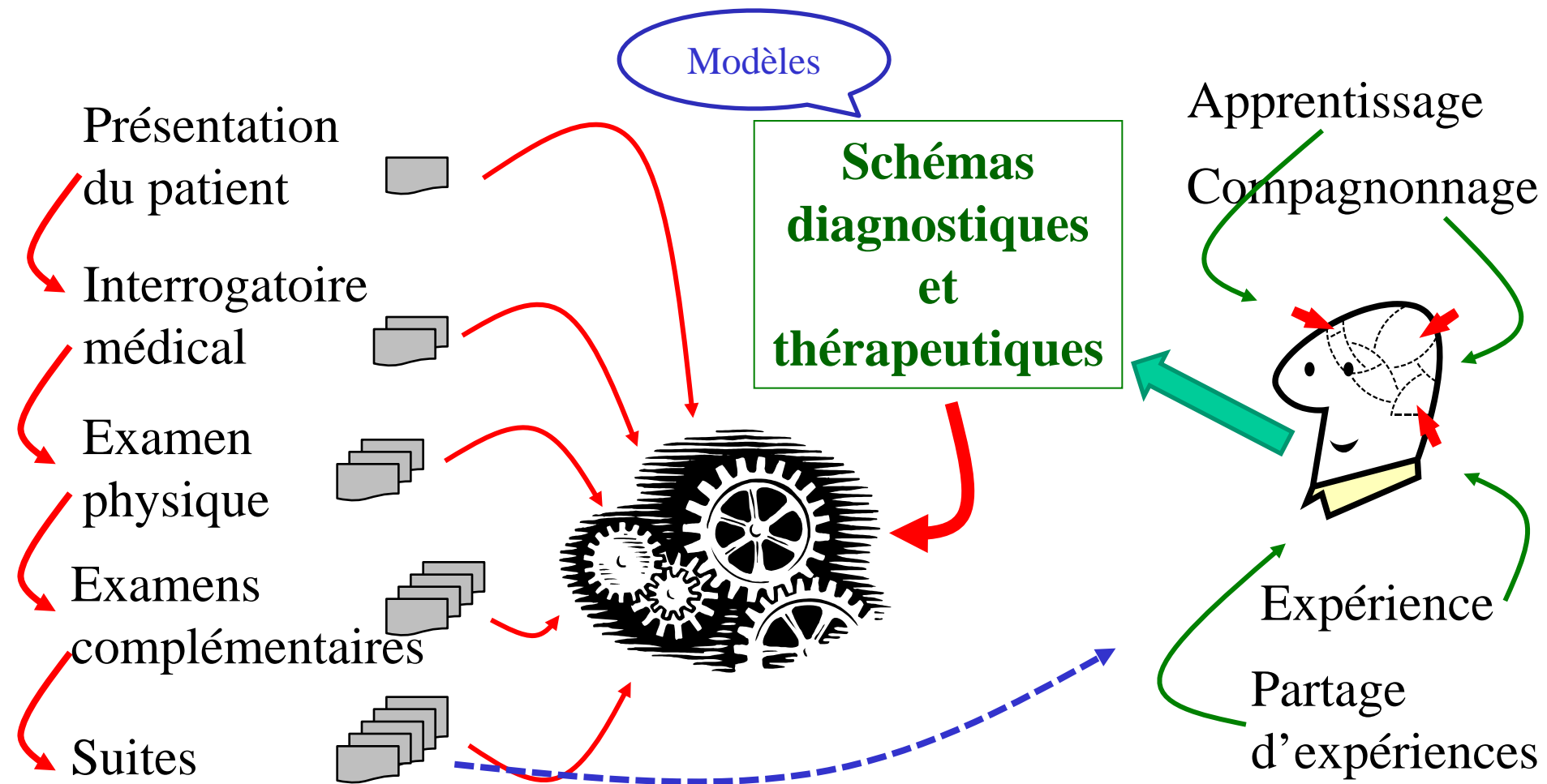
- Un modèle est toujours faux : simplification utile de la réalité
- Un modèle simple est encore plus faux, un modèle compliqué est inutilisable
 - ⇒ compréhension exacte d'un phénomène
 - Vs
 - ⇒ identification des mécanismes clefs et reproduction des observations

I.2 Démarche diagnostique et prise de décision





- Situation d'incertitude / Schéma diagnostiques et thérapeutique



Autres exemples

Raideur de
nuque

Céphalées

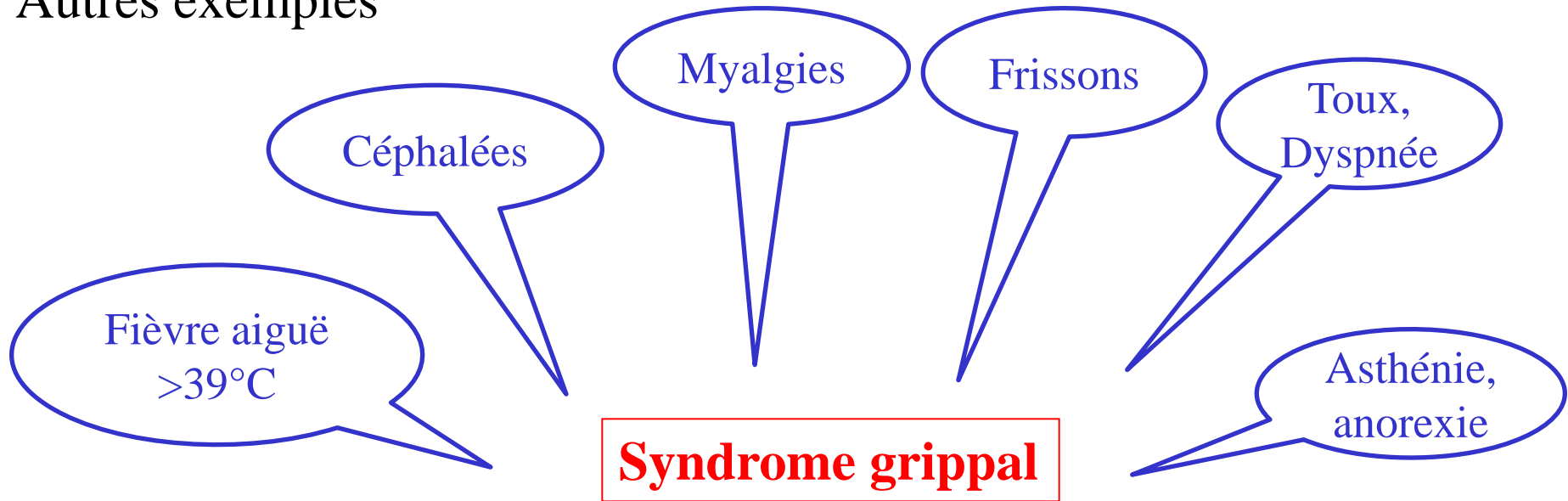
Vomissements

Syndrome méningé

Chez l'enfant?

Changement de contexte \Rightarrow changement de modèle

Autres exemples



Hors contexte d'épidémie hivernale?

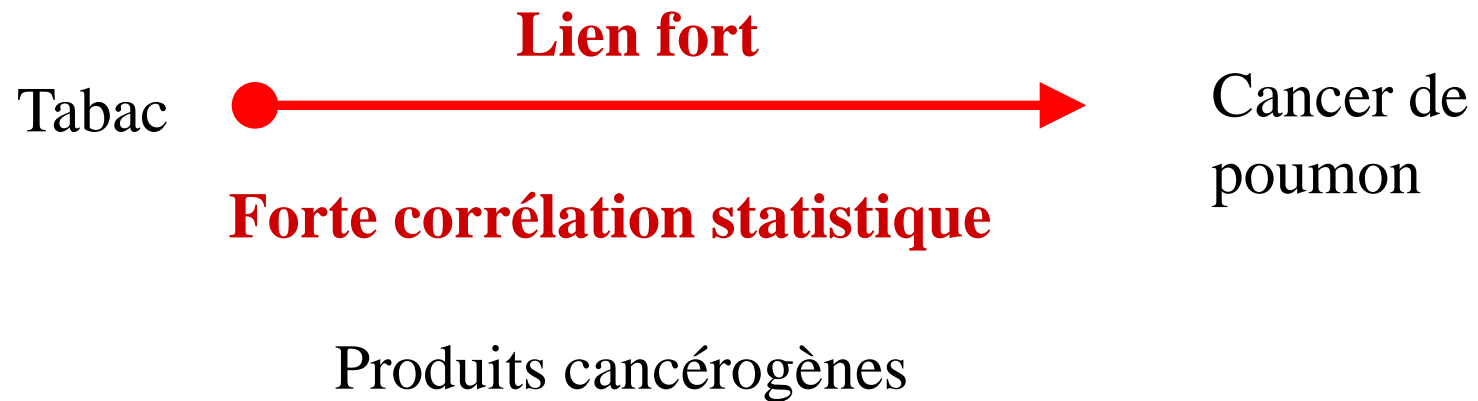
Changement de contexte \Rightarrow changement de modèle

