

# LES MODÈLES ÉPIDÉMIOLOGIQUES COMPARTIMENTAUX

Bruno Anselme

22 mars 2020

*En cette période de confinement, je vous propose quelques lignes sur les principes de la modélisation compartimentale en épidémiologie. Nous y verrons quelques concepts simples assez éclairants et en particulier le concept de  $\mathcal{R}_0$  essentiel en épidémiologie. Historiquement, d'ailleurs, le succès du modèle de Ross pour le paludisme est directement lié à l'introduction de ce paramètre.*

*Ceux qui ont vu le film Contagion ont peut-être remarqué qu'il y est fait mention de ce paramètre. Comme d'habitude, j'ai essayé de me placer à mi-chemin des mathématiques et de la biologie, d'écrire des mathématiques accessibles au biologiste et de la biologie accessible aux mathématiciens. En la matière, il y a peu de biologie et pas mal de mathématiques ; l'effort sera donc peut-être plus intense pour les biologistes.*

## 1 Le paludisme, le "Mosquito theorem" et la lutte contre les Anophèles

Le paludisme est une maladie potentiellement mortelle. Il est dû à des parasites transmis à l'homme par des piqûres de moustiques infectés.

### 1.1 Le *Plasmodium*, agent du paludisme

Le paludisme est dû à des parasites du genre *Plasmodium* transmis d'une personne à l'autre par des piqûres de moustiques *Anopheles* infectés, appelés *vecteurs du paludisme*, qui piquent principalement entre le crépuscule et le petit matin.

Il existe quatre types de parasites et donc quatre types de paludisme humain :

*Plasmodium falciparum* ;

*Plasmodium vivax* ;

*Plasmodium malariae* ;

*Plasmodium ovale*.

Les *Plasmodium falciparum* et *Plasmodium vivax* sont les plus répandus. Le *Plasmodium falciparum* est le plus mortel.

Lorsqu'un humain est infecté, le parasite se développe et prolifère dans les cellules du foie puis dans les globules rouges du sang. Là, sa prolifération fait des dégâts et il provoque les fièvres typiques de la maladie. Lorsqu'un Anophèle est infecté, le parasite se développe dans les voies digestives et entre autres dans les glandes salivaires. Lorsqu'un Anophèle infecté pique un humain, les parasites contenus dans les glandes salivaires contaminent l'homme. Lorsqu'un Anophèle pique un humain infecté, il absorbe son sang et s'en trouve contaminé.

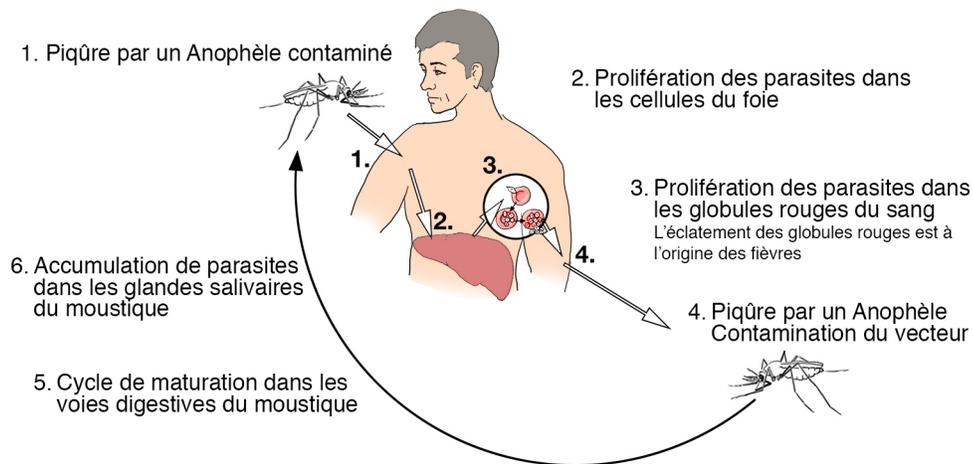


FIGURE 1 – Le cycle du parasite *Plasmodium* entre le moustique et l'homme

Au début du vingtième siècle, il était généralement admis qu'il était inutile de lutter contre les moustiques pour lutter contre cette maladie. En effet : sachant qu'il n'est pas possible d'éliminer exactement tous les moustiques, l'épidémie peut toujours repartir. Imaginons un seul humain infecté, il finira bien par infecter un moustique qui infectera d'autres humains, et ainsi de suite.

Sir Ronald Ross (prix Nobel de Physiologie Médecine en 1902) n'était pas du tout de cet avis. Et il s'est servi de mathématiques pour convaincre la population médicale.

## 1.2 Les flux entre différentes populations : un modèle compartimental

Ronald Ross a proposé en 1911 un modèle qui prend en compte les populations d'Anophèles et d'hommes. Chacune de ces deux populations est divisée en deux classes : les susceptibles  $S$  et les infectés  $I$ .

On notera  $S_h(t)$  le nombre de susceptibles humains (indice  $h$ ) à la date  $t$ , et  $I_h(t)$  le nombre d'infectés humains à la date  $t$ . Le nombre d'Anophèles (indice  $a$ ) susceptibles est noté  $S_a(t)$  et le nombre d'Anophèles infectés est noté  $I_a(t)$ .

Ross ne prend pas en compte de période intermédiaire, de latence : un infecté est instantanément infectieux.

Ross considère donc qu'il y a quatre compartiments, et que les effectifs de ces compartiments,  $I_h$ ,  $S_h$ ,  $I_a$  et  $S_a$  sont régis par des flux entre ces compartiments.

Les populations d'hommes et d'anophèles sont bien entendu obtenues en sommant les susceptibles et les infectés :

$$\begin{aligned} H(t) &= S_h(t) + I_h(t) \\ A(t) &= S_a(t) + I_a(t) \end{aligned}$$

Le modèle considère alors les paramètres suivants :

- $a$  : fréquence à laquelle un moustique pique ;
- $b_1$  : probabilité de transmission moustique-homme lors d'une piqûre ;
- $b_2$  : probabilité de transmission homme-moustique lors d'une piqûre ;
- $g_h$  : taux de guérison des humains ;
- $g_a$  : taux de guérison des anophèles ;
- $\mu_h$  : mortalité (naturelle) des humains ;
- $\mu_a$  : mortalité (naturelle) des anophèles.

### 1.2.1 Le taux de variation d'humains infectés

Nous nous intéressons aux variations de  $I_h(t)$  (les humains infectés) ; nous devons considérer le nombre de nouveaux infectés par unité de temps ainsi que le nombre de guéris.

Par unité de temps, chaque moustique réalise  $a$  piqûres. Il y a donc  $a \cdot I_a$  piqûres potentiellement infectantes, dont une proportion  $S_h/H = (H - I_h)/H$  concerne des humains susceptibles et dont une proportion  $b_1$  sont infectantes. Il y a donc par unité de temps  $a \cdot b_1 \cdot I_a \cdot \frac{H - I_h}{H}$  nouveaux humains infectés.

Et il y a par ailleurs  $g_h \cdot I_h$  guérisons et  $\mu_h \cdot I_h$  décès par unité de temps.

