



HAL
open science

Modélisation multi-échelle de la persistance des infections virales humaines : approches stochastiques et statistiques

Thomas Beneteau

► **To cite this version:**

Thomas Beneteau. Modélisation multi-échelle de la persistance des infections virales humaines : approches stochastiques et statistiques. Maladies infectieuses. Université de Montpellier, 2023. Français. NNT : 2023UMONG040 . tel-04652571

HAL Id: tel-04652571

<https://theses.hal.science/tel-04652571>

Submitted on 18 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie des Populations et Écologie

École doctorale GAIA

Unité de recherche MIVEGEC

Modélisation multi-échelle de la persistance des
infections virales humaines : approches
stochastiques et statistiques.

Présentée par Thomas BÉNÉTEAU
Le 11 décembre 2023

Sous la direction de Samuel ALIZON
et Mircea T. SOFONEA

Devant le jury composé de

Pauline EZANNO, DR, INRAE
Amandine VÉBER, DR, CNRS
Jean-Michel MARIN, PU, Université de Montpellier
Mélanie PRAGUE, CR, INRIA
Jean-Luc PRÉTET, PU-PH, Université de Franche-Comté
Samuel ALIZON, DR, CNRS
Mircea T. SOFONEA, MC, Université de Montpellier
Christian SELINGER, CR, STPH

Rapportrice
Rapportrice
Président du jury
Examinatrice
Examineur
Directeur
Co-encadrant
Invité



UNIVERSITÉ
DE MONTPELLIER

Résumé

Pour des raisons pratiques, éthiques ou scientifiques, la modélisation mathématique offre une alternative aux approches en biologie expérimentale et fondamentale dans l'analyse du vivant. Cependant, les systèmes biologiques sont difficiles à appréhender mathématiquement. Il est ainsi nécessaire de simplifier la réalité biologique afin de mettre en équation les processus en jeu. Toutefois, il n'est pas possible de décrire un système biologique dans sa globalité avec un seul modèle. La diversité des approches employées et leur complémentarité permettent d'apporter des éléments concrets sur la caractérisation des mécanismes biologiques.

À travers l'exemple des infections transitoires par les papillomavirus humains (HPV), je détaille trois approches complémentaires qui permettent de mieux cerner la dynamique infectieuse du virus. En premier lieu, je décris les motifs de détection du virus à l'aide des outils de l'analyse de survie. Dans un second temps, je m'intéresse plus spécifiquement à l'impact de phases latentes dans la dynamique infectieuse HPV en utilisant des modèles de Markov caché. Enfin, je propose un modèle stochastique pour évaluer la possibilité d'élimination des infections virales par des phénomènes autres que ceux du système immunitaire.

Dans une moindre mesure, j'ai contribué avec les membres de l'équipe à l'analyse de la première vague de l'épidémie de COVID-19 en France, avec une attention particulière pour l'estimation de la date d'émergence et de la date d'extinction de l'épidémie.

Individuellement, ces modèles ne décrivent qu'un aspect spécifique d'un phénomène biologique, mais collectivement, ils permettent d'appréhender les caractéristiques et enjeux du système étudié.

Abstract

For practical, ethical or scientific reasons, mathematical modeling offers an alternative to experimental and fundamental biological approaches to the analysis of living systems. However, biological systems are difficult to understand mathematically. It is therefore necessary to simplify biological reality in order to translate the processes involved into an equation. However, it is generally impossible to describe a biological system in its entirety with a single model. The diversity and complementarity of the approaches employed can provide concrete insights into the characterization of biological mechanisms.

Using the example of transient human papillomavirus (HPV) infections, I describe three complementary approaches to better understand the infectious dynamics of the virus. First, I describe the patterns of virus detection using survival analysis tools. Secondly, I focus more specifically on the impact of latent phases in HPV infectious dynamics, using hidden Markov models. Finally, I propose a stochastic model to evaluate the possibility of eliminating viral infections by phenomena other than those of the immune system.

To a lesser extent, I have contributed with team members to the analysis of the first wave of the COVID-19 epidemic in France, with a particular focus on estimating the date of emergence and the date of extinction of the epidemic.

Individually, these models only describe a specific aspect of a biological phenomenon, but collectively they enable us to grasp the characteristics and challenges of the system investigated.

Contexte de la thèse

Ce travail de thèse a été réalisé au sein de l'unité mixte de recherche (UMR) MIVEGEC (Maladies Infectieuses et Vecteurs : Écologie, Génétique, Évolution et Contrôle). L'encadrement et l'accompagnement ont été assurés par Samuel Alizon et Mircea T. Sofonea, ainsi que Christian Selinger.

Cette thèse est axée autour d'une thématique interdisciplinaire, à la frontière entre la virologie et les mathématiques. En particulier, j'ai cherché au cours de ma thèse à utiliser des outils variés de diverses branches des mathématiques (statistiques ou modélisation stochastique) pour apporter des éléments de compréhension de certains phénomènes biologiques sous le prisme de l'aléatoire.

Cette thèse s'est étalée sur quatre ans, de décembre 2019 à décembre 2023. Pour les trois premières années, j'ai reçu un premier financement de la Ligue contre le Cancer pour travailler sur les infections du papillomavirus humain (HPV). Au cours de ces trois ans, je me suis principalement focalisé sur les infections HPV du point de vue intra-hôte, même si dans les premiers mois de la pandémie de COVID-19 j'ai participé avec les membres de l'équipe à des travaux de modélisation de l'épidémie de SARS-CoV-2 en France. Ensuite, j'ai obtenu une prolongation pour une quatrième année de la part de la Ligue contre le Cancer pour poursuivre des travaux sur les infections HPV qui avaient été retardés par des complications administratives ou consécutives de la pandémie. Je tiens à remercier également la Société Européenne de Biologie Évolutive (ESEB : *European Society of Evolutionary Biology*) pour avoir financé un projet monté avec Katrina Lythgoe, chercheuse au *Big Data Institute* d'Oxford, portant sur la dynamique des infections chroniques du virus de l'Hépatite B. Cette bourse m'a permis de me déplacer sur place en février - mars 2022 afin de rencontrer Katrina et discuter du projet avec des virologues de l'équipe de Jane McKeating du département de Médecine de

Nuffield.

Mon travail au cours de ces quatre ans s'organise sous forme de thèse sur articles en six chapitres principaux. Le premier chapitre introduit les concepts fondamentaux de la biologie virale et les bases des outils mathématiques utilisés au cours de la thèse. S'ensuit quatre chapitres présentant les différents travaux menés dont deux articles publiés (Chapitre 4 et 5), un article actuellement en *preprint* (Chapitre 2) et un projet en cours de finition (Chapitre 3). Enfin, nous terminerons la partie principale de ce manuscrit par une discussion récapitulant et mettant en perspective les divers travaux menés durant ces quatre ans. Le contenu des articles a été reproduit à l'identique dans les parties correspondantes mais la mise en page a été uniformisée pour correspondre à celle de ce document. Les DOI des articles et du *preprint* sont fournis dans les sections correspondantes. La bibliographie est insérée à la fin de la partie centrale du document, avant les annexes. Un glossaire reprenant l'ensemble des abréviations utilisées dans le manuscrit est disponible p.210 et la liste des figures et tableaux est disponible p.212. J'ai créé les images introduisant les chapitres grâce à Midjourney. Les prompts utilisés pour les générer sont disponibles à la fin du manuscrit, après la liste des Figures et Tableaux.

Remerciements

En premier lieu, je tiens à exprimer ma plus grande gratitude à Amandine Veber et Pauline Ezanno d'avoir accepté le rôle de rapportrice de ma thèse ainsi qu'à tous les membres du jury, Jean-Michel Marin, Mélanie Prague et Jean-Luc Prétet pour avoir accepté de juger mes travaux de thèse.

Je tiens également à remercier tous les membres de mon comité de suivi de thèse, François Rousset, Nacho Bravo et Guillaume Martin qui ont gentiment accepté de me suivre et me conseiller durant ces quatre années de thèse.

Je remercie la Ligue contre le Cancer de m'avoir fait confiance et d'avoir financé mes quatre années de thèse.

Évidemment un grand merci à Samuel, Mircea pour votre soutien sans faille au cours de ces presque cinq années passées à vos côtés. Je me sens chanceux d'avoir eu des encadrants toujours disponibles (à des heures beaucoup trop tardives voire même le dimanche) et bienveillants pour répondre à mes interrogations et m'aider dans ces années d'apprenti chercheur. Votre optimisme constant a été précieux dans ces travaux remplis d'interrogations et de remise en question perpétuelle. Merci également à Christian pour ton aide dans ces premières années de thèse. Enfin, un immense merci à Yannis pour avoir généreusement pris en charge la responsabilité administrative de ma dernière année de thèse, avec le surplus de mails inopinés que cela engendre.

Un petit sentiment de nostalgie m'envahit à l'idée de quitter ce grand lieu qu'est le bâtiment Païre, plus communément appelé Kayak. Malgré la décrépitude totale du bâtiment et l'isolation inexistante, je garde un souvenir affectueux de ce local au milieu des champs à l'ambiance si

singulière, tantôt plein de vie et d'animation qui résonne à travers les couloirs, tantôt désaffecté lors des lourdes chaleurs d'été. Les flux de personnes ont été permanents et je ne me risquerai pas à lister l'ensemble des personnes impliquées dans cette effervescence perpétuelle mais je tiens à vous adresser collectivement un grand merci pour ces moments de convivialité, notamment les midis ou les rares fois où je me suis aventuré hors du bureau.

Bon courage aux futurs doctorants MIVEGEC pour la fin de vos travaux respectifs. Bien que tardif, ce fut un plaisir de venir boire des bières avec vous et de discuter ensemble de sujets souvent fumeux.

Dans un autre registre, je veux exprimer toute ma gratitude au collègue Jean Yole des Herbiers pour avoir écrit ces paroles inspirantes sur le mur extérieur des salles de technologie, phrases qui résonnent encore en moi et que j'espère avoir partagé au plus grand nombre.

Évidemment, un petit coucou à toutes les personnes qui sont ou ont été membre de l'équipe ETE, EEH ou encore d'un autre acronyme, à Montpellier comme à Paris. J'ai envie de vous remercier un à un mais je sais que je vais oublier une (probablement plusieurs) personne et je souhaite me préserver de la gêne commune que cela occasionnerais. Bref un grand merci à vous tous pour ces bières, ces repas ou ces discussions partagés au boulot comme en conférences. Bon, je me dois de faire exception à la règle (on est français après tout) pour adresser une petite dédicace aux trois piliers de bar du bâtiment Païre qui ont partagé le même bureau et ma compagnie durant beaucoup trop longtemps pour que cela n'ait pas laissé de trace dans leur esprit (notamment l'évangile selon Matthieu verset 7 :8).

Je remercie Bastien pour ses rappels réguliers aux droits salariaux des doctorants, en particulier sur le respect des 35 heures, et son bon sens paysan que l'on pourrait également qualifier de sagesse de retraité (pas si éloigné de ses goûts musicaux finalement).

Merci également à Gonché qui a partagé ce bureau avec moi pendant quelques années au péril de sa santé mentale. Merci évidemment pour tous tes TIPS (mdr tmtc) très précieux tant en R qu'en shell!

Bon courage à toi Baptiste, toi qui remonte à la capitale pour de nouvelles aventures palpitantes (j'espère que tu en profiteras pour faire des soirées quai de Seine 8.6 Picon) et repars pour une quatrième année. Merci pour tes imitations (notamment de Sarkozy) et ton sens de la ponctualité sans faille!

Je profite de ces pages pour faire une petite bise à toutes les personnes qui m'ont accompagné au cours des diverses pérégrinations scolaires dont l'aboutissement est finalement cette thèse.

Une bise lointaine vers ces contrées verdoyantes du bocage vendéen et destinée à tous ces joyeux loubards et loubardes qui sont, pour la plupart,

restés pas trop loin du nid. Bon courage à ceux et celles qui doivent ou vont devoir affronter les cris et les pleurs de leur(s) chérubin(s) dans les mois/années à venir. On se revoit vite je l'espère pour un évènement festif de grande ampleur ou pour une soirée tranquille à la cave.

Un petit coucou aux zigotos de classes préparatoires, notamment à ceux que j'ai eu le hasard de croiser à Montpellier au détour de la barbote ou lors d'autres traquenards. Je ne vous oublie pas, vous autres zinzins castrateurs de chiots ou de canards, j'espère vous revoir prochainement autour d'une bière ou d'un shooter (souvent pas très bon) en souvenir de ces années, au final sympathiques, passées à vos côtés.

Un big up à tous les guignols de la 190, essayés ici et là en France et en Navarre. On se revoit sûrement prochainement à Quedillac, dans un trou paumé en île de France ou encore dans le pays Basque (*Bainan horrela Ez zen gehiago txoria izango*). J'espère que vous pourrez, et le cas échéant penserez à, m'appeler docteur à ces occasions. Un petit bisou supplémentaire aux sept zinzins du bâtiment B pour m'avoir supporté dans maintes occasions insoutenables.

Un grand merci à Thierry, Marielle et Mona de m'avoir gentiment accueilli dans leur îlot de verdure et de tranquillité pendant les multiples périodes de confinement.

Une tendre pensée pour Diane, témoin privilégiée de mes passions et lubies incongrues. Je ne te remercierai jamais assez pour ta présence à mes côtés depuis toutes ces années et pour tout ce que tu m'as apporté de bonheur et de franches rigolades. J'ai hâte qu'on aille vivre ensemble d'autres aventures dans le pacifique ou ailleurs avec Gargouille et Bonzigue.

Je terminerai en remerciant, Maxime, papa et maman, pour votre soutien aveugle et indéfectible au cours de ces longues années d'études supérieures. Merci d'avoir toujours été présent et attentif à ma réussite. J'espère qu'à la lecture de ce manuscrit vous comprendrez un peu mieux ce que je bricole depuis quatre ans!