I.3 Critères de causalité

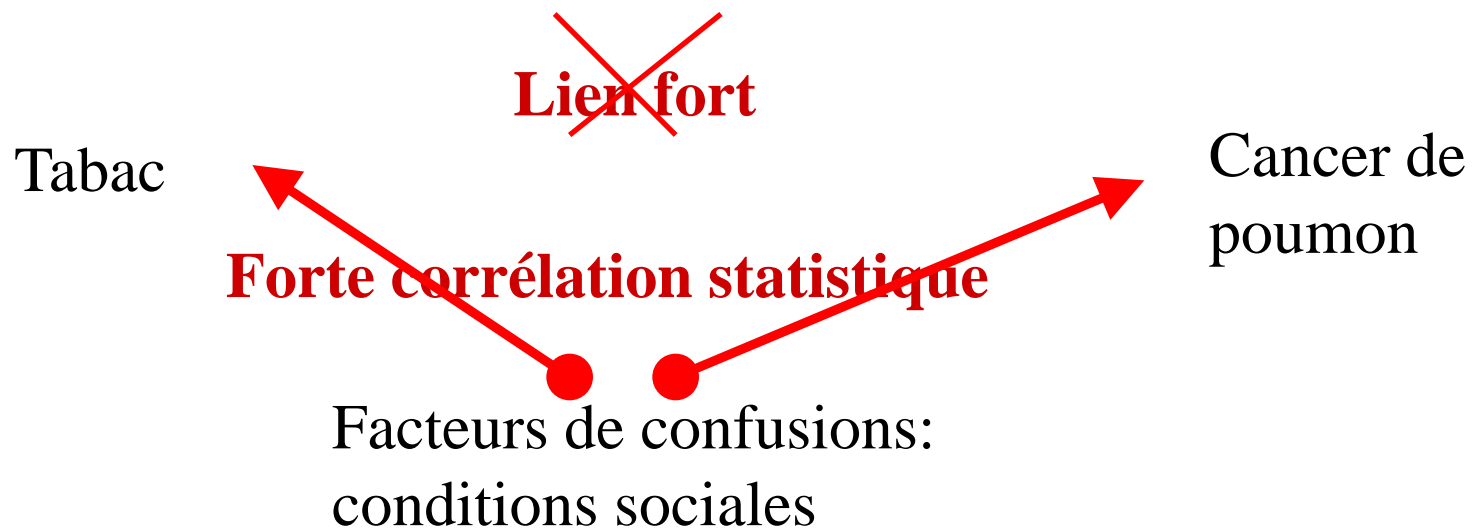
- Relation entre un facteur ou un signe et une maladie
- Le facteur est-il un cause de maladie?
- Le signe est-il une conséquence de la maladie?



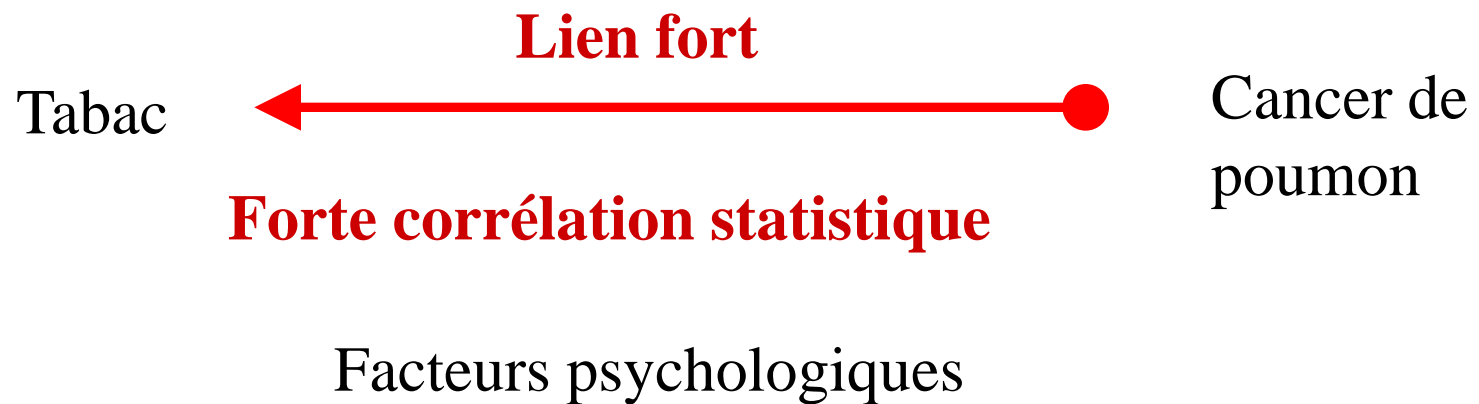
Interprétation 1



Interprétation 2



Interprétation 3



Austin Bradford Hill (1897-1991)



9 critères qui définissent la causalité (1965)

- Sans critères *sine qua non*
- Importance relative

1. Force de l'association

- Niveau de corrélation
- Importance de la différence

2. Cohérence

- Externe : Reproductibilité, constance des résultats avec d'autres travaux, équipes, populations
- Interne : protocole, prise en compte des biais

3. Temps

- Le cause précède l'effet

4. Relation dose-effet

5. Preuves expérimentales

6. Spécificité de l'association

7. Cohérence biologique

- En fonction de l'histoire naturelle de la maladie, physiopathologie

8. Plausibilité biologique

- En regard des connaissances disponibles

9. Analogie

- Avec d'autres évènements reliés à une autre pathologie

II. Modélisation Mathématique en Médecine

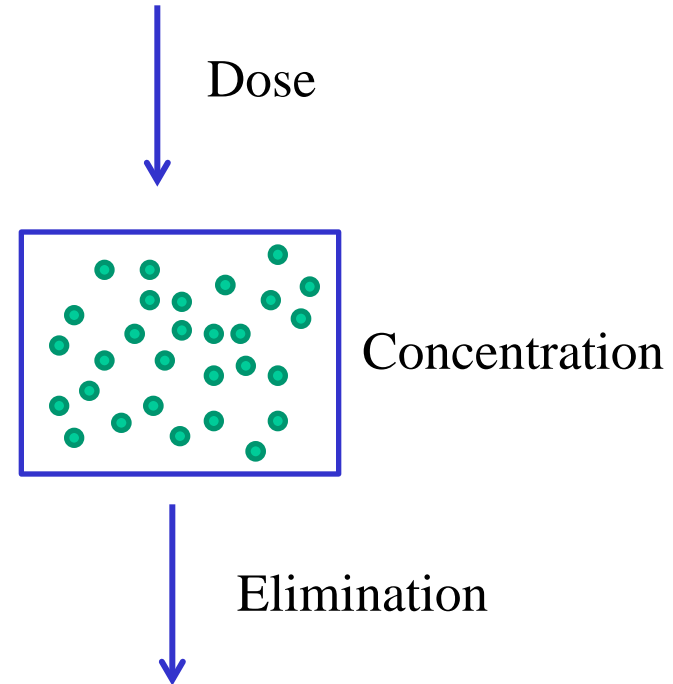
II.1 Modèles compartimentaux

Pourquoi ?