Ainsi, le taux de variation de  $I_h$  est donné par :

$$\frac{dI_h}{dt} = a \cdot b_1 \cdot I_a \cdot \frac{H - I_h}{H} - g_h \cdot I_h - \mu_h \cdot I_h$$

Ross décide de négliger  $\mu_h$  : ce terme est en effet très faible devant  $g_h$ . Les durées de vie caractéristiques (en dizaines d'années) et les temps de guérison (en dizaines de jours) sont sans comparaison, c'est-à-dire  $1/\mu_h \gg 1/g_h$ , soit  $\mu_h \ll g_h$ . Ceci permet d'assimiler  $H$  à une constante.

$$\begin{aligned} H &= S_h(t) + I_h(t) = \text{constante} \\ \frac{dI_h}{dt} &= a \cdot b_1 \cdot I_a \cdot \frac{H - I_h}{H} - g_h \cdot I_h \end{aligned} \tag{1}$$

### 1.2.2 Le taux de variation d'anophèles infectés

De la même manière, on peut construire l'équation correspondante pour  $I_a$  (les Anophèles infectés) :

$$\frac{dI_a}{dt} = a \cdot b_2 \cdot (A - I_a) \cdot \frac{I_h}{H} - g_a \cdot I_a - \mu_a \cdot I_a$$

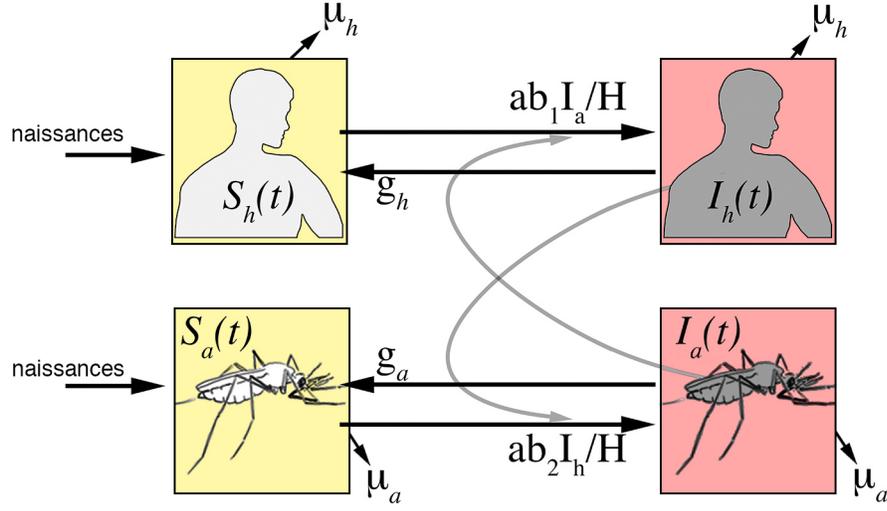


FIGURE 2 – Modèle compartimental pour le paludisme

qui exprime que par unité de temps il y a  $a \cdot S_a = a \cdot (A - I_a)$  piqûres par des anophèles susceptibles, piqûres dont une proportion  $I_h/H$  sont faites sur des hommes contaminés et dont une proportion  $b_2$  sont infectantes pour les anophèles.

Ross considère cette fois que l'on peut négliger  $g_a$  devant  $\mu_a$  : cela revient à considérer que, sur ses quelques jours de durée de vie, un anophèle femelle infecté reste infectieux jusqu'à sa mort. Par ailleurs, Ross considère que le nombre de moustiques est constant.

$$\frac{dI_a}{dt} = a \cdot b_2 \cdot (A - I_a) \cdot \frac{I_h}{H} - \mu_a \cdot I_a \quad (2)$$

### 1.2.3 Un système d'équations différentielles régissant la dynamique des quatre compartiments

Le modèle final s'écrit donc :

$$\begin{aligned} \frac{dI_h}{dt} &= a \cdot b_1 \cdot I_a \cdot \frac{H - I_h}{H} - g_h \cdot I_h \\ \frac{dI_a}{dt} &= a \cdot b_2 \cdot I_h \cdot \frac{A - I_a}{H} - \mu_a \cdot I_a \end{aligned} \quad (3)$$

Il s'agit d'un système de deux équations différentielles que nous ne chercherons pas à résoudre. En revanche, l'analyse de ce système est riche d'informations.

### 1.3 Les points d'équilibre du modèle

La première chose à faire est la recherche de points d'équilibre. Un point d'équilibre définit une situation où les effectifs des compartiments sont fixes, c'est-à-dire où :  $I_h = Cste$  et  $I_a = Cste$ .

Cela définit un point fixe sur un graphe montrant les trajectoires du système dans un repère  $I_h$  vs  $I_a$ .

Dire qu'un effectif est constant, c'est, mathématiquement, dire que sa dérivée par rapport au temps est nulle :  $dI_h/dt = 0$  et  $dI_a/dt = 0$ , équations que nous allons donc devoir résoudre.

#### 1.3.1 Détermination

Un premier point d'équilibre évident est donné par  $I_a = I_h = 0$ . Ce point (0,0) ne pose pas de problème d'interprétation biologique. Si aucun homme et aucun moustique n'est infecté, c'est qu'il n'y a pas de paludisme et il ne risque pas d'apparaître la moindre infection : il s'agit bien pour le moins d'un équilibre et peut-être d'un équilibre stable (ce qui est loin d'être vrai). Nous nommons cet équilibre l'Équilibre Sans Maladie (ESM).

Un second point d'équilibre ( $I_h^*, I_a^*$ ) peut être identifié facilement en posant :

$$\begin{aligned} 0 &= a.b_1.I_a^* \cdot \frac{H - I_h^*}{H} - g_h \cdot I_h^* \\ 0 &= a.b_2.I_h^* \cdot \frac{A - I_a^*}{H} - \mu_a \cdot I_a^* \end{aligned} \quad (4)$$

et en divisant ce système par le produit  $I_h^*.I_a^*$  (on pourra vérifier que  $I_a = 0 \Leftrightarrow I_h = 0$  et donc que l'on peut diviser par  $I_h I_a$  si l'on élimine l'hypothèse de l'équilibre ESM), ce qui donne :

$$\begin{aligned} 0 &= \frac{a.b_1}{I_h^*} - \frac{a.b_1}{H} - \frac{g_h}{I_a^*} \\ 0 &= \frac{a.b_2.A}{I_a^*.H} - \frac{a.b_2}{H} - \frac{\mu_a}{I_h^*} \end{aligned}$$

ou encore

$$\begin{cases} \frac{a.b_1}{I_h^*} - \frac{g_h}{I_a^*} = \frac{a.b_1}{H} \\ \frac{a.b_2.A}{I_a^*.H} - \frac{\mu_a}{I_h^*} = \frac{a.b_2}{H} \end{cases} \quad (5)$$

qui est un système simple de deux équations à deux inconnues ( $1/I_h^*$  et  $1/I_a^*$ ) dont la résolution (que je vous épargne) aboutit à :

$$\begin{aligned} I_h^* &= H \cdot \frac{1 - g_h \mu_a H / a^2 b_1 b_2 A}{1 + g_h H / a b_2 A} \\ I_a^* &= A \cdot \frac{1 - g_h \mu_a H / a^2 b_1 b_2 A}{1 + \mu_a / a b_1} \end{aligned} \quad (6)$$

Ces deux expressions ont le même numérateur (au facteur  $A$  ou  $H$  près) et des dénominateurs positifs.

Mais, bien entendu,  $I_h^*$  et  $I_a^*$  ne peuvent être que positifs : ce sont des effectifs d'hommes et de moustiques ! Il ne peut donc y avoir un tel équilibre différent de  $(0,0)$  que si les numérateurs sont positifs, soit :

$$A > \frac{g_h \mu_a H}{a^2 b_1 b_2} \quad \text{ou encore} \quad \frac{A}{H} > \frac{g_h \mu_a}{a^2 b_1 b_2} \quad (7)$$

Ce que nous pouvons écrire de la façon suivante :

$$\mathcal{R}_0 > 1 \quad \text{avec} \quad \mathcal{R}_0 = \frac{a^2 b_1 b_2 A}{g_h \mu_a H} \quad (8)$$

Ce "concept" de  $\mathcal{R}_0$  est tout à fait fondamental en épidémiologie, comme nous allons le voir.