Table des matières

1	Introduction	1
1.1	Éléments clés de la dynamique virale	3
1.1.1	Biologie fondamentale d'un virus	5
1.1.2	Des voies de transmission variées	7
1.1.3	Interactions hôte-virus	9
1.2	Histoire naturelle et épidémiologie des virus étudiés . . .	12
1.2.1	HPV	12
1.2.2	SARS-CoV-2	17
1.3	La modélisation mathématique, outil clé de l'épidémiologie	20
1.3.1	Modèles compartimentaux en épidémiologie . . .	22
1.3.2	Analyse de survie	34
1.3.3	Modèle multi-états et modèle de Markov caché . .	41
1.4	Objectifs et aperçu de la thèse	45
2	Caractérisation des infections transitoires HPV chez les jeunes femmes	47
2.1	Aperçu de l'article	48
2.2	Article	51
3	Inférence de la dynamique infectieuse HPV par chaînes de Markov cachées	63
3.1	Introduction	64
3.2	Matériel et méthodes	66
3.2.1	Études cliniques et participantes	66
3.2.2	Analyse statistique	68
3.3	Résultats préliminaires	72
3.3.1	Analyse descriptive	72
3.3.2	Description des scénarios	74

3.3.3	Temps de retour à l'état sain	74
3.3.4	Effet de la vaccination	78
3.4	Discussion	78
4	Modélisation du cycle de vie HPV : rôle de l'aléatoire dans la clairance des infections	82
4.1	Aperçu de l'article	83
4.2	Article	85
5	Modélisation de la première vague de COVID-19 en France	101
5.1	Aperçu de l'article	102
5.2	Article	105
6	Conclusion et perspectives	121
A	Annexes et suppléments d'articles	177
A.1	Infections transitoires HPV	177
A.2	Modèle de Markov caché pour les infections HPV	184
A.3	Modélisation du cycle de vie HPV	189
A.4	Modélisation de la première vague du COVID-19	200
B	Abréviations	210
C	Liste des figures et tableaux	212
C.1	Figures	212
C.2	Tableaux	217
C.3	Prompts midjourney	218

CHAPITRE

1

Introduction



Les infections virales occupent une place majeure dans nos sociétés modernes. En termes de santé publique, de nombreux virus font l'objet de surveillance accrue. L'expérience récente de la pandémie de COVID-19 a mis en lumière l'impact que la propagation à grande échelle d'un virus respiratoire pouvait causer à l'échelle globale. Le coût tant humain [Msemburi et al., 2023] qu'économique [Ozili and Arun, 2020] de diffusion du virus SARS-CoV-2 et l'absence d'anticipation des pouvoirs publics face à la pandémie, invitent à reconsidérer nos politiques de santé publique sur ces sujets. Aux États-Unis, plusieurs virus respiratoires font ainsi aujourd'hui l'objet de surveillance accrue, comme le virus de la grippe, le virus respiratoire syncytial ou le SARS-CoV-2 par le Centre pour le contrôle et la prévention des maladies [Center for Disease Control and Prevention, 2023].

La lutte contre les infections virales passe évidemment par des travaux de virologie fondamentaux pour étudier les mécanismes moléculaires mis en jeu lors des différentes phases de l'infection ou identifier de potentielles cibles thérapeutiques. Néanmoins, cela ne suffit pas à appréhender l'ensemble des enjeux posés par ces pathogènes. En particulier, la modélisation a joué un rôle important dans la prise de décision des politiques sanitaires publiques dans le cadre de la pandémie [McBryde et al., 2020; Brooks-Pollock et al., 2021]; la mise en équation des processus biologiques permettant notamment d'appréhender quantitativement des dynamiques parfois difficiles à évaluer instinctivement, tels que la croissance exponentielle [Hutzler et al., 2021]. L'apport de la modélisation ne se limite pas uniquement à l'échelle inter-individu. Elle permet également d'aborder des problématiques à l'échelle de l'hôte, de l'évaluation des déterminants de la persistance du virus à la progression cancéreuse d'une infection virale. La modélisation mathématique offre donc par la diversité de ses approches, des outils précieux pour évaluer la pertinence d'une hypothèse biologique, pour proposer des nouvelles pistes d'explications ou pour prédire la dynamique d'une infection virale [Keeling and Rohani, 2008; Djidjou-Demasse et al., 2020].

Ce travail de thèse aborde, à travers l'exemple principal des infections transitoires HPV, l'importance de la diversité des approches de modélisation pour la description et la caractérisation de processus biologiques.

Sans pour autant rechercher la copie parfaite du processus biologique, un modèle mathématique en épidémiologie virale se doit d'être cohérent avec les caractéristiques du virus étudié. Dans ce chapitre introductif, nous rappellerons en premier lieu les éléments clés de la biologie des virus ainsi que les différents modes de transmission d'hôtes à hôtes. Puis, nous détaillerons les manifestations et interactions d'un virus avec l'humain. Nous concluons la partie virologie par des précisions sur le cycle de vie des deux virus au cœur de ce travail, les papillomavirus humains et le SARS-CoV-2.

Sur un versant plus technique, nous présenterons ensuite des outils classiques de modélisation déterministe, statistique et stochastique en

épidémiologie virale à l'échelle intra- et inter-hôte. Nous commencerons par rappeler les bases des modèles compartimentaux, notamment en épidémiologie, ainsi que les limites et les extensions des approches classiques. Nous évoquerons ensuite les fondements de l'analyse de survie, pour terminer par les modèles multi-états et les chaînes de Markov cachées.

Ce chapitre se termine par une partie succincte, qui récapitule les objectifs et les résultats présentés dans la suite du manuscrit.

1.1 Éléments clés de la dynamique virale

Les virus sont omniprésents sur Terre. On estime le nombre de particules virales autour de 10^{31} , soit dix fois plus que le nombre de bactéries mais uniquement 0,1 % de la biomasse terrestre [Hendrix et al., 1999; Mushegian, 2020]. Leur place dans l'arbre du vivant est régulièrement débattue [Forterre, 2010, 2016; Koonin and Starokadomskyy, 2016; Forterre and Gaïa, 2022] et s'inscrit plus globalement dans la tentative de définition du vivant, centrée en particulier sur la propension d'un individu à pouvoir assurer son métabolisme [Koonin et al., 2021].

Mis en lumière en 1892 par Dimitri Ivanoski dans des plants de tabac, c'est avec les travaux de Beijerinck en 1898 que le concept de virus, vu comme un agent infectieux capable de traverser les filtres Chamberland, émergea [Dimmock et al., 2016]. La même année, Loeffler et Frosch, étudiant le syndrome pieds-mains-bouche, montrèrent que la maladie était transmise par un agent capable de passer outre ces filtres. Cela constitue la première description d'une maladie d'origine virale chez l'humain [Brown, 2003]. Le premier virus humain formellement identifié est le virus de la fièvre jaune en 1901 [Woolhouse et al., 2012]. Enfin, Frederick Twort en 1915 [Twort, 1915] et Félix d'Hérelle en 1917 [Dimmock et al., 2016] mirent en évidence l'existence de virus bactériophages. Depuis, et en particulier avec le développement des outils de séquençage génétique, de nombreuses espèces de virus ont été identifiées chez les trois grands domaines du vivant (archées, procaryotes, eucaryotes) [Dimmock et al., 2016]. On estime aujourd'hui à plus de 10 000 le nombre d'espèces recensées [ICT], mais ce nombre est largement sous-estimé. Rien que dans le microbiote intestinal humain, on identifie plus de 140 000 génomes différents de bactériophages [Camarillo-Guerrero et al., 2021].

Les virus sont des acteurs majeurs de la dynamique des êtres vivants, tant à l'échelle microscopique qu'à l'échelle macroscopique [French and Holmes, 2020]. Dans les océans, on estime qu'ils sont responsables de la mort de 20 % de la biomasse microscopique chaque jour [Suttle, 2007]. Pour les sociétés humaines, les épidémies virales ne sont en général pas aussi meurtrières, même si des épisodes de contaminations à grandes échelles ont eu lieu dans le passé et continueront d'avoir lieu [Alizon, 2020; Ebert and Fields, 2020]. La pandémie de COVID-19 en a été un exemple récent, mais de nombreux épisodes pandémiques viraux ont éga-

lement marqué l'histoire humaine, que ce soit la grippe espagnole ou les épidémies de varioles, qui ont touché les populations amérindiennes avec la survenue des européens sur le continent Américain ou la peste Antonine qui décima plus d'un quart l'Empire Romain.

Un faisceau d'indices tend à montrer que les virus sont très anciens, possiblement contemporains des premiers organismes cellulaires dont ils dépendent aujourd'hui. Trois principales hypothèses s'affrontent sur l'origine des virus. Dans la première, les virus trouveraient leurs origines dans les mêmes entités auto-répliquatives qui ont précédé l'apparition des proto-cellules, lesquelles se sont distinguées par l'acquisition de membrane et de la synthèse protéique. Selon la seconde, les virus seraient apparus par évolution régressive (pertes de structures et de fonctions, obligeant au parasitisme) d'entités cellulaires [Forterre and Gaïa, 2022]. La troisième considère quant à elle que les virus sont des « gènes échappés » ayant acquis une capacité de réplication indépendante et de fuite de leur cellule d'origine. En outre, un quatrième scénario chimérique propose une évolution des virus à partir d'entités auto-répliquatives pré-cellulaires ayant secondairement capturé des gènes de (proto-)capside contenus dans des proto-cellules [Krupovic et al., 2019]. Malgré leurs différences, tous les scénarios s'accordent sur le fait que les principaux groupes sont apparus de façon indépendante, avant de se diversifier par co-évolution avec la vie cellulaire qu'ils infectent depuis des milliards d'années. Le monde viral est donc polyphylétique et la position de ses multiples racines dans l'arbre du vivant est encore inconnue¹.

Plusieurs classifications des virus ont été proposées depuis leur découverte. La maladie et les symptômes engendrés par l'infection ne sont pas des critères suffisamment stricts pour prédire des caractéristiques communes à des virus. En effet, chez l'humain par exemple, de nombreuses infections virales se manifestent par des syndromes grippaux (fièvre, toux, myalgie, céphalée, asthénie), bien que les mécanismes moléculaires impliqués soient très différents. Une classification des virus selon l'hôte infecté est également insuffisante car même si certains virus sont très exclusifs, d'autres sont susceptibles d'infecter toute une gamme d'hôte, voire d'évoluer et de changer d'hôte. On peut descendre au niveau de la particule virale et trier les virus selon la forme de leur capsid et la présence ou non d'enveloppe virale. Toutefois, la présence d'attributs communs n'informe en rien sur la biologie ou la pathogénie d'un virus [Dimmock et al., 2016].

Deux classifications sont aujourd'hui utilisées et reposent sur des principes différents. La classification de Baltimore se focalise sur le type d'acide nucléique viral et son mode de réplication [Baltimore, 1971]. Elle comprend sept groupes viraux. Une autre classification s'inspire de la classification du vivant et regroupe les virus selon leurs séquences génétiques et a grandement profité de l'essor des outils de séquençages génétiques. Cette classification est rendue chaque année par le comité in-

1. Pour plus de détails, voir l'ouvrage à paraître (dans un futur incertain) par Samuel et Mircea sur le COVID-19

ternational de taxonomie des virus [ICT]. Ces deux définitions ne sont pas incompatibles, la classification de l'ICTV intégrant partiellement celle de David Baltimore, et peuvent s'employer conjointement.

Dans cette première partie, nous détaillerons les caractéristiques fondamentales des virus : leur dépendance à la machinerie cellulaire de l'hôte. Nous évoquerons aussi les voies empruntées pour qu'un virus se transmette d'hôte en hôte. Ensuite, nous détaillerons plus en détails les interactions entre les virus et les hôtes aux différentes échelles intra-hôtes. Enfin, nous nous pencherons plus spécifiquement sur les deux virus étudiés lors de cette thèse – les papillomavirus humains et le coronavirus responsable du COVID-19 – avec une attention portée à leur cycle de vie et à leur épidémiologie.

1.1.1 Biologie fondamentale d'un virus

À l'inverse des cellules de leurs hôtes, les virus sont dépourvus de noyaux, d'organites et de métabolisme propre. Ils sont néanmoins composés de protéines, dont certaines peuvent être des enzymes, et d'acides nucléiques – acide désoxyribonucléique (ADN) ou acide ribonucléique (ARN) – mais ne disposent pas des ressources nécessaires pour effectuer seuls un cycle de réplication complet. En particulier, non seulement ils détournent les enzymes nécessaires à la production de protéines pour l'expression de protéines virales et les polymérases de l'hôte ou celles fabriquées par l'hôte par traduction d'un ARN messager viral pour répliquer leur génome, mais ils utilisent également les ressources nucléotidiques cellulaires pour la réplication de leur génome. Enfin, comme nous le verrons ultérieurement, certains virus utilisent la bicouche lipidique de l'hôte pour envelopper leurs particules virales [Rampersad and Tennant, 2018]. Ces caractéristiques font des virus des parasites intracellulaires [Lucas, 2010].

Le cycle de vie d'un virus alterne entre deux grandes phases : une phase extracellulaire sous forme de particules virales et une phase intracellulaire durant laquelle le virus, s'il n'est pas sous forme dormante, utilise toute ou une partie de la machinerie cellulaire de l'hôte à son profit afin de produire des particules virales. Celles-ci sont ensuite relarguées pour infecter de nouvelles cellules ou de nouveaux hôtes, recommençant ainsi le cycle.