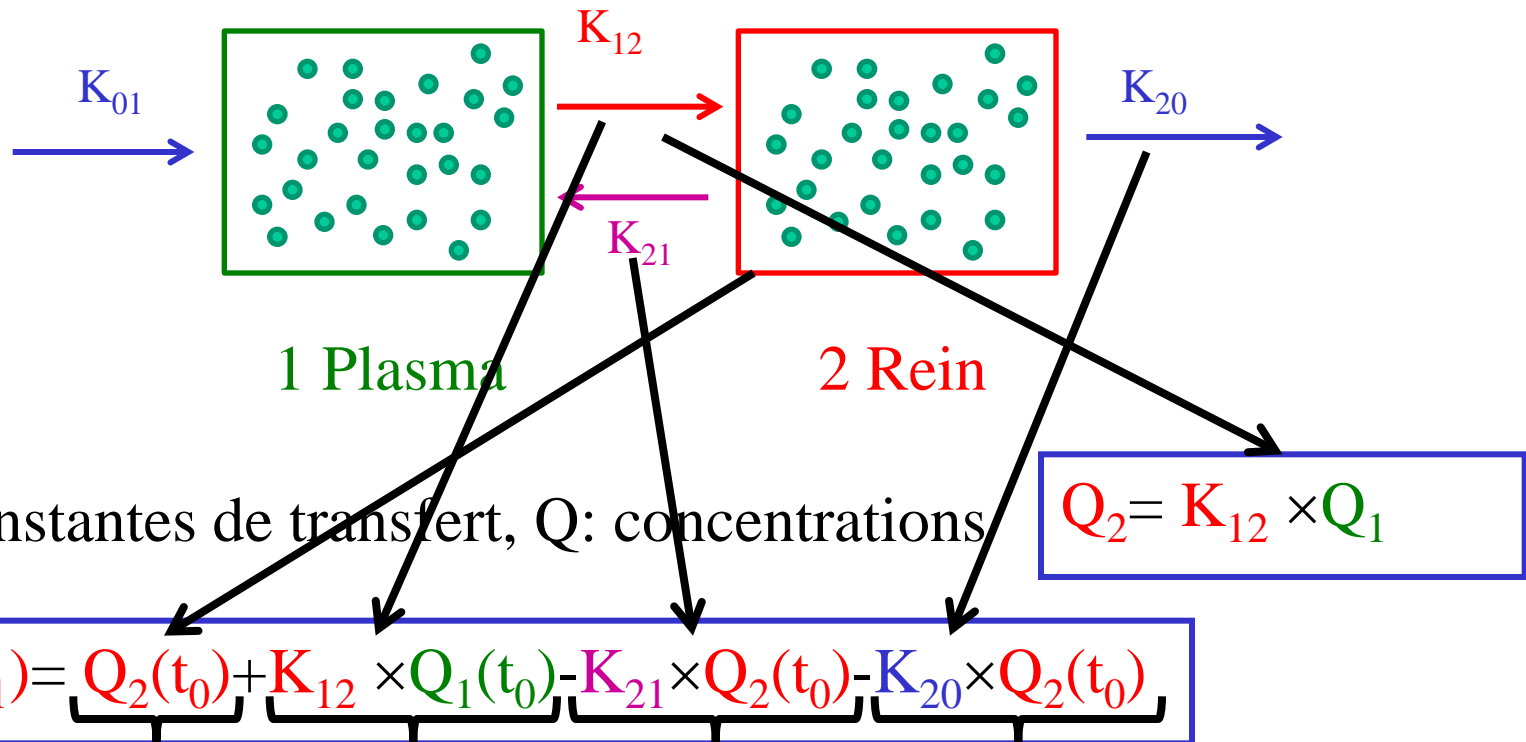
- Modèles simples
- Données plasmatiques (facile à obtenir)
- Définition et adaptation de la posologie

Définitions

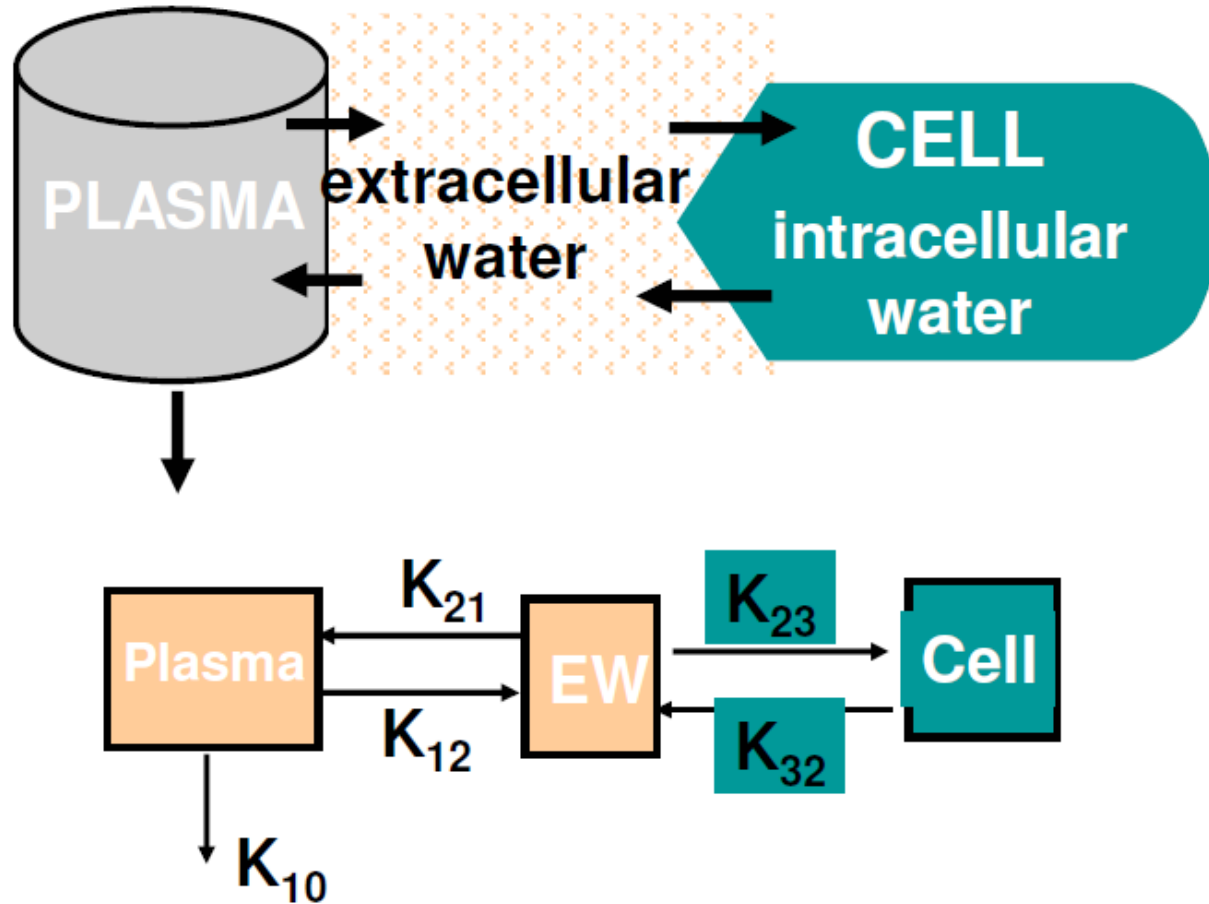
- **Compartiment:**
 - ensemble **homogène** de molécules
 - Unité fonctionnelle du modèle
 - Réalité biologique ou simplification
- **Mesures:**
 - Constantes absolues
 - Variables
 - Paramètres
 - Influences extérieures



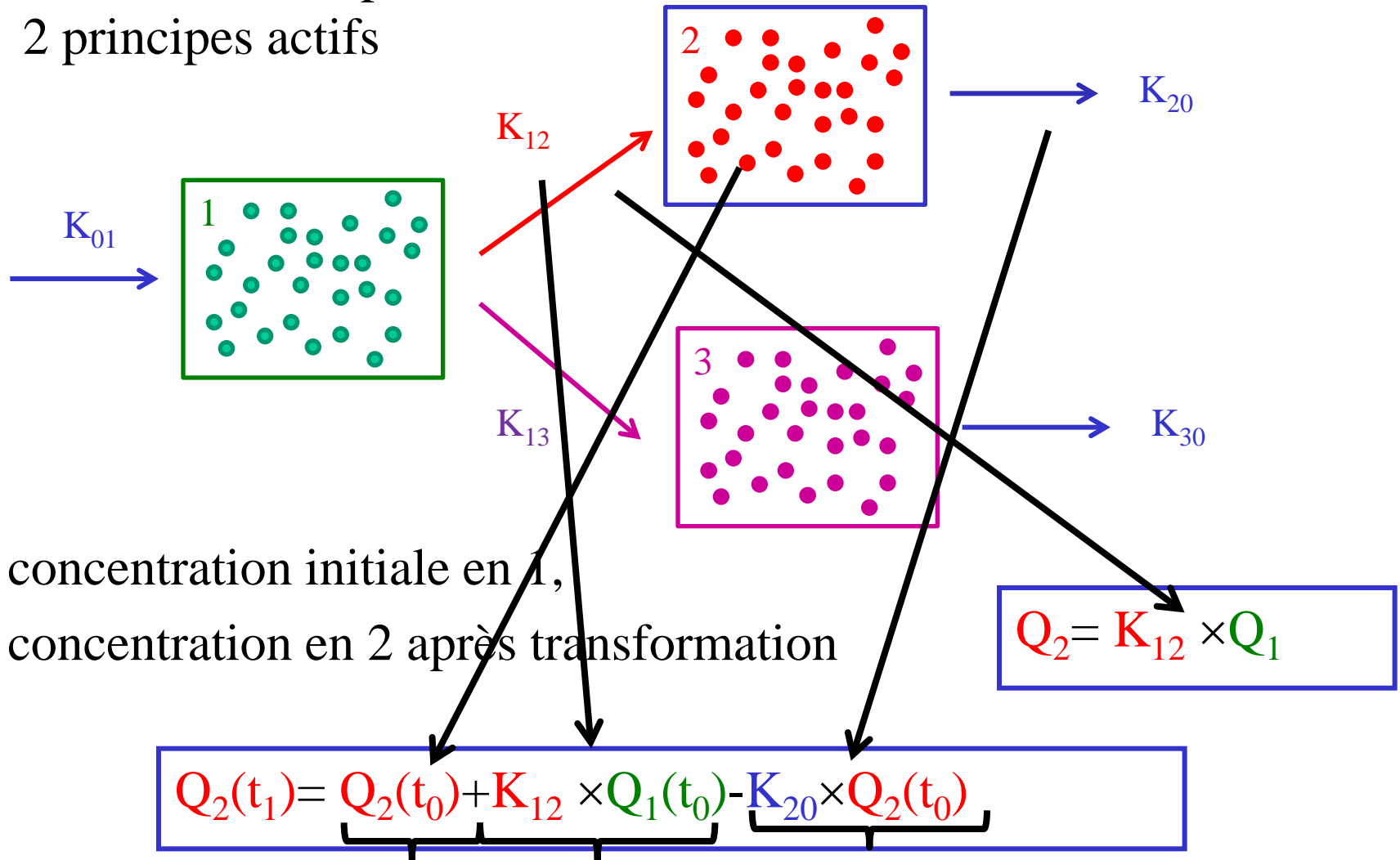
- Modèle à 2 compartiments: transfert entre 2 systèmes