### 1.3.2 Stabilité des points d'équilibre

Un point d'équilibre est dit stable si, partant de ce point, le système y revient après une petite perturbation. Il s'agit là de l'image classique d'une bille au repos au fond d'un bol qui définit un équilibre stable et d'une bille en équilibre sur la pointe d'un crayon qui définit un équilibre instable.

Mathématiquement, on étudie la stabilité du point d'équilibre d'un système d'équations différentielles en analysant la matrice jacobienne (ou matrice de Jacobi) de ce système en ce point.

Je ne détaille pas cette étude, calculatoire. Retenons que le point d'équilibre ESM est stable à la condition que :

$$\frac{A}{H} < \frac{g_h \mu_a}{a^2 b_1 b_2} \quad \text{c'est-à-dire} \quad \mathcal{R}_0 < 1$$

ce qui est la condition de la non-existence d'un point d'équilibre  $(I_h^*, I_a^*)$ .

Enfin, pour le point d'équilibre  $(I_h^*, I_a^*)$ , lorsqu'il existe (i.e. avec deux valeurs positives), l'étude montre que c'est toujours un équilibre asymptotiquement stable. Nous nommons cet équilibre l'Équilibre Endémique (EE). En effet, si cet équilibre est stable, les populations qui l'atteignent y restent, c'est-à-dire avec des effectifs non nuls et permanents d'humains et d'anophèles infectés, ce qui est la définition d'une endémie.

Par conséquent, l'équilibre ESM est stable si et seulement si  $\mathcal{R}_0 < 1$ , auquel cas EE n'existe pas. Et inversement, lorsque EE existe (donc lorsque  $\mathcal{R}_0 > 1$ ) alors il est stable et ESM est instable.

Ainsi, il ne peut y avoir qu'un seul point d'équilibre stable dans les conditions biologiques ( $0 \leq I_h \leq H$  et  $0 \leq I_a \leq A$ ), et ce point stable est :

$$\begin{aligned} ESM &= (0, 0) && \text{lorsque} \quad \mathcal{R}_0 < 1 \\ EE &= (I_h^*, I_a^*) && \text{lorsque} \quad \mathcal{R}_0 > 1 \end{aligned} \quad (9)$$

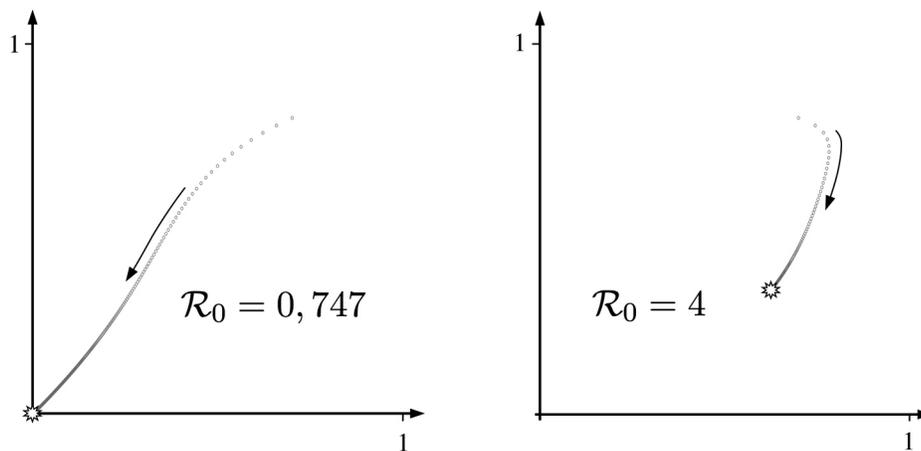


FIGURE 3 – Exemples de trajectoires obtenues numériquement pour deux valeurs différentes de  $\mathcal{R}_0$ .

### 1.3.3 Le *mosquito theorem* de Ross

Ce *théorème du moustique* énoncé par Ross en 1911 exprime :

1. Si  $\mathcal{R}_0 < 1$ , alors il existe un seul équilibre pour le modèle, l'ESM, et toutes les trajectoires tendent vers cet équilibre, c'est-à-dire vers une disparition de la maladie.
2. Si  $\mathcal{R}_0 > 1$ , alors il existe un seul équilibre, l'EE  $(I_h^*, I_a^*)$  qui est globalement asymptotiquement stable ; les trajectoires tendent vers cet équilibre, c'est-à-dire vers une installation durable de la maladie.

Et c'est de ce théorème que découle la démonstration de l'intérêt qu'il y a à lutter contre les moustiques.

## 1.4 Lutte contre le paludisme

Il ressort de ce qui précède qu'il est possible de lutter contre le paludisme en luttant contre les anophèles.

En effet, si  $\mathcal{R}_0 < 1$ , il n'y a pas d'épidémie et la maladie est jugulée, elle disparaît. Or cette condition  $\mathcal{R}_0 < 1$  peut aussi s'écrire :

$$A < A_C \quad \text{avec} \quad A_C = \frac{g_h \mu_a H}{a^2 b_1 b_2} \quad (10)$$

Nous appelons  $A_C$  l'effectif critique d'anophèles.

Il n'est ainsi pas nécessaire d'éliminer complètement les anophèles. Il suffit d'amener leur effectif en-dessous de l'effectif critique  $A_C$ . Dans ces conditions, les Anophèles sont si peu nombreuses que la probabilité de se faire infecter par quelques rares humains infectés est quasi nulle.

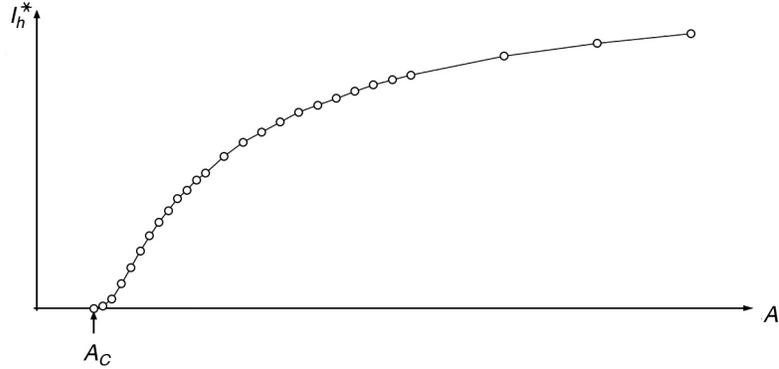


FIGURE 4 – **Evolution de l'endémie et effectif critique.** Le graphe représente l'effectif de la population humaine infectée ( $I_h^*$ ) en fonction de l'effectif total d'anophèles ( $A$ ). Lorsque  $A = A_C$  et donc  $I_h = 0$  : il n'y a plus d'endémie.

Ross envisage alors les valeurs plausibles des différents paramètres afin d'évaluer  $A_C$ . La précision est assez mauvaise, mais là n'est pas l'important. Ross a démontré que la lutte contre le paludisme pouvait (et devait) aussi se faire en luttant contre les moustiques, contrairement aux idées admises à l'époque.