Le milieu extra-cellulaire est hostile aux supports de l'information génétique que sont l'ADN ou l'ARN. Ainsi, les virus ont évolué des manières d'acheminer l'ensemble des composants nécessaires à l'initiation d'une nouvelle infection d'une cellule infectée vers une cellule saine. On nomme « capsid » le manteau protéique qui entoure une particule virale. Ce manteau est constitué de la répétition de sous-unités protéiques formant des structures symétriques complexes (par exemple hélicoïdales ou icosaédrales). Le rôle de ces sous-unités est varié : reconnaissance et emballage des composés viraux avant dissémination, protection du milieu

extérieur, identification et liaison à l'hôte et enfin libération des composés viraux une fois entrée dans la cellule. Pour éviter un relargage inopiné du génome viral, la libération de la capsidie nécessite des signaux précis, souvent liés à des changements de conformation résultant d'interactions avec certaines protéines de la membrane plasmique ou du cytoplasme.

Autour de leur nucléocapsidie, certains virus présentent également une bicouche lipidique nommée enveloppe virale (les poxviridés en ont même deux). Cette enveloppe est dérivée de la membrane plasmique de l'hôte et généralement acquise lorsque les nucléocapsides sont émises dans le milieu extra-cellulaire. Mais elle peut aussi avoir d'autres origines (appareil de Golgi, reticulum endoplasmique voir enveloppe nucléaire). Ces enveloppes n'offrent pas forcément d'avantage en termes de survie dans le milieu extérieur car elles sont sensibles à la dessiccation ; même s'il a été montré que certaines particules virales enveloppées pouvaient persister pendant plusieurs heures voir jours dans le milieu extérieur. De manière générale, les structures lipidiques sont moins stables que celles protéiques. Par contre, elles offrent une protection efficace contre le système immunitaire, en particulier les anticorps, et facilitent la pénétration du virus dans sa cellule hôte [Lucas, 2010].

Une fois dans la cellule, le virus va détourner les ressources de l'hôte à son profit pour produire de nouveaux virions. Qu'il soit d'origine virale ou cellulaire, l'ADN est répliqué de la même manière par l'ADN polymérase présent dans le noyau de l'hôte. Pour les virus à ARN, les choses sont différentes car l'hôte ne dispose pas d'enzymes susceptibles de dupliquer les brins d'ARN. On distingue alors les virus à polarité négative des virus à polarité positive ou des virus à ARN double brin. De même que l'ADN polymérase a un unique sens de lecture de l'ADN, les ribosomes ont également un unique sens de lecture des ARNm et on nomme virus à polarité positive les virus dont les gènes sont codés dans le même sens de lecture que les ribosomes. Une fois dans la cellule, ces ARN viraux vont pouvoir être traduits par les ribosomes directement. Ceci conduit à la production de l'ARN-polymérase qui permettra l'amplification du matériel génétique du virus. À l'inverse, chez les virus à polarité négative, les gènes sont codés dans le sens de lecture inverse des ribosomes. Ces ARN ne peuvent donc pas être traduits directement par l'hôte (et il en va de même pour les virus à ARN double brin). Dans ces circonstances, le virus doit apporter lors de l'entrée cellulaire des protéines virales, notamment des ARN-polymérases, permettant la transcription de leur matériel génétique en ARNm, qui pourront ensuite être traduits par les ribosomes. Les rétrovirus constituent un cas à part parmi les virus ARN. Lors de l'entrée cellulaire, les particules virales libèrent également une transcriptase inverse, qui permet de former un double brin d'ADN à partir d'un ARN simple brin. Cet ADN sera ensuite transcrit dans le noyau cellulaire. Hormis les virus à ARN à polarité positive, pour qui une seule copie d'ARN suffit à déclencher une infection, les autres virus à ARN sont dépendants de protéines pour initier l'infection. Ainsi, les particules virales

des virus à ARN à polarité négative ou ambisens déprotéinées sont non infectieuses.

Les enzymes impliquées dans la réplication du matériel génétique ont une importance évolutive majeure. L'ADN polymérase est une enzyme qui réplique très fidèlement l'ADN grâce à une exonucléase qui permet de corriger des erreurs d'appariement. On la retrouve dans l'ensemble du vivant [Kornberg and Baker, 2005]. À l'inverse, les ARN polymérases et les transcriptases inverses sont dénuées de cette fonction de relecture et correction des mésappariements. Lors de la réplication, on estime qu'une erreur a lieu toutes les $10^9 - 10^{10}$ paires de bases contre une erreur toutes les $10^3 - 10^4$ paires de bases lors de la réplication de l'ARN. On observe donc en général que les virus à ARN, comme le SARS-CoV-2, mutent plus vite que les virus à ADN, comme les papillomavirus. L'émergence de variants est donc plus fréquente chez les virus à ARN, compliquant la mise en place de politique préventive [Zeng et al., 2022].

1.1.2 Des voies de transmission variées

La plupart des virus ont une durée de vie limitée hors de leur(s) hôte(s). Néanmoins, à terme, le virus sera libéré dans le milieu extérieur, que ce soit par expulsion de l'hôte (e.g. desquamation des cellules du col de l'utérus infectées par HPV [Doorbar, 2006] ou expulsion dans des aérosols pour le SARS-CoV-2 [Wang et al., 2021; Tellier, 2022]) ou lors de la mort de l'hôte. Ainsi, il est crucial pour la survie d'un virus d'infecter un nouvel hôte.

Chez les animaux, la peau constitue une première barrière contre ces transmissions. Les voies principales de transmission pour de nombreux virus sont les voies oro-fécales et les voies respiratoires. En effet, ces deux routes contiennent des zones d'échanges privilégiées entre l'hôte et le milieu extérieur, les épithélia des voies respiratoires et des voies digestives. Ces surfaces ont évolué pour maximiser les échanges entre les deux milieux. Néanmoins, la transmission virale par contact direct, souvent cutané, reste possible (comme pour les Poxviridae).

On parle de transmission horizontale lorsque qu'un individu infecté transmet à un autre individu sain de la même espèce des particules virales, sans implication de voie spécifique à la reproduction (par exemple transmission *in utero* ou par allaitement). Même si le tropisme d'un virus n'est pas forcément lié à ses voies de transmissions privilégiées, les voies respiratoires constituent l'une des principales voies de transmission du virus. Les virus respiratoires se transmettent généralement via des gouttelettes produites naturellement lorsque l'on parle, tousse ou éternue [Wang et al., 2021]. On parle alors de transmission par aérosols. Ces aérosols peuvent être inhalés par une personne saine directement ou, dans certains, cas se déposer sur des surfaces inertes² et être ensuite

2. On parle également de fomites pour désigner une surface inerte contaminée par un pathogène

portés vers les voies respiratoires via les doigts ou les yeux. Ce mode de transmission est délicat à modéliser car il fait intervenir de multiples phénomènes physiques et biophysiques, par exemple taille et composition des gouttelettes, température, hygrométrie, mouvement convectif de l'air ambiant, stabilité des virions. Non seulement ces facteurs peuvent varier en fonction du lieu de transmission, en particulier la température et l'hygrométrie, mais ces changements de conditions physiques du milieu peuvent également entraîner des modifications des pratiques sociales. Par exemple, en hiver les gens se rassemblent plus qu'en été dans des zones souvent peu ou pas ventilées [Wang et al., 2021]. De plus, l'infection elle-même peut influencer cette transmission. En particulier, la production de mucus peut être renforcée par l'infection, par exemple par l'intermédiaire de réponse inflammatoire [Cohn et al., 1997].

L'autre route de transmission majeure est la voie oro-fécale. Les virus empruntant cette voie infectent et se multiplient dans le tractus digestif, en particulier dans la zone de l'intestin grêle. Ces virus sont ensuite excrétés en très grand nombre dans les fèces. Si la salubrité de l'habitation n'est pas satisfaisante, il peut arriver que de l'eau contaminée entre en contact avec de la nourriture ou soit directement ingérée par d'autres individus. Les politiques publiques pour lutter contre les transmissions oro-fécales s'appuient majoritairement sur la mise en place d'installations sanitaires efficaces et des travaux d'assainissement. De nombreux médecins et scientifiques ont contribué à la mise en place de politique hygiéniste dans la santé publique. On peut ainsi citer John Snow sur le choléra [Snow, 1856, 1985], Florence Nightingale sur le métier d'infirmier(ère) et l'organisation de l'hôpital [Nightingale, 1858, 1862], ou encore Ignaz Semmelweis et Louis Pasteur sur l'importance du lavage de mains en hôpital [Shorter, 1984; Vermeil et al., 2019].

À l'inverse des deux voies évoquées précédemment, la transmission par voie sexuelle n'est l'apanage que de quelques virus. Chez l'humain, on étudie principalement le virus de l'immunodéficience humaine (VIH), certains génotypes de papillomavirus (notamment les alphapapillomavirus), des virus hépatiques ou encore des herpesvirus. Pour certains de ces virus, le risque de transmission est différent selon le sexe de la personne infectée. Par exemple, une femme porteuse du VIH-1 a 2 à 3 fois moins de risque de transmettre le virus qu'un homme infecté [European Study Group on Heterosexual Transmission of HIV, 1992]. La probabilité de transmission par contact diffère également selon le virus considéré. Ainsi, ce risque est faible pour le VIH, en particulier avec utilisation de préservatifs [Patel et al., 2014], mais beaucoup plus élevé pour les HPV [Winer et al., 2008], même avec préservatif [Winer et al., 2006; Stanley, 2010].

Parmi les transmissions inter-humaines, on distingue la transmission verticale mère-fœtus/nouveau-né de la transmission horizontale. Même si des virus des microbiotes sont transmis de manière verticale [Maqsood et al., 2019], on ne connaît pour l'instant aucun virus d'intérêt sanitaire se

transmettant exclusivement par cette voie. Les virus se transmettant verticalement infectent la mère par des voies horizontales. Ces virus peuvent soit traverser la barrière placentaire pour atteindre le fœtus, soit infecter le nouveau-né pendant ou après l'accouchement. Certaines de ces infections peuvent être bénignes chez la mère mais mettre en péril la survie du nourrisson ou engendrer de graves complications. La vaccination des mères reste le moyen le plus efficace pour réduire le risque de transmission verticale [Etti et al., 2022].

Les zoonoses constituent le dernier axe majeur de transmission vers l'humain. On parle de zoonose lorsqu'un animal contamine un humain par voie directe, comme dans le cas de la rage via une morsure, ou par l'intermédiaire de vecteurs (souvent des arthropodes). Par exemple, dans le cas de la fièvre jaune, les moustiques deviennent porteurs du virus en piquant des singes infectés et transmettent le virus à l'homme lorsque les femelles moustiques viennent prélever leur repas sanguin. Le cas de la pandémie de COVID-19, d'origine zoonotique probable [Andersen et al., 2020], a mis en lumière les enjeux autour de ces sauts d'hôtes. Selon certaines études, l'émergence de nouvelles zoonoses risque de s'accroître avec les pressions humaines dont le dérèglement climatique [Mora et al., 2022].

1.1.3 Interactions hôte-virus

Manifestations d'une infection virale

À l'échelle de l'individu, les symptômes d'une infection virale sont variés et dépendent de l'hôte. Chez l'humain, les signes généralistes d'une infection virale incluent notamment fièvre, courbatures, douleurs, fatigue et diarrhée. Ces symptômes sont généralement peu discriminants et coïncident avec l'activation d'une réponse immunitaire innée. D'autres symptômes peuvent aider à préciser la localisation de l'infection (par exemple vomissement pour des infections du tract oro-fécale, éternuement et rhinorrhée pour des infections de voies respiratoires) mais ces critères sont généralement non excluants, plusieurs virus présentant la même étiologie.

De nombreuses techniques sont employées pour détecter la présence d'un virus à partir d'échantillons biologiques. Si la concentration en virus est suffisamment importante, il est possible d'observer directement les particules virales au microscope électronique. En pratique, cela est souvent trop coûteux. On peut également utiliser un microscope par fluorescence pour repérer les cellules infectées ou les particules virales à l'aide d'anticorps spécifiques au virus (aussi appelé immunoglobulines) fluorescents. Des méthodes quantitatives sont également employées pour mesurer la charge virale ou le dosage de certaines protéines d'intérêt. Par exemple, la méthode ELISA est un procédé permettant le dosage de certains antigènes/anticorps que l'on utilise pour tester la présence d'un virus dans un échantillon. Les méthodes d'amplification en chaîne par polymérase (PCR) permettent quant à elles de détecter du matériel

génétique viral (ADN, ou ARN en rajoutant une étape de transcription inverse), et ce jusqu'à des concentrations très faibles. Enfin, on peut procéder à des tests sérologiques pour regarder plus finement la réponse immunitaire de l'hôte à une infection virale, en particulier les anticorps spécifiques présents dans le sérum. Dans ce dernier cas, on ne peut cependant obtenir de l'information que sur une infection passée.

Ces approches sont complémentaires et l'information apportée par les différentes méthodes permet d'évaluer le statut d'une infection virale.

Virus et système immunitaire

Chez l'humain, il existe une collection étendue de cellules diverses interagissant entre elles avec le reste de l'organisme et capables de protéger l'organisme contre les entités extérieures : le système immunitaire. On distingue le système immunitaire inné, incluant plusieurs cellules et composants généralistes, qui constituent la première ligne de défense contre les agents extérieurs, du système immunitaire acquis, qui est activé par les lymphocytes T et B en réponse à un pathogène et donc spécifique. À l'inverse des acteurs de l'immunité innée, présents en quantités a priori constantes dans l'organisme, les cellules de l'immunité acquise sont présentes en moindre quantité avant la première infection et la mise en place d'une réponse acquise peut prendre plusieurs jours, temps durant lequel le virus peut proliférer (appelé phase d'éclipse pour le VIH). L'activation du système immunitaire acquis permet le recrutement de lymphocytes mémoires spécifiques, conférant ainsi une protection sur le long terme contre des infections par le même pathogène. Cette immunité acquise se traduit également par la présence dans le sang d'anticorps ciblant spécifiquement des protéines ou acides nucléiques viraux. Les tests sérologiques cherchent à quantifier la présence de ces anticorps. L'action des anticorps est variée. Certains neutralisent directement les particules virales, d'autres, en se fixant sur le virus, induisent indirectement sa destruction par phagocytose ou par recrutement d'anticorps complémentaires neutralisants. Ces deux mécanismes collaborent étroitement dans l'élimination de pathogènes [Clark and Kupper, 2005] et le contrôle du microbiote humain [Slack et al., 2009].