- Modèle à 3 compartiments: transferts de médicaments



- Modèle à 3 compartiments: transformation d'un médicament en 2 principes actifs

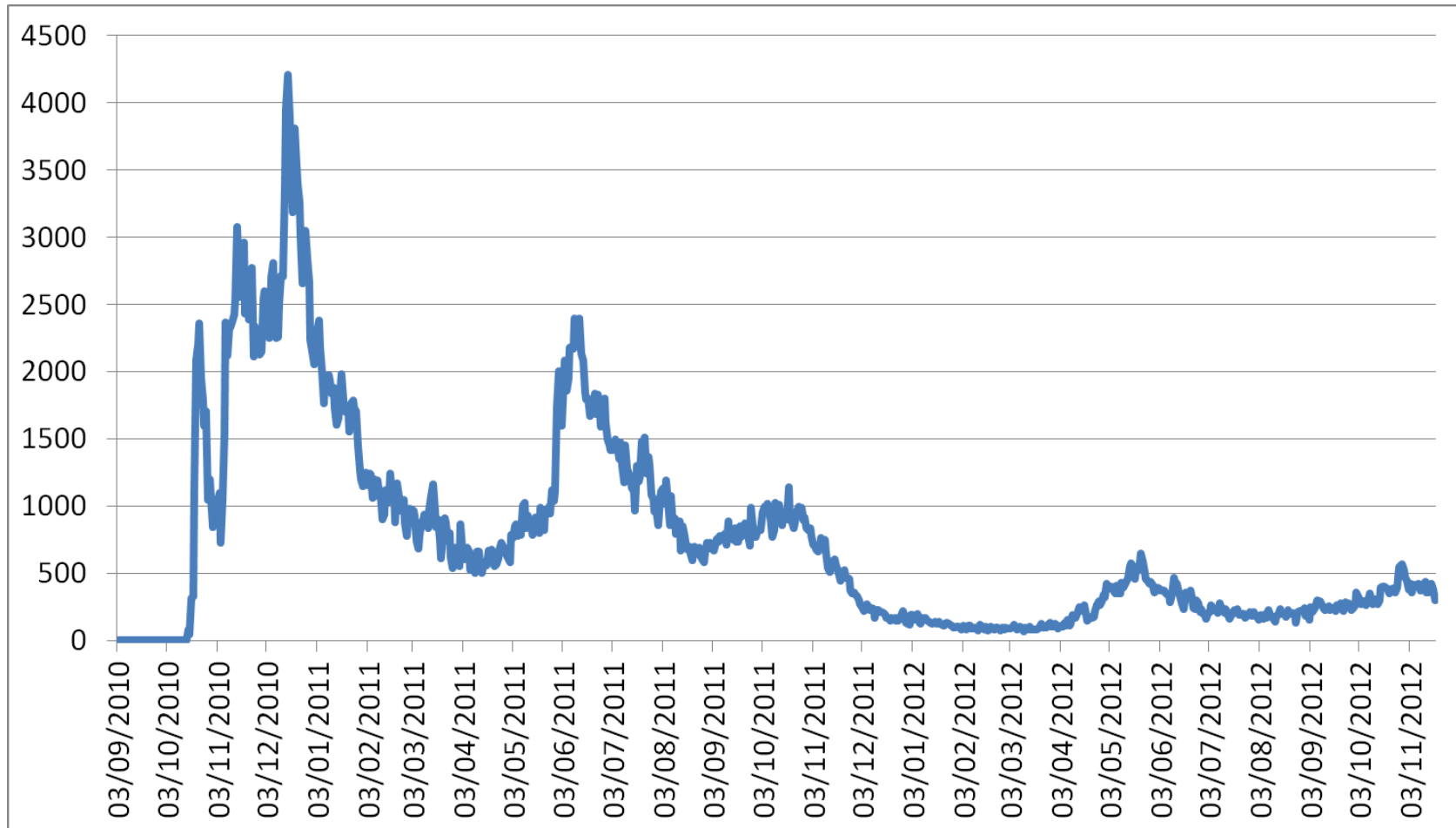


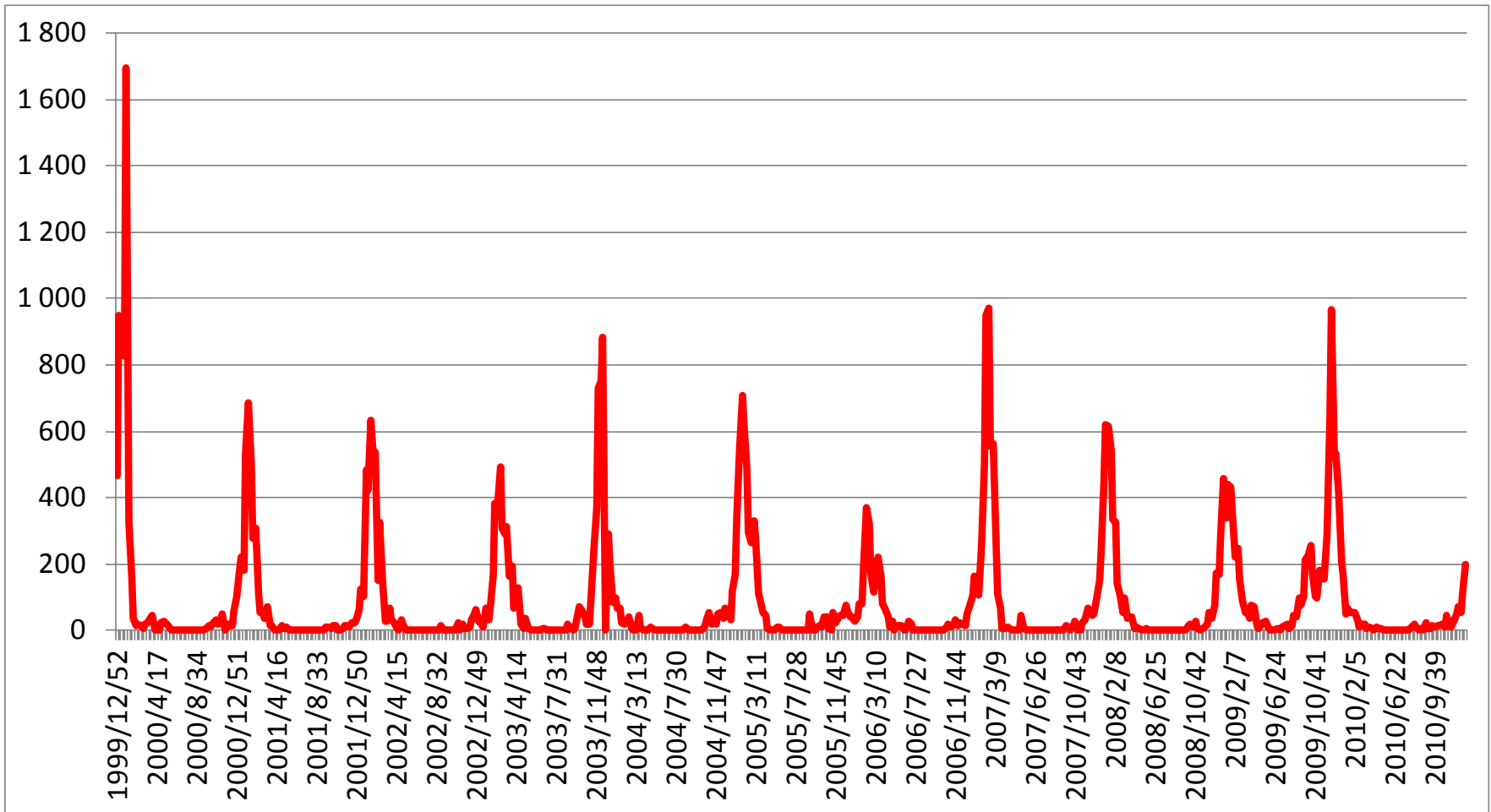
Q_1 concentration initiale en 1,
 Q_2 concentration en 2 après transformation

$$Q_2 = K_{12} \times Q_1$$

$$Q_2(t_1) = Q_2(t_0) + K_{12} \times Q_1(t_0) - K_{20} \times Q_2(t_0)$$

II.2 Epidémiologie Mathématique





Objectif de la modélisation en épidémiologie:

- Comment évolue une épidémie?

⇒ Compréhension de la dynamique des épidémies

- Comment prévoir l'évolution d'une épidémie débutante?