Ross conclut son travail par : *«En fait, toute l'épidémiologie, dont l'objet est l'étude de la variation des maladies dans le temps et dans l'espace, doit être abordée mathématiquement, quel que soit le nombre de variables impliquées, si l'on veut l'aborder vraiment. Dire qu'une maladie dépend de certains facteurs n'est pas dire grand chose, à moins d'être en mesure d'évaluer la façon dont chaque facteur influence le résultat d'ensemble. Et la méthode mathématique de traitement n'est rien d'autre que l'application d'un raisonnement soigneux au problème considéré.»*

Une bien jolie présentation de la modélisation mathématique.

## 1.5 Signification biologique du $\mathcal{R}_0$

### 1.5.1 La capacité vectorielle

On peut décrire la signification de  $\mathcal{R}_0$  en termes biologiques :

$$\mathcal{R}_0 = \frac{a^2 b_1 b_2 A}{g_h \mu_a H}$$

Écrivons  $\mathcal{R}_0$  de la manière suivante :

$$\mathcal{R}_0 = \left( \frac{A}{H} \frac{a^2}{\mu_a} b_1 \right) \left( \frac{1}{g_h} b_2 \right) \quad (11)$$

Le premier terme est le produit de la «densité vectorielle»  $A/H$ , rapport du nombre de moustiques (le vecteur) au nombre d'humains, de la probabilité

(élevée au carré)  $a$  qu'un moustique pique un humain sur un intervalle de temps donné, de la durée de vie moyenne d'un moustique infecté ( $1/\mu_a$ ), et enfin de l'infectiosité moustique→homme  $b_1$ . Il s'agit donc de la contribution du moustique au système.

Le deuxième terme est la contribution de l'hôte humain au système : il s'agit du produit de la durée moyenne d'une infection ( $1/g_h$ ) et de l'infectiosité homme→moustique  $b_2$

On écrira donc :

$$\mathcal{R}_0 = (C_V \cdot b_1) (D_h \cdot b_2)$$

$C_V$  = Capacité vectorielle

et avec  $D_h$  = Durée d'infection

La lutte contre les moustiques contribue à abaisser  $C_V$  et donc  $\mathcal{R}_0$ . Mais nous voyons aussi qu'une augmentation de  $D_h$  augmente  $\mathcal{R}_0$ . C'est ce qui pourrait expliquer que, dans certaines régions, *Plasmodium vivax* est endémique tandis que *Plasmodium falciparum* ne l'est pas.

$\mathcal{R}_0$  exprime la balance entre l'efficacité infectieuse des moustiques ( $C_V \cdot b_1$ ) et des humains ( $D_h \cdot b_2$ ). Des moustiques très virulents ne provoqueront pas d'endémie si la durée d'infection des humains est trop courte. De même, si les humains sont infectieux pendant une grande durée et que les moustiques participent peu au système, cela n'a pas d'implication.

### 1.5.2 Nombre d'infections secondaires

Nous pouvons donner une autre description biologique à  $\mathcal{R}_0$  : imaginons un humain unique infectieux accédant à une région où ne se trouvent que des susceptibles (humains comme moustiques  $S_a = A$ ). Sa durée d'infection est  $D_h$  et il subit  $\frac{A}{H}a$  piqûres par unité de temps dont une proportion  $b_2$  infectera des moustiques. Cet humain provoque donc la production de

$$D_h \cdot \frac{A}{H} \cdot a \cdot b_2$$

moustiques infectieux.

Ces moustiques vont vivre et être infectieux pendant une durée  $1/\mu_a$  et piqueront  $a$  humains par unité de temps, piqûres dont une proportion  $b_1$  infectera ces humains. Ces moustiques vont donc infecter en retour

$$a \cdot b_1 / \mu_a$$

humains, que l'on appelle infectés secondaires après l'introduction dans le système d'un premier humain infecté.

Au total donc, un humain unique infectieux accédant à une région sans maladie provoque un nombre d'*infections secondaires* égal au produit de nos deux derniers nombres : ce qui se trouve être égal à  $\mathcal{R}_0$  !

On comprend alors bien que, si  $\mathcal{R}_0 < 1$ , chaque homme infecté contribue à en infecter trop peu d'autres et la maladie ne peut pas s'installer, tandis que

si  $\mathcal{R}_0 > 1$ , alors le nombre d'humains (et de moustiques) atteints augmente (idéalement d'un facteur  $\mathcal{R}_0$  à chaque tour) et une endémie s'installe.

On appelle  $\mathcal{R}_0$  le TAUX DE REPRODUCTION DE BASE de l'infection.

## 2 Les modèles dynamiques à base SIR

Ces modèles compartimentaux sont ceux que l'on utilise pour l'étude des infections virales ou bactériennes. Le noyau de ces approches est toujours un modèle à trois compartiments  $S$ ,  $I$  et  $R$ , que l'on complexifie par la suite avec des sous groupes différents. Le système est encore décrit par des systèmes d'équations différentielles et on y retrouve l'importance du  $\mathcal{R}_0$ .

### 2.1 Le modèle SIR de Kermack et McKendrick

À la suite des travaux de Ross, Kermack et McKendrick ont développé et publié (en 1927) d'autres modèles compartimentaux reposant sur le même type d'approche.

#### 2.1.1 Un modèle initial à trois compartiments

Nous considérons trois états de la population, trois compartiments :

**Individus susceptibles  $S$**  : ce sont des individus non-infectés, pouvant être contaminés.

**Individus infectés  $I$**  : ce sont des individus infectés, qui sont contaminants.

**Individus retirés  $R$**  : *Removed* selon Kermack et McKendrick, ce sont soit des individus guéris n'étant plus contaminants et ne pouvant plus être contaminés, soit des individus morts, ce qui revient au-même du strict point de vue de la contamination. Ils sont donc retirés du système.

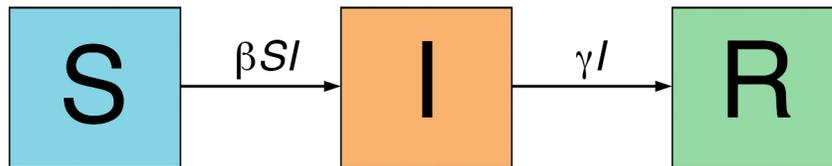


FIGURE 5 – *Modèle compartimental de Kermack et McKendrick*

Les hypothèses du modèle sont assez simples :

- La population est constante :  $N = \text{constante}$  et  $S + I + R = 1$ ; on considère que  $S$ ,  $I$  et  $R$  sont des proportions. Comme pour le modèle de Ross, cela est acceptable si l'étude (et donc l'épidémie) se fait sur une échelle de temps qui ne perturbe pas la démographie. Les populations ne

sont pas ici des effectifs, des nombres d'individus mais des densités de population.

- La contamination des susceptibles se fait avec un taux  $\beta$  (qui a la dimension de l'inverse d'un temps) et proportionnellement à la probabilité de rencontre entre susceptibles et infectés, donc proportionnellement à  $S.I$  ce qui installe un flux  $\beta SI$  du compartiment  $S$  au compartiment  $I$  ; c'est la loi d'action de masse, nommée ainsi pour son analogie forte avec la loi d'action de masse des chimistes.
- La guérison (ou la mort...) survient avec un taux  $\gamma$  (qui a aussi la dimension de l'inverse d'un temps), installant un flux  $\gamma I$  du compartiment  $I$  au compartiment  $R$ .