L'immunité varie au cours de la vie d'un individu. Les nouveau-nés et les personnes âgées sont des populations qui ont souvent une immunité plus faible et constituent donc des populations à risques [Simon et al., 2015; Bajaj et al., 2021].

On attribue généralement les dommages causés par une infection virale au virus lui-même, mais il peut arriver que les complications soient le fruit du système immunitaire. C'est notamment le cas des infections SARS-CoV-2 qui peuvent induire des emballements du système immunitaire (orage de cytokines) pouvant entraîner des défaillances de certains organes [Song et al., 2020; Montazersaheb et al., 2022]. L'immunité d'un hôte est un trait d'histoire de vie soumis à de fortes pressions de sélection

évolutives et dont les caractéristiques résultent de compromis adaptatifs. Un système immunitaire peu efficace peut ne pas suffire à contenir un pathogène conduisant à la mort de l'hôte. À l'inverse, un système faisant preuve de trop de zèle dans l'élimination d'un pathogène peut parvenir à son éradication de l'hôte, au prix de dommages collatéraux potentiellement mortels [Graham et al., 2005].

Définitions cliniques des infections virales

À l'échelle de l'hôte, une infection virale peut se manifester de multiples manières qu'on classe généralement en sept grandes catégories : infection aiguë, asymptomatique, persistante, chronique, latente, à progression lente, tumorigène. Quatre indicateurs sont classiquement utilisés pour définir ces stades : i) la production de descendants infectieux, ii) l'infection cause la mort de la cellule hôte, iii) les symptômes et iv) la durée de l'infection [Dimmock et al., 2016].

En pratique, un continuum de manifestations existe, ces catégories servant surtout de guide pour les personnels médicaux/chercheurs [Virgin et al., 2009]. Au cours de l'histoire naturelle, une infection passe fréquemment par plusieurs stades et la progression d'un stade vers l'autre est utilisée en santé publique pour la classification du risque de complications d'une infection virale. Ainsi, une infection aiguë par HPV présente peu de risque pour la personne infectée, alors qu'une infection persistante fera l'objet d'une surveillance plus approfondie pour prévenir le passage vers le stade tumorigène. Néanmoins, l'arbitrage entre aiguë et chronique est généralement arbitraire (par exemple la détection sur une durée consécutive de plus de six mois pour HPV) et ne traduit pas forcément de réalité biologique derrière.

Coévolution hôte-virus

On l'a vu, les virus sont des entités parasites intra-cellulaires. Un concept central de l'étude des relations hôtes-virus est la notion de virulence, qui peut être définie comme les dommages engendrés par le virus à l'hôte, que ce soit de manière directe ou indirecte [Alizon et al., 2009]. Ces dommages se mesurent habituellement sur la *fitness* de l'hôte. En pratique, cette dernière n'est pas toujours facile à estimer, c'est pourquoi on a parfois recours à des proxys. Dans le cas de la pandémie de COVID-19, on peut ainsi distinguer la létalité (probabilité de décès de l'hôte dans un contexte donné) de la mortalité (mesure à l'échelle de la population) de la pathogénicité (qui se mesure au niveau cellulaire) [Alizon and Sofonea, 2021].

À l'exception de certains parasites comme Ebola [Rewar and Mirdha, 2014], la mort de l'hôte est généralement considérée comme un « cul-de-sac » pour le pathogène car il ne peut alors plus se multiplier ni se transmettre. Cela a conduit certains à postuler une théorie de l'avirulence selon laquelle plus la relation entre le parasite et son hôte était ancienne,

plus l'infection est bénigne [Méthot, 2012]. Cette théorie simpliste se heurte malheureusement à divers problèmes empiriques, en particulier les cas de la tuberculose [Barberis et al., 2017], du paludisme [Cox, 2010; Drugs et al., 2004] ou de la variole [Riedel, 2005] attestés depuis plusieurs milliers d'années et dont l'infection est toujours virulente. Même dans une perspective co-évolutive, van Baalen [1998] montre que l'avirulence n'est qu'un des états finaux potentiels à long terme de l'interaction hôte-parasite. Les deux agents peuvent également entrer dans une course aux armements [Valen, 2014; Dawkins et al., 1997], qui se manifeste par une forte virulence du virus. Dans une perspective purement évolutive (donc sans évolution de l'hôte), la virulence d'un pathogène peut être maintenue par plusieurs phénomènes [Alizon and Michalakakis, 2015]. Par exemple, elle peut persister si elle agit comme un trait neutre, notamment si l'hôte est un « cul-de-sac » pour le parasite (comme la grande douve du foie pour qui l'homme est un hôte accidentel [Dalton, 1999]). Dans la même ligne, s'il n'existe pas de variation génétique associée au trait, la virulence ne pourra pas évoluer. Enfin, si la virulence est corrélée positivement à un trait lui-même corrélée positivement avec la *fitness* du parasite, il est peu probable que les interactions hôte-pathogène tendent vers l'avirulence. Ce dernier cas a été particulièrement étudié sous le prisme du compromis entre virulence et transmission [Alizon et al., 2009], selon lequel une transmission accrue du pathogène est contrebalancée par une durée infectieuse plus courte (la durée infectieuse peut être vue comme inversement proportionnelle à la virulence).

1.2 Histoire naturelle et épidémiologie des virus étudiés

1.2.1 HPV

La famille des papillomavirus (PV) comprend de nombreux virus, infectant l'humain et d'autres métazoaires (mammifères, oiseaux, reptiles). L'histoire évolutive des *Papillomaviridae* est riche comme l'illustre le fait que les papillomavirus humains ne sont pas monophylétiques (c'est-à-dire ne partageant pas un seul ancêtre commun) [Bravo and Félez-Sánchez, 2015]. On estime l'origine des premiers papillomavirus autour de 350 millions d'années, concomitantes des changements dans l'épithélium des hôtes ancestraux. Depuis, les différents papillomavirus ont co-évolué avec leurs hôtes et on détecte peu de sauts entre groupes taxonomiques [Van Doorslaer, 2013; Bravo and Félez-Sánchez, 2015].

On recense aujourd'hui plus de 200 génotypes HPV [International HPV reference center], certains étant plus étudiés que d'autres en raison de leur importance en termes de santé publique. Le Centre International de la Recherche sur le Cancer (CIRC ou IARC en anglais) distingue les génotypes dits à bas risque (LR-HPV), qui n'entraînent que rarement de

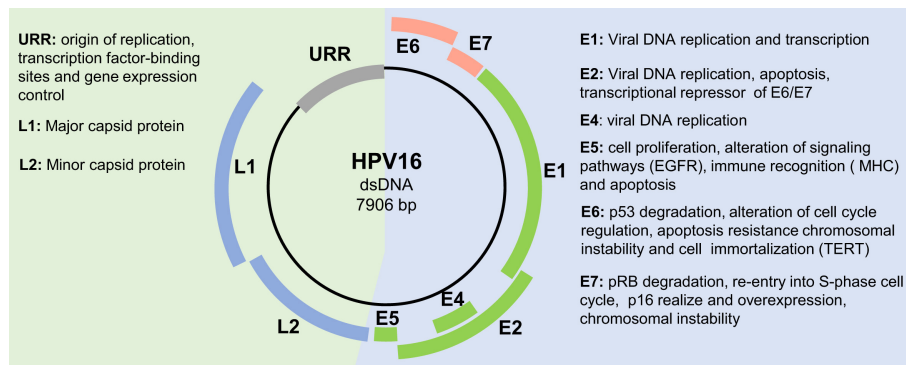


FIGURE 1.1 – Génome de HPV16 et organisation des cadres de lecture ouverte et des protéines qu’ils codent. Tous les crédits pour la production de l’image reviennent à de Sanjosé et al. [2018].

néoplasies, des génotypes dits à haut risque (HR-HPV³), qui sont impliqués dans la survenue de divers cancers [IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2012]. Toutefois, certains génotypes à bas risques sont aussi un enjeu de santé publique par leur implication dans la survenue de verrues, dont le coût de la prise en charge par les organismes publics représente de 13 % à 18 % des dépenses liées aux infections HPV [Herse and Reissell, 2011].

D’un point de vue biologique, les papillomavirus présentent une particule virale à capside icosaédrale non enveloppée d’une cinquantaine de nanomètres. Ce sont des petits virus à ADN dont le génome comprend 8 000 paires de bases. Leur taux de mutations est estimé à environ une mutation tous les 10^7 à 10^8 nucléotides par an [Nelson and Mirabello, 2023], taux comparable aux autres virus à ADN [Peck and Lauring, 2018]. Ce génome est organisé sous une forme circulaire à deux brins, que l’on nomme également épisome. On a identifié 8 à 9 cadres de lecture ouverte (ORF) selon le genre considéré, codant des protéines exprimées tardivement (*late*) L1 et L2, et des protéines exprimées plus précocement (*early*) E1, E2, E4, E5⁴, E6, E7. Les ORFs L1, L2 et E1, E2 sont généralement bien conservés entre les génotypes et sont respectivement impliqués dans l’encapsidation et dans la réplication du génome viral [Doorbar et al., 2015]. Pour les autres ORFs, on observe une plus grande diversité. Ils sont impliqués dans l’entrée cellulaire, l’évasion immunitaire ou encore la libération des virions. En particulier, le rôle des protéines E6 et E7 a été mis en avant dans la progression cancéreuse [Doorbar et al., 2012; Tomaić, 2016; de Sanjosé et al., 2018]. Pour plus de détails sur le génome HPV, on peut se référer à la Figure 1.1.

Les papillomavirus humains (HPV) infectent divers épithéliums muqueux (génitaux ou oraux) et cutanés (dont les manifestations les plus courantes sont les verrues). De par leur importance en termes de santé pu-

3. HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66

4. Absent chez les betapapillomavirus

blique, les alphapapillomavirus ont été plus étudiés que les autres genres HPV. En effet, ils sont impliqués dans la survenue de nombreux cancers, dont presque tous ceux du col de l'utérus [Doorbar et al., 2012]. Nous nous focaliserons par la suite sur le cycle de vie des alphapapillomavirus dans le cadre des infections génitales. Il est généralement admis comme indispensable pour l'infection que le virion atteigne les couches basales de l'épithélium. Cela serait facilité par la présence de micro-lésions dans le tissu épithélial [Stubenrauch and Laimins, 1999]. L'entrée du génome viral dans la cellule est complexe et fait intervenir de nombreux intermédiaires. La protéine L1, sur la capsid virale, s'attache à la membrane basale ou à la surface des cellules des couches basales par l'intermédiaire d'un polysaccharide (le sulfate d'héparane). Cela induit des changements dans la conformation de protéines L2 à la surface du virion [Kines et al., 2009]. Le clivage de la zone N-terminale de la protéine F2 va permettre au virus de s'attacher à une seconde protéine de la membrane plasmique des cellules basales, variable selon les génotypes HPV et les cellules infectées. L'environnement de la lame basale jouerait également un rôle important dans le changement conformationnel des protéines de capsides, changement nécessaire à l'entrée du virus dans la cellule [Kines et al., 2009]. Le virus est ensuite internalisé par endocytose, puis dirigé vers le noyau cellulaire [Graham, 2017]. Il a été suggéré que les réponses de cicatrifications de ces micro-lésions pourraient favoriser la propagation de cellules infectées dans le tissu [Doorbar et al., 2012]. En effet, il est souvent postulé pour les génotypes à haut risque qu'une étape de division cellulaire est nécessaire à l'entrée du génome viral dans le noyau [Pyeon et al., 2009]. Cette entrée est suivie par une phase d'amplification du génome virale [Doorbar et al., 2015]. Toutefois, le génome HPV reste maintenu à des niveaux bas dans les cellules basales. On estime ce chiffre entre 50 et 100 [Moody, 2017], avec une limite maximale estimée à 200 copies par cellule [Doorbar, 2006]. En pratique ce chiffre est assez mal connu et est issu d'estimations pour d'autres papillomavirus non-humains. Ces premières phases de réplifications virales font intervenir les protéines E1 et E2 [Graham, 2017].

Les PV sont des virus non-lytiques qui suivent le devenir des cellules qu'ils infectent. Dans les cellules souches, le virus a une activité métabolique limitée aux premières phases de répllication et de maintenance du génome viral [Doorbar, 2006; Graham, 2017]. En migrant vers la partie apicale du tissu dans les cellules parabasales, on observe une hausse progressive de l'activité métabolique du virus et il commence à se répliquer en grand nombre dans les cellules en cours de maturation. L'expression des protéines L1 et L2 intervient tardivement, lors de l'assemblage des particules virales et de la libération des virions [Doorbar, 2006]. Finalement, la desquamation des cellules hôtes à la surface permet la libération des particules virales dans le milieu extérieur [Doorbar et al., 2012].

Les HPV sont très répandus sur tous les continents et constituent une infection sexuellement transmissible très courante. Les suivis trans-

versaux estiment une prévalence autour des 20 % chez les femmes de moins de 25 ans. Cette prévalence a tendance à baisser avec l'âge pour se stabiliser vers une prévalence autour des 10 % pour les femmes de plus de 55 ans [Serrano et al., 2018]. Le pic de prévalence concorde avec le début des rapports sexuels, puis diminue sensiblement avec l'âge [Schiffman and Solomon, 2013]. On estime en particulier que pratiquement l'ensemble des êtres humains sera exposé à au moins un génotype génital au cours de sa vie. Des modèles suggèrent que 80 % des adultes de plus de 45 ans ont été ou sont infectés par au moins un génotype génital [Chesson et al., 2014a].