⇒ Utilisation d'une la mécanique comprise

- Peut-on protéger les populations susceptibles?

⇒ Qui? Quelle proportion? Comment?

- Observations:

- Population saine

Susceptibles

- Symptomatologie clinique

Infectés

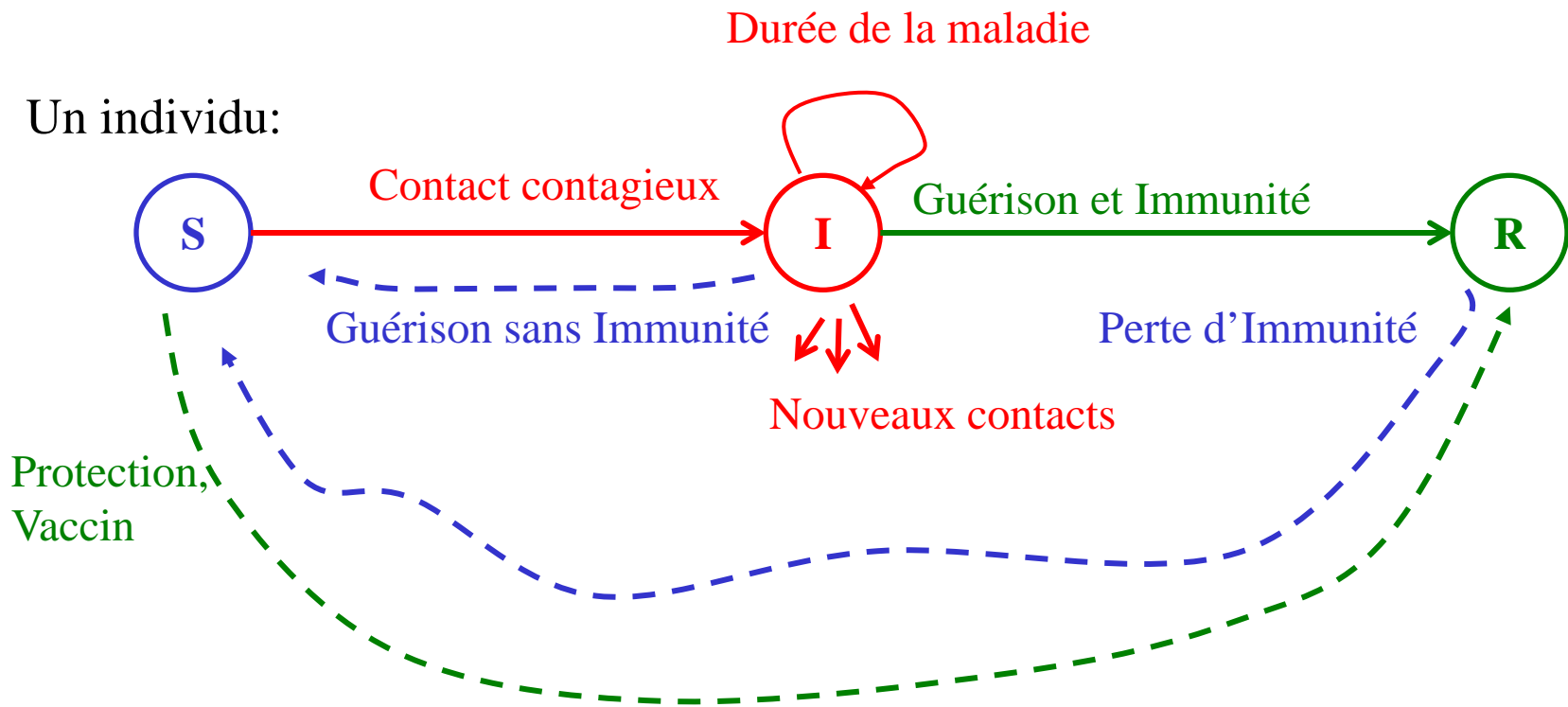
- Connaissance de la physio-pathologie

**Contagiosité,
immunité,
mortalité**

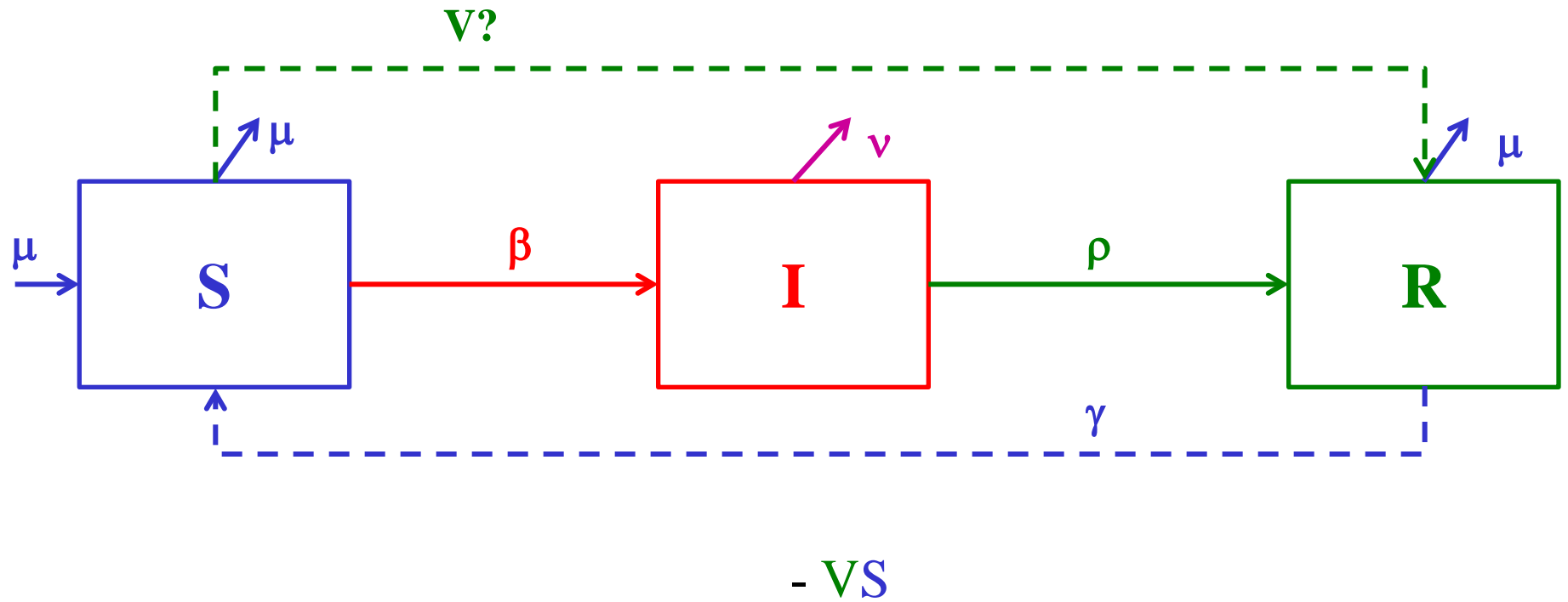
- Connaissance des mécanismes immunitaires

Résistants

Modèle compartimental

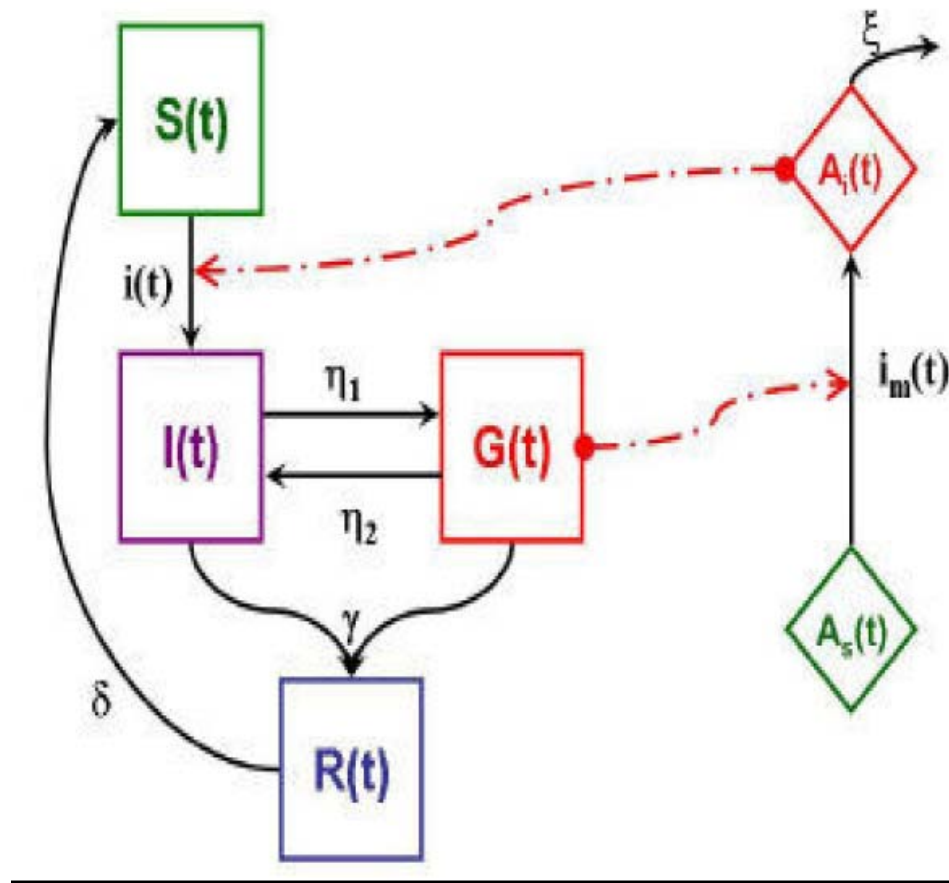


Population N:



+VS

- *Autre exemple: le paludisme*



Nombre de reproduction de base

- Paramètre clef en épidémiologie
- Noté R_0 ou Z_0
- Définition:

Nombre de cas secondaire à partir d'un cas, au début de l'épidémie

- Dépend des paramètres:
 - Contagiosité, Durée de la maladie, Mortalité
- Interprétation
 - Si $R_0 > 1$ l'épidémie continue
 - Si $R_0 < 1$ l'épidémie s'éteint

Exemples:

- Rougeole: 15 à 20
- Variole: 3
- Choléra Haiti: 2 à 3
- SARS: 2
- Grippe: 2 à 4
- Hépatite B: 1,1
 - Sous groupes à haut risque: 8,8

Stratégies de protection (ex. vaccin, eau potable...)

⇒ Quelle proportion de la population susceptible faut-il protéger pour éteindre l'épidémie?

$$p > 1 - 1/R_0$$

- Rougeole: 15 à 20 ⇒ 93 à 95%
- Grippe: 2 à 4 ⇒ 50 à 75%
- Variole: 3 ⇒ 67%
- Choléra: 2 à 3 ⇒ 50 à 67% (500 000 à 700 000 à Haiti)
- SARS: 2 ⇒ 50%
- Hépatite B: 1,1 ⇒ 10%
- Sous groupes haut risque: 8,8 ⇒ 89%

Attention à l'interprétation des modèles:

- ⇒ Contexte (ex. Choléra Haiti \neq Bangladesh)
- ⇒ Ecriture du modèle (manque de connaissance, facteurs non pris en compte)
- ⇒ Estimation des paramètres (disponibilité des données)

Exemple de prédictions erronées:

- Choléra en Haiti
- Grippe H1N1
- nvCJD

III. Modèles statistiques en médecine

III.1 Introduction

- Analyses statistiques simples: Analyse **Univariée**
⇒ relation entre **un** facteur de risque et la maladie

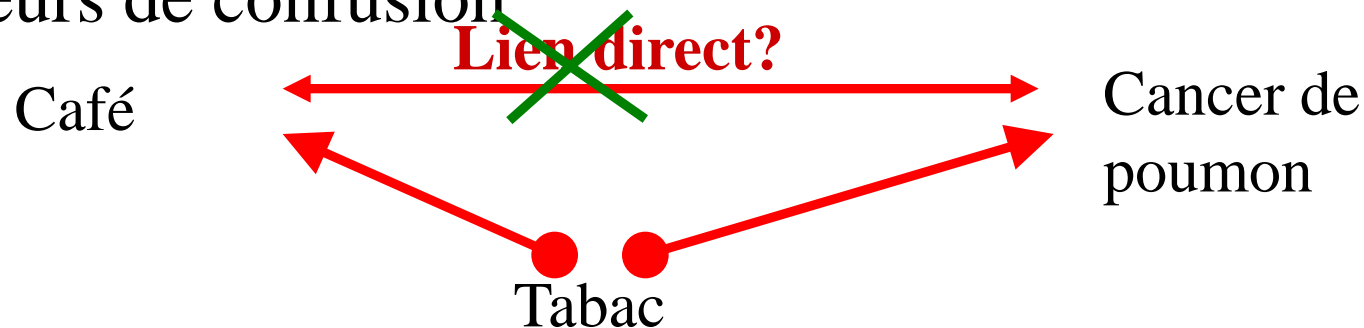
Ex. Quels sont les facteurs de risque d'Echinococcose alvéolaire en France?

- Avoir un chien
- Toucher un renard
- Consommer des baies sauvages
- Avoir un jardin potager
- Etre en zone d'endémie
- Etre chasseur
- Etre agriculteur

Analyse univarié donne la relation entre **un seul** facteur de risque et la maladie.

Problèmes:

- Facteurs de confusion



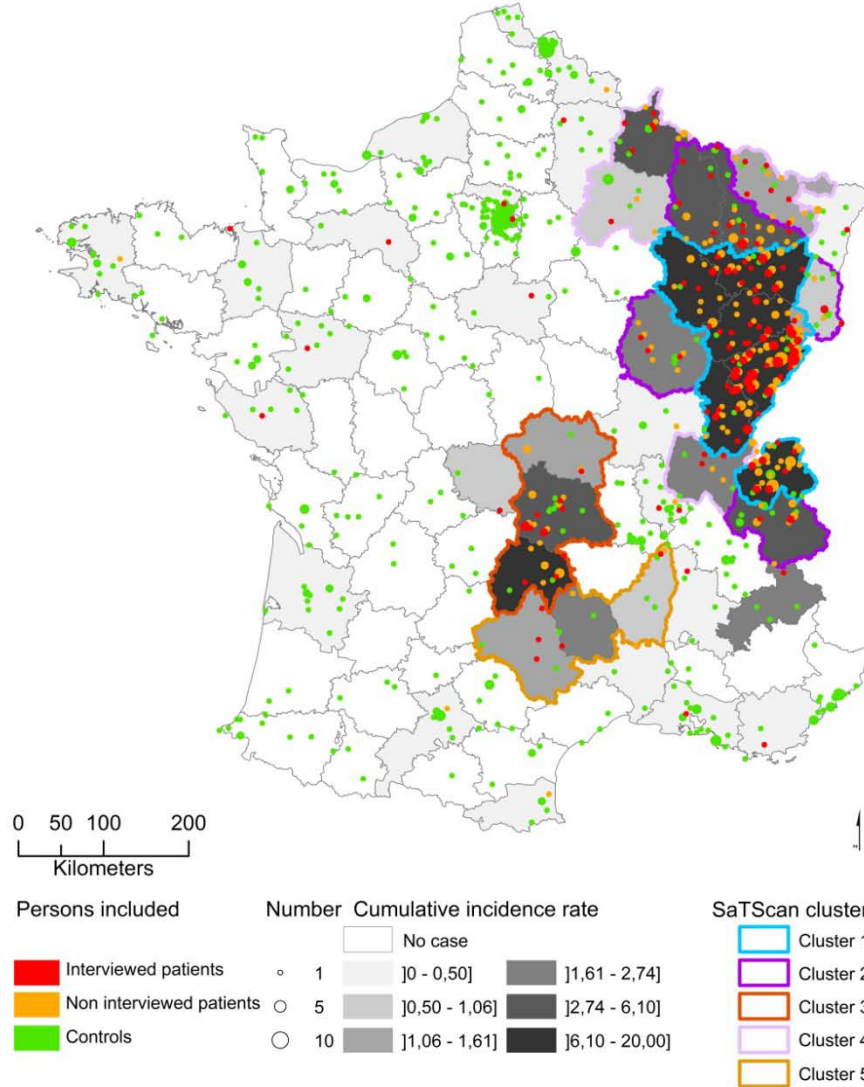
⇒ Ajustement sur les facteurs de confusion

- Comment déterminer le risque de maladie **pour plusieurs facteurs** à la fois ? ⇒ **Modèle statistique**