Ce qui permet d'écrire le modèle :

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I \\ R + I + S &= 1\end{aligned}\tag{12}$$

### 2.1.2 Le comportement du modèle

Comme le modèle est abordé à effectif constant, on peut ne considérer que les deux premiers compartiments,  $S$  et  $I$ , le troisième se déduisant des deux premiers :  $R = 1 - I - S$ . Notre étude revient donc à :

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I\end{aligned}\tag{13}$$

On peut, dans la seconde équation, remplacer  $\beta SI$  par  $-dS/dt$  et  $I$  par  $-\frac{1}{\beta S}dS/dt$ , ce qui donne :

$$\frac{dI}{dt} = -\frac{dS}{dt} + \frac{\gamma}{\beta} \frac{1}{S} \frac{dS}{dt}$$

Soit :

$$\frac{dI}{dt} = \frac{d}{dt} \left( -S + \frac{\gamma}{\beta} \ln S \right)$$

et donc :

$$I + S - \frac{\gamma}{\beta} \ln S = \text{Constante}\tag{14}$$

Nous avons déjà rencontré ce type de situation avec le modèle de Lotka et Volterra : les trajectoires du plan de phase  $(S, I)$  sont des courbes de niveau de la surface d'équation  $Z = I + S - \frac{\gamma}{\beta} \ln S$ . Mais plus simplement encore,

nous voyons que les trajectoires sont sur des courbes du plan  $(S, I)$  d'équation :  $I = \frac{\gamma}{\beta} \ln S - S + k$ , où  $k$  est une constante.

Les trajectoires du plan de phase respectent toutes :  $0 < S < 1$  ;  $0 < I < 1$  ;  $0 < S + I < 1$ . Ce qui fait que notre analyse ne se fait que dans le triangle de sommets  $(1, 0)$ ,  $(0, 1)$ ,  $(0, 0)$ .

Nous avons donc maintenant un canevas (qui dépend des valeurs de  $\gamma$  et  $\beta$ ), nous donnant le support des trajectoires. Nous savons également que  $dS/dt$  est toujours négatif (système d'équation (13)), donc que les trajectoires sont parcourues de la droite vers la gauche, les vecteurs tangents sont tous orientés de droite à gauche.

### 2.1.3 Évolution du système – théorème du seuil

Imaginons une grande population exempte d'infection,  $S = 1, I = 0$ , nous sommes au point extrême droit de notre triangle dans le plan de phase. Si nous introduisons un petit nombre d'individus infectés dans le système, nous restons dans un premier temps à proximité de ce point en  $(1 - \varepsilon, \varepsilon)$ .

On s'intéresse au rapport :

$$\frac{\beta}{\gamma}$$

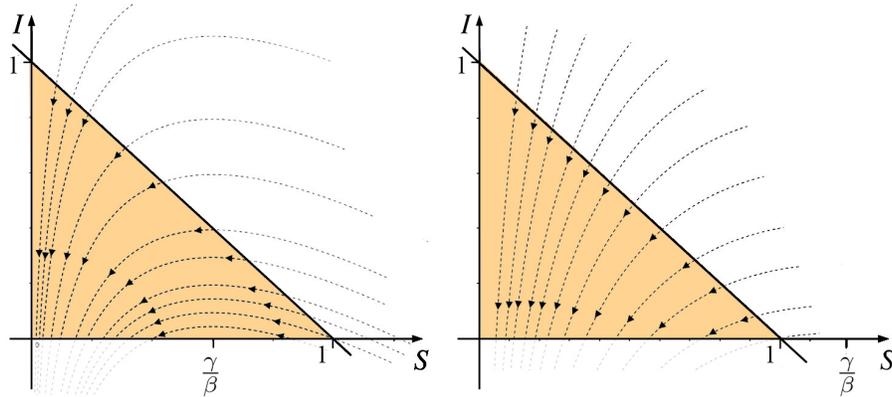


FIGURE 6 – Trajectoires du système pour deux valeurs différentes de  $\frac{\beta}{\gamma}$

Comme le montre la figure (6), toutes les trajectoires admettent un maximum pour  $S = \frac{\gamma}{\beta}$ . Ainsi, si ce terme est inférieur à 1, ce maximum est atteint dans le triangle (à gauche sur la figure) tandis que le maximum est hors champ lorsque  $\frac{\gamma}{\beta} > 1$ . Comme toutes les trajectoires vont de la droite vers la gauche, cela montre que dans le premier cas il y a une épidémie (le nombre de  $I$  augmente) suivi d'une extinction de l'épidémie ( $I$  revient à 0) ; dans le second cas, il n'y a jamais d'épidémie, les trajectoires sont toutes descendantes.

En résumé :

- (1) les trajectoires montrent un maximum pour  $I$  lorsque  $S = \frac{\gamma}{\beta}$

- (2) le maximum n'est dans le triangle que si  $\frac{\gamma}{\beta} < 1$
- (3) toute épidémie finit par s'éteindre, avec  $S$  non nul

Nous définissons alors :

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

ce qui nous permet d'énoncer le théorème du seuil (analogue au *Mosquito theorem*) :

Partant de  $(1 - \varepsilon, \varepsilon)$ , il n'y aura pas d'installation possible d'une épidémie si  $\mathcal{R}_0 < 1$ . En revanche, si  $\mathcal{R}_0 > 1$ , il y aura déclenchement d'une épidémie.

Et dans ce modèle, il n'y a pas d'endémie possible : toute épidémie finit par disparaître. En effet :

Dans le cas d'une épidémie (si  $\mathcal{R}_0 > 1$ ), le système évolue vers un point de coordonnées  $(S_\infty, 0)$ , c'est-à-dire avec plus aucun individu infecté ( $I = 0$ ) et avec :

$$S_\infty = \lim_{t \rightarrow \infty} S(t)$$

Il subsiste, en fin d'épidémie,  $S_\infty$  individus susceptibles. Tous les autres sont dans le compartiment  $R$ , c'est-à-dire soit immuns soit morts.

#### 2.1.4 Signification biologique de $\mathcal{R}_0$

Il est également possible ici de donner une signification à  $\mathcal{R}_0$ .

Considérons la situation où un individu unique est infecté dans une population de grand effectif  $N$ . Dans cette situation, nous avons  $I = \varepsilon$  et  $S = 1 - \varepsilon$ , où  $\varepsilon = \frac{1}{N}$  est petit. Si nous considérons alors que  $S \approx 1$ , l'équation différentielle pour  $I$  devient :

$$\frac{dI}{dt} = \beta I - \gamma I$$

Ce qui nous mène à une croissance malthusienne, exponentielle, de la population d'infectés avec un taux de croissance *per capita* égal à  $\beta - \gamma$ . L'installation de l'épidémie n'est possible que si ce taux est positif, ce qui revient à  $\beta > \gamma$  ou encore  $\mathcal{R}_0 > 1$ .

Nous pouvons aussi nous demander combien d'individus infectés supplémentaires notre infecté unique initial peut produire. L'individu reste infecté pendant la durée  $\tau = \frac{1}{\gamma}$ . Et pendant cette durée  $\tau$ , il va infecter une proportion  $\tau\beta SI$  de la population (d'effectif  $N$ ).

Ce qui fait un nombre d'individus infectés égal à :

$$\tau\beta S.I.N$$

Rappelons nous que  $S \approx 1$ , que  $I = \varepsilon = \frac{1}{N}$ . Il vient donc que le nombre de cas secondaires que peut provoquer un individu unique infecté est :

$$\tau\beta \text{ c'est-à-dire } \frac{\beta}{\gamma} \text{ c'est-à-dire } \mathcal{R}_0$$

Ici aussi, donc, notre paramètre  $\mathcal{R}_0$  a la signification du taux de reproduction de base de l'infection (nombre de cas secondaires provoqués par un infecté unique). Ce paramètre est une clé essentielle de l'épidémiologie. L'une des premières questions à résoudre devant une épidémie est celle de la valeur de  $\mathcal{R}_0$

### 2.1.5 Les prédictions du modèle SIR

La simulation (discrétisation du système d'équations différentielles dans un tableur) de la figure (7) montre bien l'apparition puis la disparition de l'épidémie, avec *in fine* un nombre S non nul et un nombre R assez conséquent. De nombreuses questions peuvent alors se poser :

- quelle sera la durée de l'épidémie ?
- quelle sera la valeur du pic de l'épidémie ?
- quelle sera la valeur finale de S, nombre d'individus non touchés ?
- quel sera le nombre de morts ?
- et bien entendu : quels sont les moyens d'action ?