Néanmoins, ces infections sont généralement asymptomatiques, bien que pouvant se manifester sous forme de papillomes ou de verrues. Le virus n'étant pas lytique, il n'induit généralement pas d'inflammation [de Sanjosé et al., 2018]. De plus, on ne parvient souvent plus à détecter la présence du virus au bout de 3 ans [Insinga et al., 2007; Schiffman et al., 2011a; Ramanakumar et al., 2016] après le début de l'infection. On nomme généralement cette perte de détection « clairance » de l'infection. Cette clairance a généralement été attribuée à l'élimination du virus mais des travaux récents suggèrent que le virus pourrait cependant persister à des niveaux peu voir pas détectables [Gravitt, 2012; Gravitt and Winer, 2017; Doorbar, 2023]. Il est classiquement admis que le système immunitaire est impliqué dans le passage d'une infection productive vers un stade latent. Cette phase latente est cependant réversible et des perturbations extérieures ou des changements immunitaires peuvent induire un retour à des phases productives [Gravitt and Winer, 2017].

Il peut arriver que ces infections persistent et induisent des stades cancéreux dans le tissu infecté. Il a ainsi été démontré que la présence du virus est nécessaire à la survenue des cancers du col de l'utérus [Bo-shart et al., 1984; zur Hausen, 1996; Bosch et al., 2002]. Une zone a été identifiée comme étant particulièrement à risque de carcinogénèse au niveau de la jonction entre l'exocol et l'endocol [Herfs et al., 2012]. Plus spécifiquement, des cellules d'une population bien définie, les cellules cuboïdes, ont été retrouvées dans de nombreux cancers et sont suspectées d'être associées à la survenue de cancers du col de l'utérus [Herfs and Crum, 2015]. Ces différences de manifestations cliniques entre les génotypes à bas risques (verrues) et les génotypes à haut risque (néoplasie) semblent refléter des stratégies différentes de transmission et propagation dans l'épithélium ainsi que des interactions variables avec le système immunitaire de l'hôte [Doorbar, 2005].

À ce titre, les HPV sont les virus les plus oncogènes et parmi les plus meurtriers, en particulier chez les femmes où 8.6 % (contre 0.8 % chez les hommes) des cancers leur sont attribuables, pour un fardeau estimé à plus de 650 000 nouveaux cas diagnostiqués chaque année [Plummer et al., 2016b; de Martel et al., 2017], dont une grande majorité (plus de 500 000) de cancers du col de l'utérus [de Martel et al., 2017]. Même parmi les génotypes dits à haut risque, tous n'ont pas la même propension à

causer la survenue de cancers. Près de 70 % des cancers du col de l'utérus sont attribués aux génotypes 16 et 18 [Sanjose et al., 2010] ce qui explique l'attention prioritaire qui leur est portée.

Ce chiffre de 650 000 nouveaux cancers diagnostiqués par an dans le monde attribuable aux HPV cache de fortes disparités. Par exemple, 87 % des femmes diagnostiquées du cancer de l'utérus ont été recensées dans des pays en développement [de Martel et al., 2017]. Ce chiffre témoigne des inégalités en termes de moyens alloués à la prévention des infections HPV et aux suivis gynécologiques réguliers [Bruni et al., 2022].

En 2020, l'Organisation Mondiale de la Santé (OMS) lance un programme ambitieux d'élimination (à échelle de 4 cas pour 100 000 femmes) des cancers du col de l'utérus dans le monde. Pour 2030, il vise à vacciner entièrement (au moins deux doses) 90 % des filles de moins de 15 ans, dépister 70 % des femmes à 35 et 45 ans, traiter 90 % des femmes atteintes de lésions pré-cancéreuses et prendre en charge 90 % des femmes atteintes de ce cancer [Ghebreyesus and Jakab, 2020; Wilailak et al., 2021]. D'après les projections mathématiques de l'OMS, atteindre ces objectifs 90-70-90, en particulier dans les pays à ressources limitées, permettrait de réduire de 97 % l'incidence moyenne des cancers du col de l'utérus d'ici 2120. Néanmoins, cet objectif est encore loin d'être atteint car même pour des pays aux ressources importantes, comme la France ou le Japon, le taux de vaccination reste faible [Bruni et al., 2022].

Le dépistage régulier des femmes à partir de 25 ans par l'identification des lésions pré-cancéreuses (par cytologie liquide par exemple [Denton, 2007]) et/ou du génotype HPV permet également de détecter précocement les premiers stades de progression cancéreuse. Si pris à temps, il est possible de détruire ou d'enlever chirurgicalement ces lésions pré-cancéreuses pour éviter des complications plus sévères [Martin-Hirsch et al., 2010]. Toutefois, la mise en place de ces campagnes de dépistage nécessite des moyens humains et matériels importants [Wilailak et al., 2021]. Sur le long terme, la vaccination reste le moyen le plus efficace pour prévenir et éliminer, ou du moins réduire la prévalence à des niveaux faibles, les cancers du col de l'utérus [Schiffman et al., 2005; Wilailak et al., 2021].

Une des raisons qui fait que la vaccination reste le meilleur moyen préventif contre les infections par HPV [Markowitz et al., 2018] est que même les préservatifs n'offrent pas une protection complète contre les transmissions horizontales. Il a ainsi été estimé qu'une utilisation régulière et correcte des préservatifs permettait d'obtenir environ 60 % de protection contre les infections HPV [Winer et al., 2006]. Bien que la voie principale de transmission HPV soit par pénétration, d'autres voies de transmission existent : par exemple des contacts génitaux sans pénétration, contacts main/bouche - parties génitales [Liu et al., 2016]. Il a également été montré que les particules résistaient bien à la dessiccation et pouvaient persister sur des surfaces inertes, notamment du matériel médical, jusqu'à une voire deux semaines [Roden et al., 1997; Ding et al.,

2011]. Néanmoins, des incertitudes demeurent sur le potentiel infectieux des copies virales retrouvées sur les surfaces inertes car il n'existe pas à l'heure actuelle de modèles *in vitro* pour tester leur infectivité [Ma et al., 2013].

Plusieurs vaccins sont aujourd'hui commercialisés. Depuis 2006, un vaccin tétravalent commercialisé par Merck & Co., le Gardasil, cible les deux génotypes les plus impliqués dans la survenue de cancers (HPV16 et HPV18) [Joura et al., 2015] ainsi les deux génotypes à bas risques (HPV6 et HPV11) les plus impliqués dans la survenue de verrues génitales [Aubin et al., 2008; Chan et al., 2009]. L'année suivante, un vaccin bivalent ciblant HPV16 et HPV18 a également été commercialisé par GSK. Depuis 2015, une version nonavalente du Gardasil est commercialisée. Elle cible en plus les génotypes 31, 33, 45, 52 et 58 [Joura et al., 2015]. C'est ce dernier qui est recommandé par la Haute Autorité de Santé [2020] pour toute nouvelle vaccination.

1.2.2 SARS-CoV-2

Le coronavirus 2 du syndrome respiratoire aigu sévère (SARS-CoV-2) est un virus à ARN simple brin à polarité positive, enveloppé. On estime la taille de son génome à 30 000 paires de bases. Il est organisé en 14 ORFs qui codent 29 protéines virales dont notamment la protéine spike (S), cible privilégiée pour l'élaboration des vaccins [Kyriakidis et al., 2021]. Le virus entre dans la cellule hôte via des interactions entre la protéine S et des protéines de la membrane cellulaire, les enzymes de conversion de l'angiotensine 2. S'ensuit l'entrée du matériel génétique viral dans le cytoplasme qui est directement traduit par les ribosomes de l'hôte. Les premières protéines virales synthétisées forment des complexes de transcription et réplication du génome viral. Les ARNm nouvellement transcrits servent de modèles pour la traduction des protéines virales. Les nouvelles particules virales sont assemblées dans le réticulum endoplasmique et l'appareil de Golgi puis larguées par exocytose hors de la cellule [Yang and Rao, 2021].

La transmission d'hôte à hôte s'effectue majoritairement par aérosols, en particulier dans les particules les plus fines. Ces particules aéroportées sont expulsées de l'hôte infecté, notamment lors d'éternuement ou d'épisodes de toux, mais également dès lors que la personne parle, chante ou respire. Bien que les particules virales ne soient présentes qu'en concentration limitée dans l'air ambiant, de l'ordre de $2.42 \times 10^3 - 9.4 \times 10^4$ copies par m^3 d'air selon les protocoles, le volume d'air respiré par une personne moyenne favorise grandement les infections. De plus, ces concentrations dissimulent de grandes variations. Selon l'organisation de la salle, la ventilation de la pièce, la présence de personnes infectées dans les environs, le risque d'infection varie grandement [Tellier, 2022].

Les coronavirus sont des pathogènes généralistes qui infectent de nombreux mammifères. En particulier, les chauves-souris sont un réservoir

naturel de nombreuses espèces de coronavirus [Forni et al., 2017]. On a également détecté la présence de coronavirus chez divers animaux domestiques, notamment les chameaux [de Wit et al., 2016; Forni et al., 2017], et sauvages [Zhang et al., 2020].

Deux espèces de coronavirus humains ont été découvertes vers la fin des années 1960, HCoV-229E et HCoV-OC43. Avec les espèces HCoV-NL63 et HCoV-HKU1 découvertes dans les années 2000, elles sont responsables de 10 à 30 % des infections des voies respiratoires supérieures chaque année. Ces infections sont caractérisées par des symptômes respiratoires légers, semblables à un rhume [Yang and Rao, 2021]. Comparativement, le SARS-CoV-2 peut entraîner des séquelles sévères chez l'humain (les symptômes sont généralement plus graves chez les hommes que chez les femmes), en particulier chez les personnes âgées [O'Driscoll et al., 2021]. Deux autres espèces de coronavirus susceptibles de causer des complications sévères avaient été identifiées par le passé.

En novembre 2002, le premier cas du syndrome respiratoire aigu sévère (SARS) est identifié dans la région du Guandong en Chine. Rapidement, de nouvelles personnes sont infectées et le virus se propage dans le pays et dans plusieurs autres pays [de Wit et al., 2016]. En mars 2003, un coronavirus, le SARS-CoV, est identifié comme responsable de la pandémie [Zhong et al., 2003; Ksiazek et al., 2003]. On reporte plus de 8 000 cas et 774 décès entre novembre 2002 et juillet 2003 [de Wit et al., 2016]. Dix ans plus tard, un nouveau coronavirus est identifié au Moyen-Orient (MERS-CoV pour *Middle-East Respiratory Syndrome-related Coronavirus*) et se propage dans et hors de la péninsule Arabique [Zaki et al., 2012]. Entre 2010 et 2020, on note 2 500 infections et 800 décès [Yang and Rao, 2021]. En décembre 2019, un nouveau coronavirus est identifié dans la région du Wuhan en Chine impliqué dans des cas de pneumonie [Wu et al., 2020; Zhou et al., 2020]. Rapidement, le virus s'est propagé dans le pays et dans le monde. Dès février 2020, on reporte la présence du virus dans au moins 26 pays [Mahase, 2020]. La majorité des personnes infectées présente des symptômes légers et guérissent de la maladie. On estime un taux de létalité (proportion de décès par rapport au nombre total de cas) général autour de 3,67 % en Chine. Toutefois, ce chiffre cache de grandes disparités. La létalité est bien plus faible pour les personnes de moins de 60 ans (1,3-1,4 %) comparativement à celles de plus de 60 ans (4,5-6,4 %) voir les personnes de plus de 80 ans (13,4 %) [Verity et al., 2020]. Bien que la létalité reportée du SARS-CoV-2 soit moindre que celle du SARS-CoV ou du MERS, en moins de trois mois de pandémie, il a causé plus de mort que les deux précédents combinés depuis leur découverte [Mahase, 2020]. En août 2023, d'après les données du site internet [Our World in Data](#) [2023], on estimait autour de sept millions de morts dans le monde des suites d'une infection au SARS-CoV-2, dont près de 170 000 en France.

Comme d'autres virus à ARN, les coronavirus présentent un taux de mutation et de recombinaison génétique élevé de l'ordre de 1×10^{-6} –

2×10^{-6} mutations par nucléotide et par cycle de réplication. Depuis la survenue de la pandémie, des campagnes de séquençage massives dans de nombreux pays ont permis de renseigner en temps réel l'émergence de nouvelles mutations et les changements dans les dynamiques infectieuses consécutives aux nouveaux variants, en particulier liées aux mutations dans la protéine S. Les premiers mois de la pandémie ont vu relativement peu de changements phénotypiques des infections causées par le SARS-CoV-2 même si, assez tôt, une substitution dans la protéine spike, la D614G, a été repérée et associée à un avantage du mutant dans la transmissibilité [Jackson et al., 2022]. Ce variant a rapidement été majoritaire en Europe. À partir d'octobre 2020, on note l'apparition de nouveaux variants se distinguant par la présence d'un grand nombre de mutations non-synonymes dans la protéine S [Carabelli et al., 2023]. Ces mutations ont généralement été associées à des augmentations dans la transmissibilité des infections, leur virulence et leur échappement immunitaire (résistance aux anticorps, notamment vaccinaux) [Jackson et al., 2022]. Cinq variants ont fait l'objet d'attentions particulières de par leur propagation intense dans certaines régions, pour les variants Alpha, Beta et Gamma, ou à l'échelle globale, pour les variants Delta et Omicron.

Bien que d'intérêt majeur en santé publique, il est compliqué d'évaluer de l'évolution de la virulence du SARS-CoV-2. Les remplacements du variant B.1.177 par le variant Alpha et celui d'Alpha par le variant Delta ont été associés à des augmentations dans la sévérité de la maladie [Alizon and Sofonea, 2021]. À l'inverse, le remplacement de Delta par Omicron a induit une diminution des symptômes, mais une augmentation des taux d'infectiosité [Carabelli et al., 2023].