III.2 An

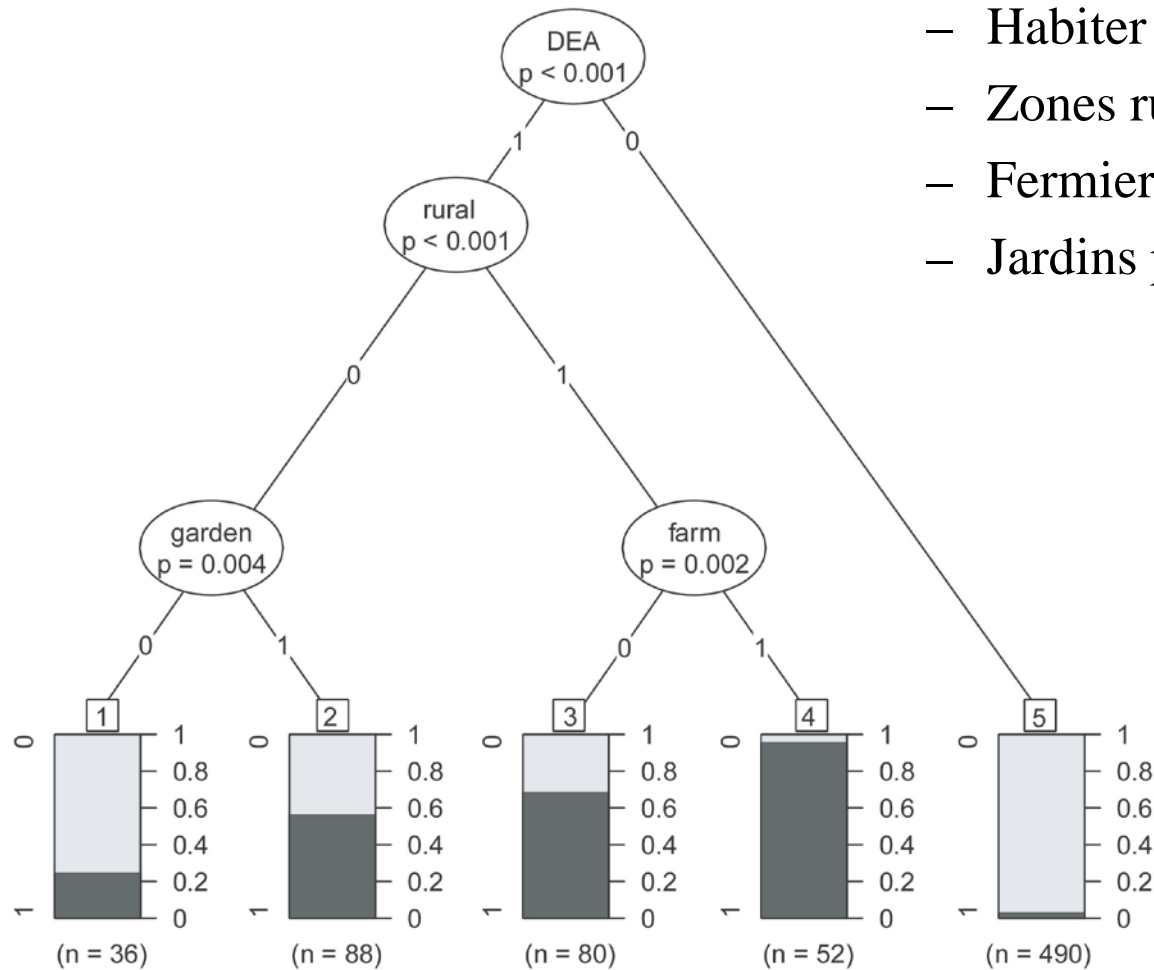
Exemple: cor

- Anciennes
 - Passage
 - Consom
 - Manipul



en France

- Modification du « modèle »:
 - Habiter en zone d'endémie,
 - Zones rurales
 - Fermiers
 - Jardins potagers en zone urbaines



Lecture critique d'article: la pratique courante

1. D'abord une analyse univariée
2. Puis une analyse multivariée

Exemple: survie des athlètes olympique (Zwiers, BMJ 2012)

Facteurs étudiés:

- Sexe,
 - Age,
 - Nationalité
 - Intensités de l'effort: Cardio-vasculaire, Statique, Dynamique
 - Contact physique, Collision
- } Facteurs d'ajustements

Table 1 | Hazard ratios of mortality for athletes in disciplines with different intensities of exercise

Intensity	Univariate analysis*		Multivariate analysis†	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
Cardiovascular:				
Low	Reference		Reference	
Moderate	1.01 (0.96 to 1.07)	0.71	1.04 (0.95 to 1.15)	0.40
High	0.98 (0.92 to 1.04)	0.46	1.05 (0.89 to 1.25)	0.58
Static:				
Low	Reference		Reference	
Moderate	0.94 (0.89 to 0.99)	0.02	0.93 (0.85 to 1.01)	0.09
High	0.99 (0.94 to 1.04)	0.62	0.95 (0.85 to 1.07)	0.40
Dynamic:				
Low	Reference		Reference	
Moderate	0.94 (0.89 to 0.99)	0.03	0.94 (0.87 to 1.01)	0.09
High	0.97 (0.92 to 1.02)	0.19	0.94 (0.83 to 1.06)	0.34

*Adjusted for sex, year of birth, and nationality.

†Additionally includes all types of exercise intensity (cardiovascular, static, and dynamic) in model.

Table 2| Hazard ratios of mortality for athletes in disciplines with different risk of bodily collision and physical contact

Sport type	Univariate analysis*		Multivariate analysis†	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
Bodily collision:				
No	Reference		Reference	
Yes	1.11 (1.06 to 1.15)	<0.001	1.03 (0.98 to 1.09)	0.25
Physical contact:				
Low	Reference		Reference	
Moderate	0.97 (0.93 to 1.02)	0.25	0.96 (0.92 to 1.01)	0.16
High	1.16 (1.11 to 1.22)	<0.001	1.13 (1.06 to 1.21)	<0.001

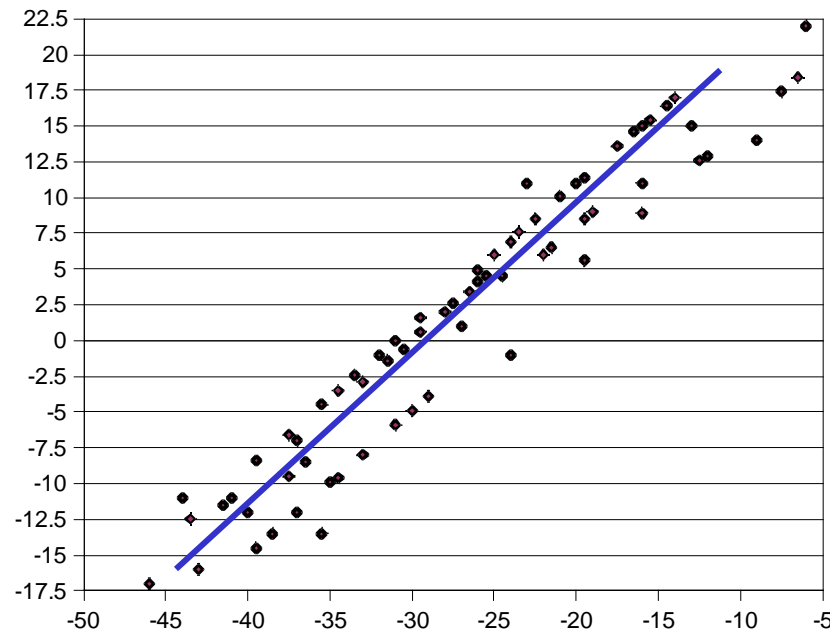
*Adjusted for sex, year of birth, and nationality.

†Additionally includes both bodily collision and physical contact in model.

- Entre Analyse univariée et analyse multivariée, des facteurs ont une relation à la maladie modifiée
- **Seule l'analyse multivariée est interprétable**
- **Les résultats dépendent du modèle utilisé:**
 - **Facteurs analysés**
 - **Population étudiée et contexte**
 - **Méthode choisie**

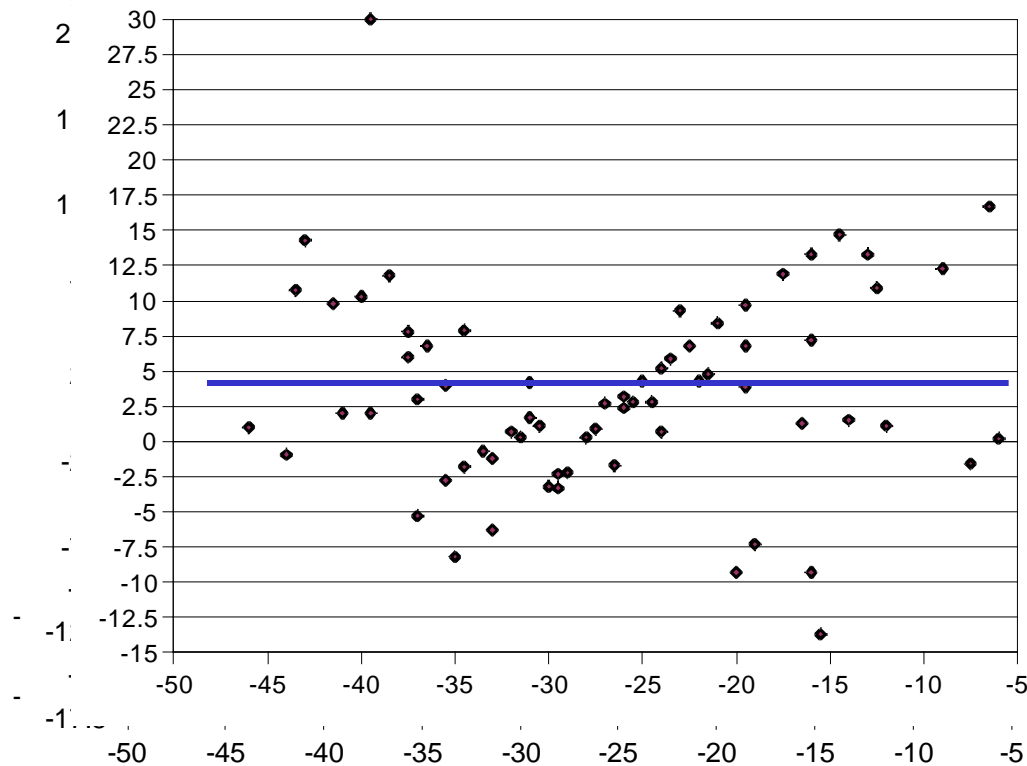
Principales méthodes dans les articles

- Méthodes autour du modèle linéaire
(régression linéaire, régression logistique, régression de Poisson, modèle de Cox...)
- Principe générale: la relation entre la maladie et les facteurs est linéaire (sous réserve de quelques modifications)



$$\text{Maladie} = a + b_1 \times \text{Facteur}_1 + b_2 \times \text{Facteur}_2 + b_3 \times \text{Facteur}_3$$