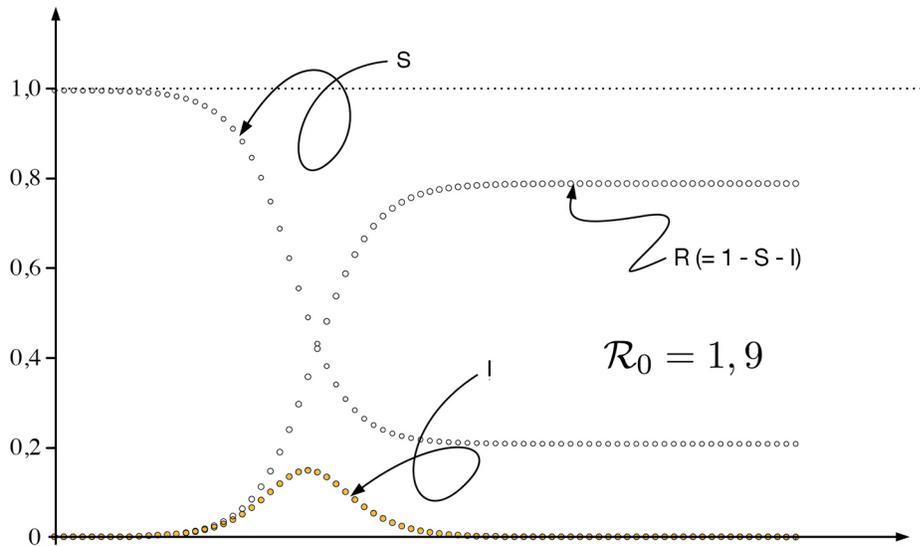


FIGURE 7 – Simulations d'un modèle SIR avec  $\mathcal{R}_0 = 1,9$ .

Si nous reprenons l'équation (14) et que nous faisons tendre  $t$  soit vers 0 soit vers l'infini, nous obtenons :

$$\begin{aligned}
 t \text{ quelconque} \quad I + S - \frac{\gamma}{\beta} \ln S &= K \\
 t \rightarrow 0 \quad 1 &= K \\
 t \rightarrow \infty \quad S_\infty - \frac{\gamma}{\beta} \ln S_\infty &= K = 1
 \end{aligned}
 \tag{15}$$

D'où l'on peut tirer :

$$I = 1 - S + \frac{\gamma}{\beta} \ln S \quad (16)$$

et

$$\frac{\gamma}{\beta} = \frac{\ln S_\infty}{S_\infty - 1} \quad (17)$$

Ainsi, nous pouvons en déduire des informations précieuses :

Le maximum de l'épidémie ( $\frac{dI}{dt} = 0$ ) est atteint lorsque  $S = \gamma/\beta = 1/\mathcal{R}_0$

En utilisant l'équation (16), nous pouvons alors trouver la valeur maximale de  $I$  (lorsque  $S = \gamma/\beta$ ), soit :

$$I_{max} = 1 - \frac{\gamma}{\beta} + \frac{\gamma}{\beta} \ln \frac{\gamma}{\beta} \quad (18)$$

ou encore

$$I_{max} = 1 - \frac{1}{\mathcal{R}_0} - \frac{1}{\mathcal{R}_0} \ln \mathcal{R}_0 \quad (19)$$

On comprend tout l'intérêt qu'il y a à connaître la valeur de  $\mathcal{R}_0$  !

Et comme on peut, cliniquement, avoir un ordre de grandeur de la durée  $\tau$  d'une infection (temps moyen de guérison ou temps moyen de survie), on a une estimation de  $\gamma = 1/\tau$ , ce qui permet, *in fine*, de déterminer la valeur de  $\beta$ , paramètre important de l'infection.

Enfin, concernant le nombre de morts, on peut connaître la valeur de  $R_\infty$  puisque  $I_\infty = 0$  et donc  $R_\infty + S_\infty = 1$ . Si l'on connaît la létalité de la maladie (nombre de morts par cas d'infections) on peut connaître le nombre de morts attendus en fin d'épidémie.

## 2.2 Peut-on appliquer cela à l'infection à Coronavirus ?

L'infection à Coronavirus est modélisée à l'aide de modèles compartimentaux beaucoup moins grossiers que le modèle SIR. Mais il y a tout de même moyen de dégager quelques idées de ce modèle SIR.

### 2.2.1 Valeur de $\mathcal{R}_0$ et nombre d'infectés au moment du pic

Connaître  $\mathcal{R}_0$ , c'est pouvoir prédire la valeur du pic, et le nombre de susceptibles au moment du pic.

La valeur de  $\mathcal{R}_0$  est très difficile à connaître ; elle peut être estimée à partir de la croissance de l'épidémie en début d'infection. L'analyse de l'infection sur le bateau de croisière *Diamond Princess* a permis également d'approcher ce paramètre.

Pour une valeur de 2,4 à 2,5 comme on nous l'annonce (mais cette valeur, plausible, n'est pas du tout certaine, certains vont jusqu'à 5), cela fait 40% des individus non-infectés au moment du pic (le pic d'épidémie survient lorsque  $S$  est égal à  $\gamma/\beta$  c'est à dire  $1/\mathcal{R}_0$ ) ! C'est à dire un taux d'infection cumulée de près de 60% de la population au moment du pic.

### 2.2.2 Valeur de $\mathcal{R}_0$ et fin de l'épidémie

En fin d'épidémie,  $S$  tend vers une valeur  $S_\infty$  solution de l'équation (17) ou encore solution de :

$$S_\infty = 1 + \frac{1}{\mathcal{R}_0} \ln S_\infty \quad (20)$$

Il y a une solution évidente qui est  $S_\infty = 1$ , mais cette solution ne nous intéresse pas : elle correspond à l'absence de toute infection.

Il est possible de montrer que lorsque  $\mathcal{R}_0 > 1$  (épidémie), il y a une solution unique différente de 1. On peut construire graphiquement cette solution en repérant l'intersection entre la droite d'équations  $y = x$  et la courbe d'équation  $y = 1 + \frac{1}{\mathcal{R}_0} \ln x$ .

On constate alors que  $S_\infty$  est une fonction décroissante de  $\mathcal{R}_0$  (ce que l'on peut démontrer, également). Nous comprenons ici encore que si nous voulons que le moins possible de personnes soient atteintes ( $S_\infty$  le plus grand possible) nous avons intérêt à avoir la valeur la plus faible possible de  $\mathcal{R}_0$ .

Pour  $\mathcal{R}_0 = 2, 4$ , le modèle SIR prédit une valeur de  $S$  de 0, 12, ce qui signifie 88% de personnes touchées par l'infection. Ceci pour dire qu'il peut y avoir un grand nombre d'infectés si rien n'est fait. La valeur obtenue (88%) ayant peu de signification, le modèle étant trop grossier.