Rapidement, des vaccins ont été développés pour lutter contre les infections de SARS-CoV-2 et limiter les risques de complications des suites d'une infection virale. Des études ont évalué que des centaines de milliers de morts ont été évitées sur une période de 9 mois grâce à la vaccination généralisée [Steele et al., 2022]. Ces vaccins se sont révélés relativement efficaces contre la majorité des variants même si la protection immunitaire a tendance à décroître avec le temps et à être moindre envers certains variants [Lin et al., 2022; Zeng et al., 2022]. Des études ont montré que l'utilisation de *boosters* (ou rappels vaccinaux) permet d'augmenter l'efficacité de la protection contre les variants Delta et Omicron [Zeng et al., 2022].

Le SARS-CoV-2 est encore en circulation trois ans après la survenue de la pandémie. Les instances de santé publique continuent de suivre l'évolution de la propagation du virus, en particulier le dernier variant en date BA.2.86⁵, tout comme d'autres virus respiratoires (grippe, virus respiratoire syncytial) [Center for Disease Control and Prevention, 2023], proposant de nouvelles campagnes vaccinales pour les personnes les plus à risques [Ministère de la santé et de la prévention, 2023]. Cependant, la

5. La rédaction du présent document a lieu fin décembre 2023.

suppression des indicateurs de suivis force à utiliser des nouveaux outils de surveillance du virus, par exemple à travers le séquençage des eaux usées [Huisman et al., 2022].

1.3 La modélisation mathématique, outil clé de l'épidémiologie

La modélisation mathématique est la mise en équation de phénomènes réels. En biologie, les modèles peuvent être employés dans presque toutes les circonstances pour répondre à des problématiques diverses [Djidjou-Demasse et al., 2020].

Bien avant la découverte des virus, les outils de modélisation en épidémiologie ont éclairé les décisions dans la mise en place de politiques de prévention. En 1760, Daniel Bernoulli développe un modèle déterministe de l'épidémie de variole et montre l'augmentation de l'espérance de vie consécutive à la variolisation [Bernoulli and Chapelle, 2023]. De nombreux chercheurs et chercheuses ont ouvert la voie de l'épidémiologie mathématique moderne [Brauer, 2008]. On peut notamment penser à Ronald Ross qui développa un modèle mathématique de propagation de la malaria et milita en faveur de l'utilisation de la modélisation mathématique dans ce qui s'appelait encore la pathométrie [Ross, 1997; Bacaër, 2011]. Il collabora également avec Hilda Hudson avec qui ils établirent les prémisses du modèle épidémiologique de base, le SIR [Ross and Hudson, 1997a,b; Magal and Ruan, 2014]. Ce modèle a été repris et enrichi avec les travaux pionniers de Kermack et McKendrick sur les modèles compartimentaux [Kermack et al., 1997a,b,c] que nous détaillerons ultérieurement.

Les systèmes biologiques ne peuvent être décrits tels quels en équation mathématique. Ce sont souvent des processus impliquant de grands nombre de variables d'états (par exemple conditions environnementales, caractéristiques du pathogène) avec une forte intrication des chaînes causales. Conjointement, ces deux propriétés induisent un grand nombre de paramètres, qui lui même introduit des problèmes d'identifiabilité et de mesure, et la non-linéarité du système mathématique compliquant les analyses analytique. Les modèles mathématiques sont donc des représentations simplifiées, sous l'angle mathématique, des réalités biologiques. Ce qui peut induire une certaine distance entre les conclusions du modèle et les applications biologiques. La confrontation du modèle aux données et connaissances biologiques est cruciale pour éviter de tomber dans certains écueils et le dialogue avec des biologistes plus fondamentaux est souvent nécessaire pour ne pas perdre de vue les enjeux réels dans le domaine étudié.

Les systèmes biologiques sont souvent trop complexes pour être appréhendés dans leur globalité par un seul modèle. La modélisation consiste ainsi en un vaste panel d'approches méthodologiquement diverses et com-

plémentaires. Toutes ces approches ont en commun la simplification du processus biologique par la formalisation mathématique, mais des qualités et des faiblesses, voir des difficultés, différentes inhérentes au formalisme choisi. Les outils employés ainsi que le degré de complexité du modèle sont généralement liés à la question biologique que l'on se pose et aux données accessibles. Il a été longtemps évoqué la difficulté en modélisation de phénomènes biologiques de construire des modèles qui soient à la fois précis, réalistes et généraux [Levins, 1966]. L'augmentation dans l'un ou deux de ces critères conduisant généralement à une dégradation dans le troisième critère. Ainsi, trois types de modèles sont généralement considérés maximisant deux des trois critères évoqués précédemment. Bien que critiquée [Matthewson, 2011], cette trichotomie illustre bien l'enjeu en modélisation des systèmes biologiques de la pluralité des approches. En particulier, dans ma thèse j'ai privilégié les modèles stochastiques aux approches déterministes pour les travaux plus théoriques. Quand les données le permettait, j'ai complété mes analyses à l'aide de modèles statistiques. À partir de l'échelle moléculaire, les systèmes biologiques ne sont pas intrinsèquement aléatoires. Néanmoins, le nombre d'acteurs impliqués dans ces systèmes, l'ensemble des possibilités d'actions de ces acteurs ou une combinaison des deux induit une superposition d'un grand nombre de chaînes causales qui complique l'analyse du phénomène. Cela induit de la contingence et donc de la variabilité dans les scénarios possibles. Même si incorrect sur le principe, la stochasticité permet de rendre compte de cette variabilité induite par un ensemble de micro-processus à l'œuvre à l'aide d'un nombre limité de variables et de paramètres qu'une approche plus réaliste du processus biologique en jeu. De nombreux modèles stochastiques ont des équivalents déterministes, souvent plus facile à analyser ou simuler numériquement, les algorithmes de simulation de processus aléatoires (Gillespie, tau-leap) étant gourmands en temps et en ressources informatiques. Bien que présentant des similitudes, notamment lorsque le nombre de réactions est grand, ces deux approches diffèrent dans leur prédiction de la tendance au long terme. Les approches déterministes par essence produisent un unique résultat étant donné une condition initiale et un jeu de paramètres donné. Les modèles stochastiques fournissent un ensemble de trajectoires qu'il convient d'analyser pour estimer leur loi de probabilité ou leurs moments. Les systèmes déterministes ne sont pas dénués d'incertitude pour autant. En particulier, des petites variations dans le choix des valeurs des paramètres peut induire des changements de dynamique majeurs. On peut rendre compte de cette incertitude à l'aide de diverses méthodes d'analyse de sensibilité (par exemple par décomposition de la variance [Sobol, 2001; Saltelli et al., 2007, 2010]). Néanmoins, aussi fidèle soit l'approximation d'un processus stochastique par un système déterministe, elle induit un biais structurel qui peut s'avérer délicat à gérer. Ce biais est une manifestation directe de l'inégalité de Jensen ⁶[Smeach and Gold, 1975].

6. Soit f une fonction convexe et X une variable aléatoire, l'inégalité de Jensen

La complémentarité des modèles employés permet de répondre à des enjeux auxquelles les approches intuitives ou expérimentales ne permettent pas d’apporter d’éléments [Djidjou-Demasse et al., 2020]. La modélisation, comme toute démarche en recherche, demeure néanmoins une entreprise humaine. En cela, les choix de modélisation n’émergent pas nécessairement de choix raisonnés mais plus d’habitudes voir de la popularité d’une approche dans un domaine. La modélisation en biologie est ainsi une discipline profondément inter-disciplinaire à l’interface entre les domaines de la biologie et des mathématiques appliquées.

Dans la suite du manuscrit, je développe les bases des outils utilisés au cours de ma thèse au travers d’exemples jouets qui serviront à illustrer les concepts clefs des méthodes évoquées. En premier lieu, nous verrons les modèles compartimentaux, notamment à travers l’exemple type du SIR/SIS, tant de manière déterministe que stochastique ainsi que les extensions que l’on peut apporter à ce modèle de base. Ensuite, nous discuterons des bases de l’analyse de survie et de son utilisation pour les données longitudinales. Enfin, nous détaillerons le principe des chaînes de Markov cachées et les méthodes d’inférence des paramètres du modèle.

1.3.1 Modèles compartimentaux en épidémiologie

On parle de modèles compartimentaux lorsque l’on divise la population étudiée en plusieurs groupes supposés parfaitement homogènes sur le plan clinico-épidémiologique. Les individus d’une catégorie peuvent généralement changer de catégorie au cours du temps : on parle alors de transition d’un état vers un autre. On considère que les individus d’un même groupe partagent les mêmes caractéristiques et les mêmes taux de transitions indépendamment de leur entrée dans le compartiment. Les taux de transitions entre catégories dépendent classiquement de paramètres fixes, exceptée la force d’infection, et du nombre d’individus dans cette sous-population. Il n’y a pas de limite au nombre de catégories utilisables ni au nombre de transitions. Néanmoins, augmenter le nombre de compartiments augmente le nombre de transitions possibles et donc le nombre de paramètres nécessaires. Sauf connaissances a priori, cela nécessite d’estimer de nombreux paramètres à partir de données biologiques avec le risque qu’ils ne soient pas identifiables⁷.

Dans cette section, nous introduisons d’abord des modèles de base en épidémiologie, que sont les modèles SIR et SIS dans leur version déterministe. Nous évoquerons également des extensions de ces modèles classiques, en particulier dans le cadre du COVID-19. Ensuite, nous parlerons des spécificités des processus de naissances et morts, notamment en épidémiologie. Pour conclure cette partie, nous discuterons des limites biologiques des lois exponentielles inhérentes à ces approches classiques

énonce que $f(E[X]) \leq E[f(X)]$.

7. On dit qu’un modèle est identifiable si il est possible d’inférer la vraie valeur des paramètres du modèle à partir d’un nombre fini d’observations.

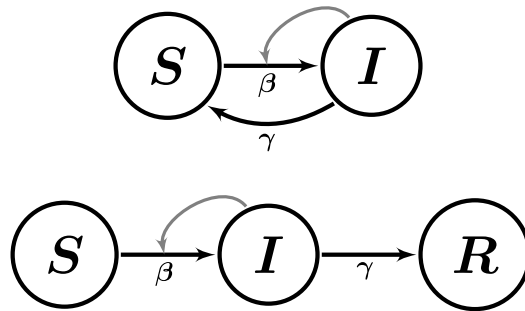


FIGURE 1.2 – **Modèles SIS et SIR.** Le diagramme du haut correspond au modèle SIS et celui du bas au modèle SIR. La catégorie **S** correspond aux individus susceptibles (personnes à risque de contracter la maladie), le groupe **I** aux personnes infectées qui peuvent transmettre le pathogène et l'ensemble **R** aux personnes immunisées ou décédées (qui ne peuvent plus être contagieux).

et évoquerons des outils permettant de se défaire des certaines propriétés peu réalistes du point de vue biologique.

Les modèle de base en épidémiologie : SIR et SIS

Dans le cadre d'infections aiguës, durant lesquelles le virus persiste chez l'hôte quelques jours ou semaines, on peut modéliser la propagation d'une maladie dans une population à l'aide d'un modèle SIR ou SIS, où la lettre **S** correspond aux individus sensibles (personnes à risques de contracter la maladie), **I** aux personnes infectées qui peuvent transmettre le pathogène et **R** aux personnes immunisées ou décédées (individus retirés de la chaîne de transmission). On utilisera un modèle SIR dans le cadre d'infections qui génèrent une immunité parfaite contre les réinfections chez les hôtes infectés. En revanche, pour décrire la dynamique de maladies qui ne confèrent pas cette immunité, on utilisera plutôt le modèle SIS [Grassly et al., 2005]. Les premiers travaux sur ces modèles ont été publiés dans les années 1930 via une série de trois articles de William Ogilvy Kermack et Anderson Gray McKendrick [Kermack et al., 1997a,b,c]. On peut représenter graphiquement ces deux modèles à l'aide de diagrammes, comme dans la Figure 1.2.

Mathématiquement, on peut décrire la dynamique du système à l'aide d'un système d'équations différentielles ordinaires minimaliste, qui permet néanmoins de capturer la forme de la vague épidémique. Dans un contexte d'immunité parfaite, sans démographie, on écrit le modèle SIR fréquence-dépendant ⁸

8. En modélisation fréquence-dépendant, le taux de contact est indépendant de la taille totale de la population mais dépend de la proportion d'individus infectés.

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta SI}{N} \\
\frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{aligned}
\tag{1.1}$$

avec β le taux de transmission, γ le taux de guérison et $N = S + I + R$ le nombre total d'individu dans la population. Dans un modèle SIS, on suppose à l'inverse que les individus redeviennent directement sensibles une fois la période infectieuse terminée, on écrit

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta SI}{N} + \gamma I \\
\frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I
\end{aligned}
\tag{1.2}$$

Implicitement, en posant le modèle de la sorte, on fait plusieurs hypothèses majeures pour notre système d'étude, en premier lieu celle de champ moyen. On considère la population comme homogène de sorte qu'aucune structure spatiale ne régit la répartition des individus, cela revient à une distribution ponctuelle des individus. Ainsi, chaque individu interagit avec la même probabilité avec chacun des autres individus de la population générale [Keeling and Rohani, 2008]. Ensuite, on suppose que les infectés guérissent de l'infection au taux γ , ce qui revient à dire que chaque individu passe en moyenne une durée $1/\gamma$ dans le compartiment I avant de le quitter. Cette quantité $1/\gamma$ peut s'interpréter comme la durée d'infectiosité. Enfin, on se place dans un cadre sans démographie, c'est-à-dire que la population totale d'individus reste constante au cours du temps, de sorte que, soit $N \in \mathbb{N}^*$

$$\forall t \in \mathbb{R}_+, \quad S(t) + I(t) + R(t) = N
\tag{1.3}$$

Cette dernière hypothèse traduit généralement le fait que l'échelle de temps durant lequel a lieu l'épidémie est beaucoup plus courte que l'échelle de temps démographique de la population. Autrement dit, en première approximation, on peut négliger la dynamique démographique.