- La **force du lien** entre chaque facteur et la maladie est donné par son **paramètre b**
- Le **sens du lien** entre chaque facteur et la maladie est donné par son **paramètre b**:
 - Si $b=0$ pas de lien
 - Si $b<0$ facteur protecteur
 - Si $b>0$ facteur de risque



- A partir des **paramètres b** on peut estimer, pour chaque facteur étudier:
 - des **Odd Ratios** (Régression logistique)
 - **Rapport d'Incidence Standardisé** (Régression de Poisson)
 - des **Risques Relatifs** (Modèle de Cox)

Ajusté (en tenant compte) sur les autres facteurs

Survie des athlètes olympiques (BMJ 2012) : modèle de Cox

Table 2| Hazard ratios of mortality for athletes in disciplines with different risk of bodily collision and physical contact

Sport type	Univariate analysis*		Multivariate analysis†	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
Bodily collision:				
No	Reference		Reference	
Yes	1.11 (1.06 to 1.15)	<0.001	1.03 (0.98 to 1.09)	0.25
Physical contact:				
Low	Reference		Reference	
Moderate	0.97 (0.93 to 1.02)	0.25	0.96 (0.92 to 1.01)	0.16
High	1.16 (1.11 to 1.22)	<0.001	1.13 (1.06 to 1.21)	<0.001

*Adjusted for sex, year of birth, and nationality.

†Additionally includes both bodily collision and physical contact in model.

Cancer chez l'enfant (BMJ 2011) : régression logistique

Exposition in utero aux radiations et ultra-sons

Table 3 | Odds ratios (95% confidence intervals) for childhood cancer associated with any in utero exposure to radiation and ultrasound scans, UK Childhood Cancer Study, 1976-96*

Cancer	Radiation exposure			Ultrasound exposure		
	No of cases exposed	No of controls exposed	Odds ratio†	No of cases exposed	No of controls exposed	Odds ratio†
All cancers	120	185	1.14 (0.90 to 1.45)	2282	4234	0.93 (0.79 to 1.09)
Leukaemia:	48	62	1.36 (0.91 to 2.02)	1074	2034	0.87 (0.68 to 1.11)
Acute lymphoid	36	53	1.20 (0.76 to 1.88)	913	1726	0.83 (0.63 to 1.09)
Acute myeloid	11	8	2.44 (0.95 to 6.33)	144	273	1.28 (0.64 to 2.56)
Chronic myeloid	1	1	–	15	33	0.47 (0.06 to 3.56)
Lymphoma:	16	30	1.06 (0.55 to 2.06)	182	310	1.25 (0.80 to 1.95)
Non-Hodgkin's	13	18	1.48 (0.66 to 3.32)	131	224	1.30 (0.75 to 2.25)
Hodgkin's	3	12	–	46	80	1.03 (0.46 to 2.34)
Brain/central nervous system	25	41	1.06 (0.64 to 1.77)	408	731	1.08 (0.77 to 1.52)
Sarcoma	10	15	1.13 (0.49 to 2.61)	142	271	0.69 (0.39 to 1.22)
Peripheral neural tumours	7	13	1.00 (0.37 to 2.67)	155	288	0.88 (0.46 to 1.69)
Renal	5	6	1.64 (0.48 to 5.59)	130	252	0.56 (0.28 to 1.13)
Retinoblastoma	2	5	–	64	118	0.84 (0.19 to 3.71)
Gonadal	1	6	–	51	92	1.04 (0.34 to 3.18)
Hepatic	0	3	–	22	41	0.88 (0.11 to 6.80)

*Participants with trisomy 21 and one participant with radiotherapy excluded from analysis. Results not shown for categories with fewer than five exposed cases.

†Adjusted for sex of child, age at diagnosis, UKCCS study region, birth weight, maternal age, early infancy radiation (1 year lag), and early infancy ultrasound scans (1 year lag); radiation and ultrasound exposure each adjusted for other.

Cancer chez l'enfant (BMJ 2011): Exposition précoce

Table 5 | Odds ratios (95% confidence intervals) for childhood cancer associated with any exposure to diagnostic radiation or ultrasound scans in early infancy (0-100 days), UK Childhood Cancer Study, 1976-96*

Cancer	Early infancy radiation exposure				Early infancy ultrasound scan exposure			
	0 year lag		2 year lag		0 year lag		2 year lag	
	No of cases/ controls exposed	Odds ratio† (95% CI)	No of cases/ controls exposed	Odds ratio† (95% CI)	No of cases/ controls exposed	Odds ratio† (95% CI)	No of cases/ controls exposed	Odds ratio† (95% CI)
All cancers	68/102	1.16 (0.83 to 1.62)	50/75	1.19 (0.82 to 1.74)	49/58	1.92 (1.25 to 2.95)	25/33	1.55 (0.89 to 2.70)
Leukaemia:	34/49	1.39 (0.87 to 2.23)	27/40	1.35 (0.81 to 2.27)	16/35	0.95 (0.49 to 1.86)	9/25	0.68 (0.30 to 1.53)
Acute lymphoid	29/43	1.41 (0.85 to 2.35)	26/34	1.55 (0.90 to 2.67)	13/33	0.74 (0.36 to 1.52)	9/25	0.65 (0.29 to 1.45)
Acute myeloid	5/5	1.63 (0.42 to 6.35)	1/5	–	3/2	–	0/0	–
Chronic myeloid	0/0	–	0/0	–	0/0	–	0/0	–
Lymphoma	7/3	5.14 (1.27 to 20.78)	7/3	5.14 (1.27 to 20.8)	1/0	–	1/0	–
Non-Hodgkin's	6/2	6.85 (1.31 to 35.69)	6/2	6.85 (1.31 to 35.7)	1/0	–	1/0	–
Hodgkin's	1/0	–	1/0	–	0/0	–	0/0	–
Brain/central nervous system	9/16	0.90 (0.36 to 2.31)	6/10	0.94 (0.31 to 2.92)	6/7	1.90 (0.59 to 6.17)	4/2	–
Peripheral neural tumours	5/12	0.26 (0.05 to 1.33)	2/8	–	8/5	8.79 (1.64 to 47.15)	1/2	–
Sarcoma	4/8	–	3/5	–	4/3	–	3/3	–
Renal	4/6	–	3/5	–	6/4	3.77 (0.87 to 16.36)	5/0	–
Hepatic	2/1	–	1/0	–	4/0	–	2/0	–
Gonadal	1/2	–	0/1	–	2/2	–	0/1	–
Retinoblastoma	1/2	–	0/0	–	2/2	–	0/0	–

*Participants with trisomy 21 and one participant with radiotherapy excluded from analysis. Results not shown for categories with fewer than five exposed cases.

†Adjusted for sex of child, age at diagnosis, UKCCS study region, maternal age, birth weight, and in utero exposure to radiation and ultrasound scans; early infancy diagnostic radiation and ultrasound scans each adjusted for other.

Facteurs de risque de Choléra, fin 2010 - début 2011, Haïti (EID2011): régression de Poisson

Facteurs	Rapport d'incidence standardisé [CI95%]	P-value
Etre en aval de la rivière Meille	4.91 [1.47-16.47]	0.012
Distance à Mirebalais (km)	0.99 [0.94-1.04]	0.594
Présence d'un camp de réfugiés	0.10 [0.01-1.12]	0.063
Commune en plaine costale	4.60 [2.28-9.30]	0.0001

Attention à l'interprétation des modèles:

- ⇒ Contexte (ex. Population \neq Sous-groupe, biais de sélection)
- ⇒ Choix de la méthode
- ⇒ Ecriture du modèle (ex. manque de connaissance, facteurs non pris en compte)

■ Références

- Jean Bouyer: *Epidémiologie quantitative*, Ed. INSERM
- Coll. (CIMES): *Biostatistiques*, Ed. Omnisciences
- AJ Valleron: *L'épidémiologie en France*, Ed. EDP Sciences
- A Flahault, P Zylberman: *Des épidémies et des hommes*, Ed. De la Martinière
- A Flahault, JY Nau: *A(H1N1) journal de la pandémie*, Ed. Plon
- A Flahault: <http://blog.ehesp.fr/blog/tag/epidemiologie/>