### 2.2.3 Combattre l'épidémie par l'immunité de groupe

C'est la conséquence directe de ce qui précède : si l'on ne fait rien, un nombre de plus en plus grand d'individus seront immunisés. Lorsque 60% de la population aura été atteinte, la décroissance viendra d'elle même. C'est la stratégie (très critiquée) un temps envisagée par les autorités britanniques.

### 2.2.4 Combattre l'épidémie par la vaccination

Si un vaccin existe, il est possible de lutter contre l'épidémie sans nécessairement vacciner toute la population. C'est l'immunité de groupe à laquelle on donne un petit coup de pouce.

Supposons que nous vaccinons une proportion  $V$  de la population. Alors une proportion  $VS$  est protégée, et seule une proportion  $(1 - V)S$  est sensible à l'infection. Le modèle s'écrit alors de la même manière, mais avec un terme  $(1 - V)S$  au lieu de  $S$ . Ce qui donne le système :

$$\begin{aligned} \frac{dS}{dt} &= -\beta(1 - V)SI \\ \frac{dI}{dt} &= \beta(1 - V)SI - \gamma I \end{aligned} \quad (21)$$

Ce qui est exactement le même modèle SIR, avec une nouvelle valeur

$$\beta' = (1 - V)\beta$$

Ce qui nous mène alors à une nouvelle valeur pour  $\mathcal{R}_0$  :

$$\mathcal{R}'_0 = (1 - V)\mathcal{R}_0$$

La vaccination sera un succès si nous parvenons à vacciner suffisamment de personnes pour amener ce  $\mathcal{R}'_0$  en-dessous de 1. Ce qui sera réalisé si :

$$V > 1 - \frac{1}{\mathcal{R}_0}$$

Dans le cas de la variole, ce fut difficile, car avec  $\mathcal{R}_0 = 3$ , il a fallu parvenir à vacciner plus de 66,7 % de la population pour éradiquer la maladie. Dans le cas de la rougeole, c'est peine perdue. La vaccination est bien entendu essentielle pour protéger la population et sauver des vies, mais avec  $\mathcal{R}_0 = 16$ , éradiquer la rougeole supposerait une vaccination de plus de 93,7 % de la population, ce qui semble irréaliste. Il y a une endémie inévitable de rougeole.

Dans le cas du coronavirus, la valeur de  $\mathcal{R}_0 = 2,4$  rend possible une éradication par vaccination à un taux de 41,7 %.

### 2.2.5 Les mesures d'hygiène, les "gestes barrière"

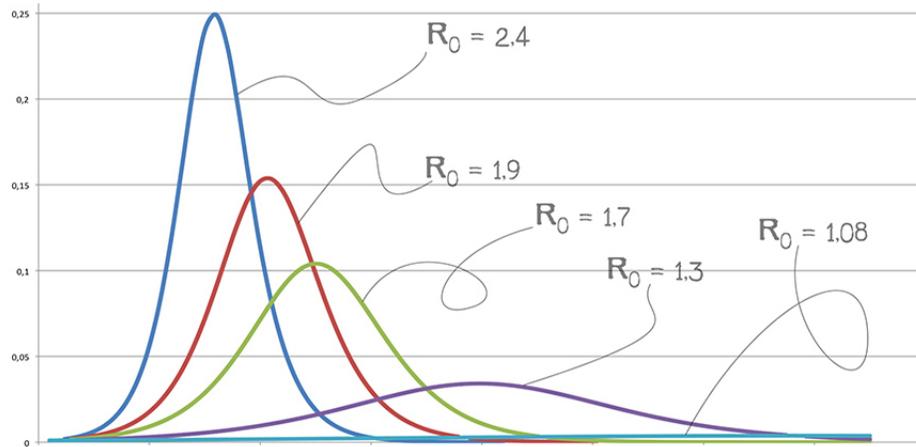


FIGURE 8 – Evolution du nombre d'infectés dans un modèle SIR pour différentes valeurs de  $\mathcal{R}_0$ . La diminution de  $\mathcal{R}_0$  a été obtenue en ne jouant que sur la valeur de  $\beta$

En dehors de la vaccination, il reste l'espoir de jouer sur la valeur de  $\mathcal{R}_0 = \frac{\beta}{\gamma}$ . C'est à dire, soit augmenter les valeur de  $\gamma$ , soit diminuer la valeur de  $\beta$ .

Augmenter la valeur de  $\gamma$  signifie "soigner les malades", ou encore diminuer la durée de la maladie, ou encore "guérir" les malades. Toutes choses que l'on ne sait que très peu faire à l'heure actuelle. Des espoirs sont fondés sur certaines molécules, elles permettront peut-être d'augmenter  $\gamma$ .

En l'absence de possibilité réelle de soigner la maladie (pas d'antiviraux disponibles) il ne reste que l'objectif de diminuer la valeur de  $\beta$ , c'est à dire enrayer la transmission entre personnes. Les moyens sont simples à énoncer : (i) dépister précocement les malades avec des tests plus nombreux (comme en Allemagne) permet de soigner et d'isoler plus tôt, (ii) les gestes d'hygiène, les "gestes barrière" ralentissent la progression du virus, de même que (iii) les mesures de confinement. C'est le seul levier dont on dispose. Les gestes barrière sont donc tout à fait essentiels. Ce n'est pas un pis aller, c'est le moyen de diminuer l'impact de la maladie.

Lorsque l'on réalise des simulations (avec un tableur, par exemple), on constate bien l'effet d'une diminution de la valeur de  $\beta$  : le pic est abaissé, il est retardé, et l'infection est étalée dans le temps (figure (8)). On sait par ailleurs que la diminution de la valeur de  $\mathcal{R}_0$  permet une augmentation de la valeur de  $S_\infty$  et donc une diminution de la valeur de  $R_\infty$  : le nombre final de personnes touchées est moindre.

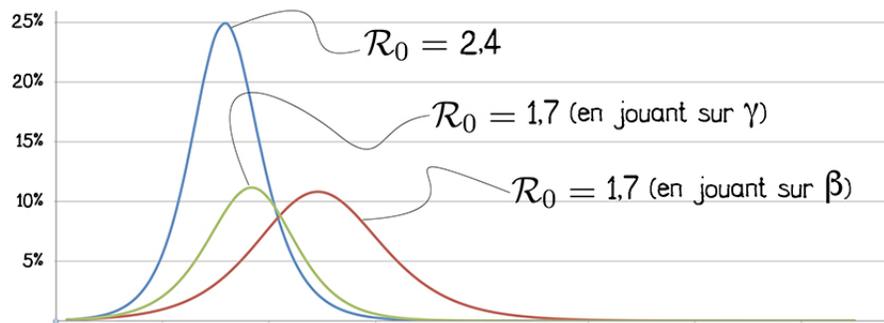


FIGURE 9 – Effets comparés d'une baisse de  $\beta$  et d'une hausse de  $\gamma$  pour faire baisser  $\mathcal{R}_0$ .

Si la diminution de  $\mathcal{R}_0$  était effectuée en jouant sur  $\gamma$  au lieu de  $\beta$ , l'effet serait comparable, mais avec beaucoup moins de décalage dans le temps (figure (9)). En jouant sur  $\beta$ , on étale la courbe et on la décale ; en jouant sur  $\gamma$ , on ne fait presque que l'étaler. Pouvoir soigner les malades permettrait donc de réduire la durée de l'épidémie.

Toutefois, on voit sur le graphe représentant  $I_{max}$  en fonction de  $\mathcal{R}_0$  (figure (10), en utilisant l'équation (19)) que la pente de la courbe est assez faible. Il faut un gros effort sur  $\beta$  pour espérer un effet significatif sur  $I_{max}$ . Les consignes de confinement ne doivent pas être prises à la légère. Plus draconiennes sont les mesures de confinement, meilleurs sont les résultats, ce qui bien entendu tient plus du bon sens que des mathématiques.