Il n'est pas possible de trouver une solution explicite au modèle SIR (équation 1.1). En revanche, le modèle SIS sans démographie (équation 1.2) peut s'écrire comme une fonction logistique :

$$\frac{dI}{dt} = (\beta - \gamma) I - \frac{\beta}{N} I^2
\tag{1.4}$$

Si $\beta > \gamma$, on peut montrer que la dynamique des infectés au cours du temps suit la fonction suivante :

$$\begin{aligned} \frac{dI}{dt} &= (\beta - \gamma)I - \frac{\beta}{N}I^2 \\ \Leftrightarrow I(t) &= \frac{(\beta - \gamma)N}{\beta + [(\beta - \gamma)N/I_0 - \beta]e^{-(\beta - \gamma)t}} \end{aligned} \quad (1.5)$$

Ainsi, quand $t \rightarrow \infty$ et $\beta > \gamma$, la fonction $I(t)$ converge vers un état d'équilibre endémique $I(\infty) = N(1 - \gamma/\beta)$. À l'inverse, si $\beta < \gamma$, l'infection s'éteindra à un taux exponentiel proportionnel à $\beta - \gamma$.

Même sans solutions explicites, il est possible d'extraire de l'information sur la dynamique du système SIR à l'aide des équations 1.1. On peut ré-écrire la dynamique des infectés comme suit :

$$\frac{dI}{dt} = \left(\frac{\beta}{N}S - \gamma \right) I \quad (1.6)$$

Il en découle en particulier que si $s(0) < \gamma/\beta$, avec $s(0) = S(0)/N$, la dynamique du nombre d'infectés est décroissante et l'infection ne se propage pas. Inversement, si $s(0) > \gamma/\beta$, une épidémie a lieu [Hethcote, 2000]. Contrairement au modèle SIS, on peut montrer que dans un modèle SIR l'épidémie est vouée à s'éteindre. Brièvement, dans un SIR, le nombre d'individus susceptibles décroît forcément, remplacés à terme par des individus qui ne prennent pas part à la transmission. La chaîne de transmission finit donc par casser par manque d'individus sensibles qui ne peuvent plus être infectés.

Ces deux approches sont modèles conceptuels très utiles dans la description des épidémies. En particulier, l'effet de seuil des différents modèles est un concept clef de l'analyse des dynamiques épidémiques. On définit ainsi le nombre de reproduction de base R_0 comme le nombre d'infections secondaires causées par un unique individu infectieux dans une population entièrement susceptible. Dans notre cas, la quantité $R_0 = \beta/\gamma$, nous renseigne sur le potentiel épidémique d'un pathogène car si $R_0 > 1$, une épidémie a lieu. Néanmoins, il ne renseigne pas directement sur la vitesse de propagation de la maladie. Au début de l'épidémie, le temps de doublement peut s'écrire comme une fonction du nombre de reproduction de base et du temps de génération (temps écoulé entre l'infection d'un cas primaire et ses cas secondaires). À taux de génération équivalent, le pathogène avec le plus grand R_0 se propage plus vite dans la population que les autres. Enfin, le nombre de reproduction de base ne nous renseigne sur la dynamique épidémique que durant ses premiers instants. À mesure que le pathogène se propage, le nombre de susceptibles diminue dans la population et le potentiel de propagation du pathogène diminue. Il se peut également que le taux de transmission diminue au cours du temps (par exemple avec le port du masque), le taux de guérison étant généralement moins variable. On utilise alors le nombre de reproduction effectif $R_e = R_0 c(t)S(t)/S(0)$, avec $c(t)$ l'efficacité des mesures de contrôle, pour décrire la dynamique épidémique dans ses phases ultérieures.

En pratique, les modèles de base SIR ou SIS sont souvent trop simplistes pour décrire la dynamique infectieuse de nombreux virus. Par exemple, certains présentent des périodes de latence durant lesquelles ils persistent dans l'hôte sans être infectieux ou encore des périodes asymptomatiques durant lesquelles l'individu est infectieux sans symptômes manifestes. Il peut donc s'avérer nécessaire d'enrichir le modèle de base pour prendre en compte ces spécificités. On utilise par exemple souvent le modèle SEIR en ajoutant un compartiment « exposé », contenant des individus infectés mais non infectieux. Cela permet de décrire les virus présentant une phase de latence. Dans d'autres circonstances, on peut évaluer la propension d'un variant à envahir une population en le distinguant de la souche déjà présente à l'aide d'un modèle multi-type. De manière générale, les modèles compartimentaux sont des outils flexibles. On peut aisément rajouter de nouveaux compartiments, modifier les transitions entre compartiments selon le choix de modélisation. La représentation en diagramme de flux permet de visualiser rapidement et efficacement le type de modèle considéré, comme par exemple dans la Figure 1.3. Néanmoins, à mesure que le nombre de compartiments ou transitions augmente, l'inférence des paramètres du modèle se complique et les soucis d'identifiabilité se renforcent.

Même dans ces cas plus compliqués, le nombre de reproduction de base est toujours une donnée cruciale à estimer pour évaluer le risque épidémique d'un pathogène. D'après [Kermack et al. \[1997a,b,c\]](#), le nombre de reproduction peut s'écrire comme le produit de la probabilité de transmission d'un pathogène, le nombre de contacts par unité de temps et la durée de la phase contagieuse. Néanmoins, le calcul direct de sa valeur est souvent impossible pour des systèmes d'équations non-linéaires aux structures de population complexes. La méthode de la *Next Generation Matrix* [[Diekmann et al., 1990](#)] permet parfois de simplifier grandement l'expression du R_0 et, dans tous les cas, de le définir analytiquement en linéarisant le système d'EDO autour d'une valeur à l'équilibre. On peut ainsi décomposer la Jacobienne du système d'EDO comme la différence entre la matrice \mathbf{F} (qui récapitule les taux d'apparitions des individus dans les différentes classes infectieuses) et la matrice \mathbf{V} (qui décrit les flux d'individus entre les différentes classes infectieuses et vers les compartiments non-infectieux), avec \mathbf{F} une matrice carrée dont tous les éléments sont positifs ou nuls et \mathbf{V} une matrice carrée qui vérifie $s(-\mathbf{V}) < 0$, avec $s(\cdot)$ la limite spectrale d'une matrice (c'est-à-dire la partie réelle maximale des valeurs propres [[Hirsch and Smale, 1974](#)]), et dont l'inverse $\mathbf{V}^{-1} \geq 0$ [[Hurford et al., 2009](#)]. Si on dispose de ces deux matrices \mathbf{F} et \mathbf{V} alors par application du théorème de *Next Generation* on peut calculer le R_0 comme le rayon spectral, c'est-à-dire la plus grande valeur propre, de la matrice \mathbf{FV}^{-1} [[Diekmann et al., 1990](#); [Hurford et al., 2009](#)].

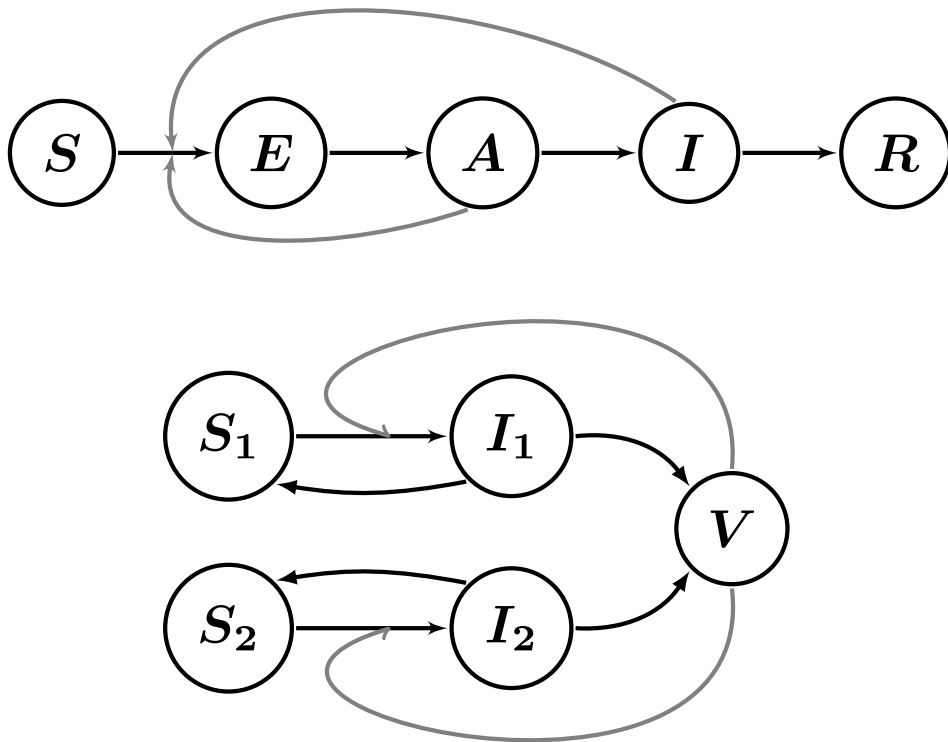


FIGURE 1.3 – Diagrammes de flux pour des modèles SEAIR et $S_1S_2I_1I_2V$. Le diagramme du haut décrit le modèle SEAIR et celui du bas le modèle $S_1S_2I_1I_2V$. Le premier modèle peut être utilisé pour décrire la dynamique d'un pathogène présentant une phase d'incubation non-infectieuse et une phase asymptomatique infectieuse (par exemple dans le cas du COVID-19, voir chapitre 5). Le second cas peut être employé pour décrire l'infection d'un tissu contenant deux lignées cellulaires qui peuvent être infectées par une même particule virale.

Processus Markovien

Par essence, étant donné une condition initiale et un jeu de paramètres donné, les outils déterministes produisent une unique solution. La description du système par des équations différentielles ordinaires décrit la dynamique du processus pris dans sa globalité et met généralement de côté l'hétérogénéité individuelle au profit d'une tendance moyenne. Du point de vue de la modélisation, il peut s'avérer intéressant de passer par une description du processus centrée sur l'individu pour ensuite déduire des tendances à l'échelle de la population.

Utilisés à l'origine pour étudier le maintien des familles nobles françaises [Bienaymé, 1845] et anglaises [Watson and Galton, 1875] à travers la survie de descendants mâles porteurs du patronyme de la famille, les processus de naissance et morts sont des outils généralistes qui permettent de décrire la dynamique de nombreux processus biologiques à l'échelle intra-hôte comme inter-hôte. Fondamentalement, les approches déterministes et stochastiques ne sont pas incompatibles. De fait, de nombreux modèles stochastiques classiques ont des équivalents déterministes [Allen, 2015; Meleard and Bansaye, 2015]. En particulier, on peut voir les systèmes d'équations différentielles ordinaires comme une approximation de processus stochastiques dans une population de taille infinie [Kurtz, 1970, 1971, 2003].

Avant de décrire plus en profondeur les caractéristiques des processus de naissance et morts, nous commencerons par définir les caractéristiques de bases de processus stochastiques plus généraux que sont les chaînes de Markov, que nous reverrons dans la partie 1.3.3. Comme nous le détaillerons ultérieurement, les processus de naissance et mort sont des cas particuliers de chaînes de Markov définis sur \mathbb{N} et dont l'amplitude de transitions vaut ± 1 . Les éléments abordés ici ne visent qu'à introduire brièvement le sujet, pour plus de détails on peut se rapporter aux travaux de Athreya and Ney [1972], Haccou et al. [2005] ou encore Allen [2010].

Chaînes de Markov discrètes et continues. Tirant leur nom du nom de leur inventeur Andrei Markov [Markov, 2006, 1954], les chaînes de Markov sont des objets mathématiques stochastiques qui décrivent les changements d'états d'un processus au cours du temps. Ce sont des outils généraux qui peuvent être employés pour modéliser des dynamiques épidémiques, la progression d'une maladie ou des changements de bases nucléotidiques [Allen, 2010; Yoon, 2009]. Une caractéristique fondamentale de ces outils est la propriété de Markov : la loi des états futurs ne dépend du passé que par l'état présent. Pour cette raison, on dit des chaînes de Markov que ce sont des processus sans mémoire. Mathématiquement, cela se traduit dans la loi conditionnelle du processus en fonction des états passés.

Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov en temps discret définie comme une suite de variables aléatoires à valeurs dans \mathcal{E} , espace dénombrable non

numérique, qui vérifie la propriété de Markov suivante : soit $i_0, \dots, i_n, j \in \mathbb{Z}$, on a

$$\mathbb{P}[X_{n+1} = j | X_0 = i_0, \dots, X_n = i_n] = \mathbb{P}[X_{n+1} = j | X_n = i_n]. \quad (1.7)$$

Si les probabilités de transitions sont indépendantes du temps écoulé, on parle alors de chaîne de Markov homogène. Les transitions d'une chaîne de Markov homogène peuvent être alors décrites à l'aide d'une matrice de transition P telle que

$$P_{i,j} = \mathbb{P}[X_1 = j | X_0 = i] \quad ; \quad i, j \in \mathbb{Z} \quad (1.8)$$

Étant donné une condition initiale $X_0 = i_0$ et une matrice P , on peut décrire la loi d'une chaîne de Markov discrète par

$$\mathbb{P}[X_0 = i_0, \dots, X_n = i_n] = \mathbb{P}[X_0 = i_0] P_{i_0, i_1} \dots P_{i_{n-1}, i_n}. \quad (1.9)$$

La matrice P d'une chaîne de Markov est très informative sur le processus sous-jacent et permet notamment de calculer rapidement la probabilité de transition du processus à l'itération n , soit $\mathbb{P}[X_n = j | X_0 = i] = P_{i,j}^{(n)}$ avec $P^{(n)}$ la matrice P à la puissance n . En particulier, si la chaîne de Markov est irréductible, on peut trouver sa probabilité stationnaire.