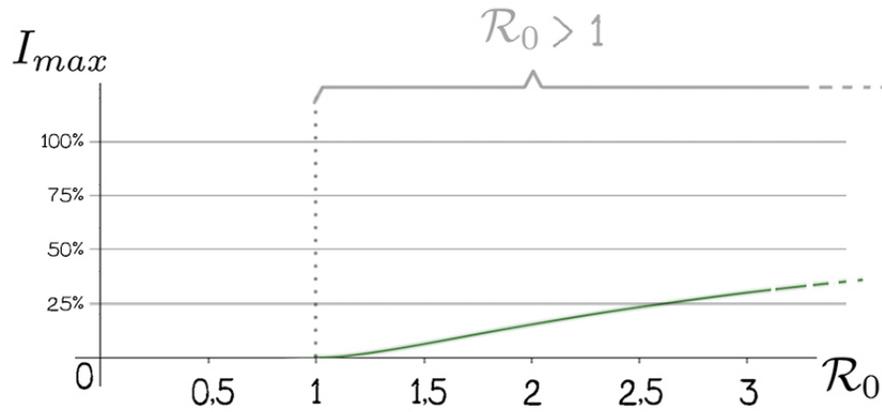


FIGURE 10 – Evolution de la valeur du pic d’infection ( $I_{max}$ ) en fonction de  $\mathcal{R}_0$ .

## 2.3 Diversité des modèles compartimentaux

### 2.3.1 Aspects démographiques

Dans le modèle de Kermack et McKendrick, il n’y a pas de fluctuation démographique : il ne tient pas compte des naissances et des décès. Il est possible d’intégrer ces aspects démographiques en définissant une natalité  $\nu$  et une mortalité  $\mu$ , avec toutefois deux contraintes :

- la mortalité est la même pour chaque sous-population ;
- la population reste constante et donc  $\nu = \mu$ .

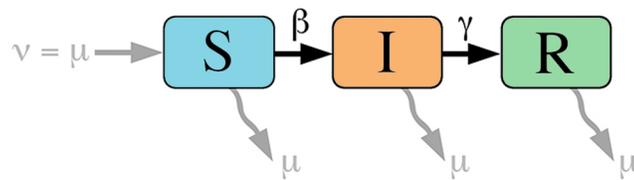


FIGURE 11 – Modèle compartimental  $SIR$  avec démographie, c’est-à-dire avec prise en compte des naissances et des décès.

Dans ces conditions, le modèle est modifié de la façon suivante :

$$\begin{aligned}
 \frac{dS}{dt} &= \nu N - \beta SI - \mu S = \mu(N - S) - \beta SI \\
 \frac{dI}{dt} &= \beta SI - \gamma I - \mu I \\
 \frac{dR}{dt} &= \gamma I - \mu R \\
 R + I + S &= 1
 \end{aligned}
 \tag{22}$$

### 2.3.2 Modèle SIRS

Ce modèle reprend les données du modèle SIR (avec démographie), en ajoutant une modification : les membres de la classe  $R$  peuvent revenir à l'état  $S$ , libres de toute infection mais à nouveau susceptibles d'être infectés (et nous conservons l'égalité  $S + I + R = 1$ ).

Nous introduisons donc juste un paramètre  $\lambda$ , taux de ce retour  $R \rightarrow S$ .

$$\begin{aligned}
 \frac{dS}{dt} &= \mu(N - S) - \beta SI + \lambda R \\
 \frac{dI}{dt} &= \beta SI - \gamma I - \mu I \\
 \frac{dR}{dt} &= \gamma I - \mu R - \lambda R
 \end{aligned}
 \tag{23}$$

Ce type de modèle peut aboutir à une situation endémique stable comme le montre le portrait de phase de la figure 12.

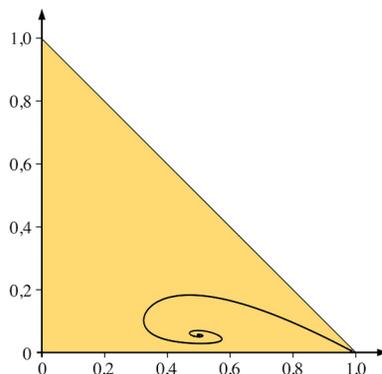


FIGURE 12 – Portrait de phase d'un modèle SIRS aboutissant à un équilibre endémique.

### 2.3.3 Autres modèles

Donnons un dernier exemple avec le modèle SEIR : nous introduisons maintenant une nouvelle classe  $E$  (pour *exposed*) d'individus infectés mais non infectieux. Le modèle s'écrira alors (avec toujours  $R + I + E + S = 1$ ) :

$$\begin{aligned} \frac{dS}{dt} &= \mu(N - S) - \beta SI \\ \frac{dE}{dt} &= \beta SI - (\mu + \varepsilon)E \\ \frac{dI}{dt} &= \varepsilon E - (\gamma + \mu)I \\ \frac{dR}{dt} &= \gamma I - \mu R \end{aligned} \tag{24}$$

On peut aussi faire évoluer ce modèle vers une structure SEIRS, etc.

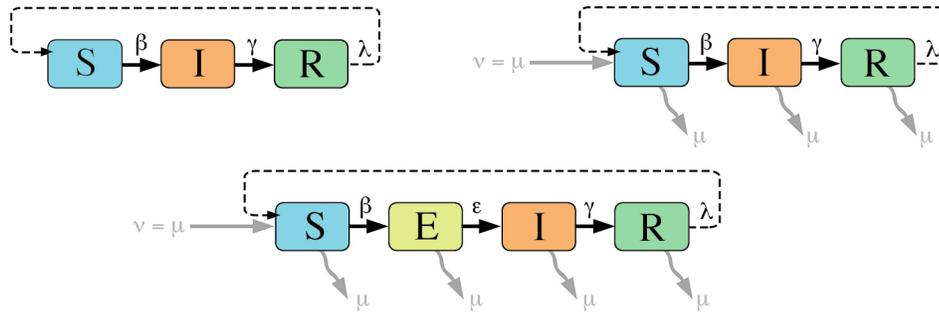


FIGURE 13 – Différents modèles dérivés du modèle SIR, avec ou sans retour à l'état  $S$  (la ligne en tiretés est facultative), avec ou sans structure démographique.

## 2.4 Un modèle pour le coronavirus

Si l'on se reporte par exemple à l'article de l'inserm :

**Expected impact of school closure and telework to mitigate COVID-19 epidemic in France** (Laura Di Domenico, Giulia Pullano, Pietro Coletti, Niel Hens, Vittoria Colizza),

on trouve l'utilisation d'un modèle SEIIR dans lequel les personnes infectées sont réparties en deux sous-groupes :

$I_p$  : les infectieux présymptomatiques

$I_s$  : les infectieux symptomatiques.

Cela permet d'affiner le suivi de différents compartiments en ajustant de nombreux paramètres sociétaux. Le principe reste toutefois le même, les simulations étant juste plus lourdes à mener.

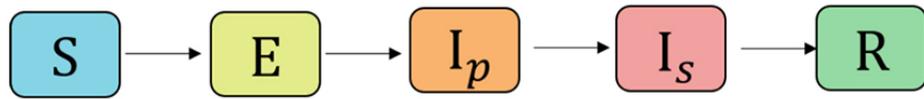


FIGURE 14 – Un modèle SEIIR.