La notion de chaîne de Markov peut être généralisée en temps continu. Soit $(X_t)_{t \in \mathbb{R}_+}$ un processus de Markov défini en temps continu. Il satisfait la propriété de Markov dans le sens où, soit $(\mathcal{F}_t^X)_{t \geq 0}$ une filtration engendrée par un processus X , $(\mathcal{F}_t^X)_{t \geq 0} := \sigma(\{X_s, s \leq t\})$ avec $\sigma(X_s, s \leq t)$ la tribu engendrée par l'ensemble des trajectoires du processus X sur l'intervalle $[0, t]$, on a

$$\mathbb{P}[X_{t+s} | \mathcal{F}_t^X] = \mathbb{P}[X_{t+s} | X_t] \quad (1.10)$$

Si la probabilité $\mathbb{P}[X_t = j | X_s = i]$ ne dépend de s et de t que par la différence de temps $t - s$, alors on dit que le processus est homogène. Pour de tels processus, on peut écrire :

$$\mathbb{P}[X_t = j | X_s = i] = P_{i,j}(t - s) \quad (1.11)$$

où $P_{i,j}(t)$ est une matrice carrée $\mathcal{E} \times \mathcal{E}$ de taille potentiellement infinie qui vérifie, $\forall i, j \in \mathcal{E}, \forall t > 0$,

$$P_{i,j}(t) \geq 0 \quad ; \quad \sum_j P_{i,j}(t) = 1. \quad (1.12)$$

Les matrices de transition $P(t)$ vérifient également l'équation de Chapman-Kolmogorov :

$$P(0) = I_{\mathcal{E}} \quad ; \quad P(t+s) = P(t)P(s) = P(s)P(t) \quad (1.13)$$

En pratique, pour décrire un processus Markovien, il n'est pas nécessaire de connaître l'information de $P(t)$ pour tous $t > 0$. De la relation

1.13, on peut déduire l'entièreté de la connaissance de $P(t)$ de la seule connaissance de la dérivée à droite de P en 0.

Soit une matrice $(Q_{i,j}; i, j \in \mathbb{N})$ appelée générateur infinitésimal du processus $(X_t)_t$ qui vérifie :

$$Q_{i,j} \geq 0 \text{ si } i \neq j \quad ; \quad Q_{i,i} = - \sum_{j \in \mathcal{E} \setminus \{i\}} Q_{i,j} \leq 0 \quad (1.14)$$

On a alors pour $\tau \rightarrow 0$:

$$\begin{aligned} P_{i,j}(\tau) &= \tau Q_{i,j} + o(\tau) \quad \text{si } i \neq j \\ P_{i,i}(\tau) &= 1 - \tau q_i + o(\tau) \end{aligned} \quad (1.15)$$

avec $q_i = -Q_{i,i}$.

En particulier, on peut réécrire les équations de Kolmogorov tel que $P(t)$, $t \geq 0$, est l'unique de solution de :

$$\begin{aligned} dP(t)/dt &= QP(t) && \text{(Rétrograde)} \\ dP(t)/dt &= P(t)Q && \text{(Progressive)} \end{aligned} \quad (1.16)$$

Les équations de Chapman-Kolmogorov sont cruciales pour établir des relations de récurrence entre des probabilités et ainsi calculer de nombreuses quantités d'intérêt comme par exemple les moments d'une loi de probabilités pour un processus.

Processus de naissance et mort. Mathématiquement, les processus de naissance et mort sont des processus markovien de saut à valeurs dans \mathbb{N} dont les amplitudes de saut sont égales à 1. Un processus de saut est décrit par des trajectoires presque-sûrement⁹ continues à droite et limitées à gauche (càdlàg) et constantes entre des instants de sauts isolés. Soit $(X_t, t \geq 0)$ un processus de saut à valeurs dans un espace \mathcal{E} dénombrable, on peut le décrire à l'aide d'un couple de variables aléatoires $(T_n, Z_n), n \geq 0$ où T_n est le n -ième instant de saut et $Z_n \in \mathcal{E}$ la position du processus juste avant le saut au temps T_{n+1} . Par définition, on pose $T_0 = 0$ et les instants de saut forment une suite croissante telle que $T_n < T_{n+1}$ si $T_n < +\infty$. Si on suppose également que les instants de sauts ne s'accumulent pas en temps fini, c'est-à-dire que $\lim_{n \rightarrow \infty} T_n = +\infty$, on peut définir le processus X_t comme :

$$X_t = \sum_{n \geq 0, T_n < +\infty} Z_n \mathbf{1}_{[T_n, T_{n+1}[}(t) \quad (1.17)$$

avec $\mathbf{1}_{[A,B]}$ la fonction indicatrice qui prend la valeur 1 sur l'intervalle $[A, B]$ et 0 sinon.

Les processus de naissance et mort sont des processus de saut particulier dont les amplitudes des sauts sont égales à ± 1 . Pour un processus de naissance et mort à un seul type, on note généralement les transitions

9. C'est-à-dire que la propriété est vraie partout en dehors d'un ensemble négligeable.

$$\begin{cases} i \rightarrow i + 1 & \text{au taux } \lambda_i \geq 0 \\ i \rightarrow i - 1 & \text{au taux } \mu_i \geq 0 \end{cases} \quad (1.18)$$

Si on se place dans un contexte sans immigration ni démographie, alors l'état 0 est absorbant et $\lambda_0 = 0$. La matrice du générateur Q vaut donc $Q_{i,i+1} = \lambda_i, Q_{i,i-1} = \mu_i, Q_{i,i} = -(\lambda_i + \mu_i)$. On peut montrer qu'un processus de naissance et mort est défini sur \mathbb{R}_+ si $\forall i$

$$\lambda_i \leq \lambda_{i+1}. \quad (1.19)$$

Plusieurs quantités sont particulièrement d'intérêt pour caractériser le processus. En premier lieu, la probabilité d'extinction, notée $\mathbb{P}_i[\exists t > 0, X_t = 0]$, est cruciale pour évaluer le risque de propagation d'une épidémie ou le risque de disparition d'une population. Pour l'évaluer, on commence par calculer la somme

$$U_\infty = \sum_{k=1}^{\infty} \frac{\mu_1 \dots \mu_k}{\lambda_1 \dots \lambda_k}. \quad (1.20)$$

Si $U_\infty = +\infty$, alors le processus s'éteint presque-sûrement en temps fini. Sinon, la probabilité d'extinction vaut

$$\mathbb{P}_i[\exists t > 0, X_t = 0] = (1 + U_\infty)^{-1} \sum_{k=i}^{\infty} \frac{\mu_1 \dots \mu_k}{\lambda_1 \dots \lambda_k}. \quad (1.21)$$

Pour un processus dont la probabilité d'extinction vaut 1, on peut calculer le temps moyen à l'extinction (T_0) :

$$\mathbb{E}_n[T_0] = \sum_{k=1}^{n-1} \left(\sum_{i \geq k+1} \frac{\lambda_{k+1} \dots \lambda_{i-1}}{\mu_{k+1} \dots \mu_i} \right) \quad (1.22)$$

Comme on l'a vu précédemment, on peut réduire l'étude de la dynamique d'un modèle SIS à la seule étude de la dynamique du compartiment **I**. En reprenant les notations de l'équation 1.2 (SIS sans démographie), le générateur du processus s'écrit

$$\begin{cases} \lambda_i & = \frac{\beta}{N} i(N - i) \\ \mu_i & = \gamma i \end{cases}. \quad (1.23)$$

On vérifie aisément que la propriété de non-explosion de l'équation 1.19 est bien vérifiée dans le cas d'un SIS. Pour le modèle SIR, on ne peut pas simplifier la dynamique du processus par la description d'un seul compartiment. On utilise alors un processus de branchement multi-type défini par

$$\mathbb{P}[\Delta S(t) = i, \Delta I(t) = j | (S(t), I(t))] = \tag{1.24}$$

$$\begin{cases} \frac{\beta}{N}S(t)I(t)\tau + o(\tau) & (i, j) = (-1, 1) \\ \gamma I(t)\tau + o(\tau) & (i, j) = (0, -1) \\ 1 - \frac{\beta}{N}S(t)I(t)\tau - \gamma I(t) + o(\tau) & (i, j) = (0, 0) \\ o(\tau) & \text{sinon} \end{cases}$$

avec $\tau \rightarrow 0$, $\Delta S(t) = S(t+\tau) - S(t)$ et $\Delta I(t) = I(t+\tau) - I(t)$. Pour un processus de branchement multi-type, des outils similaires peuvent être employés en organisant la matrice Q pour prendre en compte l'ensemble des combinaisons possibles. En renormalisant ces processus de naissance et mort par un facteur $1/K$, $K \rightarrow +\infty$, il est possible de calculer un équivalent déterministe à notre processus [Kurtz, 1971, 1970, 2003]. Notamment, on peut montrer qu'un modèle SIS stochastique converge bien vers un équivalent logistique. Néanmoins, les deux modèles présentent des résultats asymptotiques différents. En particulier, on peut montrer qu'un modèle SIS stochastique s'éteint presque-sûrement en temps fini. De plus, à partir de l'inégalité de Jensen, on peut montrer que l'espérance du processus stochastique pour tout $t \in \mathbb{R}_+$ est inférieure ou égale à la valeur de l'équivalent déterministe [Smeach and Gold, 1975].

Simulation de processus stochastique. Il peut arriver que le processus soit trop compliqué à appréhender directement. Les approches de Monte-Carlo sont des méthodes numériques visant à estimer une valeur approchée d'une quantité à l'aide de la simulation de processus aléatoires [Metropolis and Ulam, 1949]. Soit $(X_t)_{t \in \mathbb{R}_+}$, un processus quelconque défini sur \mathcal{E} . Connaissant son générateur Q , il est aisé de construire un algorithme de simulation de trajectoires de ce processus [Gillespie, 1977]. Soit une condition initiale $i \in \mathcal{E}$ tel que $X_0 = i$

- i) On démarre de l'état $X_0 = i$ et on attend un temps exponentiel T_1 de paramètre q_i , le processus $X_t = X_0$ pour tout $t \in [0, T_1[$.
- ii) Au temps T_1 , le processus saute de l'état i vers l'état j , $j \neq i$ avec probabilité $Q_{i,j}/q_i$.
- iii) On attend un temps exponentiel T_2 de paramètre q_j , le processus $X_t = j$ pour tout $t \in [T_1, T_2[$.
- iv) Au temps T_2 , le processus saute de l'état j vers l'état k , $k \neq j$ avec probabilité $Q_{j,k}/q_j$.
- v) ...

Cette approche fonctionne bien lorsque les sauts ne sont pas trop rapprochés entre eux. En effet, lorsque les q_i deviennent trop grands, le temps entre chaque saut est faible et cela peut induire des problèmes numériques, tant dans le calcul du prochain saut que dans le stockage mémoire de la trajectoire. Des méthodes permettent d'approximer le processus stochastique sur des temps de pas temps fixes, comme par exemple

la méthode de *tau-leap*, qui moyenne l'ensemble des sauts qui ont lieu sur un intervalle de temps par un événement aléatoire suivant une loi de Poisson [Cao et al., 2006]. Des approches de simulations combinant Gillespie et *tau-leaping* permettent la génération rapide et précise de trajectoires stochastiques [Danesh et al., 2023a].

On peut ainsi estimer des quantités d'intérêt par génération d'un grand nombre de trajectoires indépendantes suivant la même loi de probabilité. L'estimateur de la moyenne empirique est un exemple d'utilisation de méthodes Monte-Carlo. De manière générale, il est possible d'estimer un grand nombre de quantités d'intérêt à l'aide de ces méthodes [Rubinstein and Kroese, 2017].

Extension des approches classiques

D'un point de vue biologique, les modèles classiques SIR et SIS présentent des limitations dans leur application. En premier lieu, la dynamique des deux approches, qu'elle soit stochastique ou déterministe, repose sur des temps de séjour distribués exponentiellement. Or la loi exponentielle est une loi « sans-mémoire » dans le sens où un individu nouvellement arrivé dans un compartiment aura la même chance de le quitter qu'un individu arrivé plusieurs unités de temps avant ce dernier.

Comme évoqué précédemment, une solution pour limiter ce problème peut consister en l'ajout de nouveaux compartiments, ce qui permet d'induire une certaine forme de latence. Toutefois, l'ajout de trop nombreux compartiments peut faire perdre son sens biologique au modèle et induire des problèmes d'identifiabilité [Yates et al., 2017].

Pour induire de la mémoire sur les états précédents, on peut également utiliser des processus de Markov de plus grand ordre [Raftery, 1985]. Dans ce contexte, la dynamique future du processus ne dépend plus uniquement de la valeur au temps présent, mais également des itérations précédentes. Évidemment, cela limite grandement les outils disponibles pour analyser le processus. Cependant, à l'aide des outils numériques de simulations par méthode de Monte-Carlo, il est possible d'obtenir des estimations suffisamment précises de la dynamique du processus d'intérêt [Ky and Tuyen, 2018].

Pour les modèles déterministes, le formalisme par systèmes d'équations aux dérivées partielles (EDP) permet également de prendre en compte explicitement le temps d'entrée dans un compartiment [Reyné et al., 2022].

De plus, dans les approches classiques, on fait généralement l'hypothèse que les individus d'un même compartiment ont les mêmes propriétés infectieuses et les mêmes taux de guérisons. En pratique, cette hypothèse est rarement vérifiée dans le cas d'une épidémie touchant des individus très différents, par exemple avec des écarts d'âges marqués. Pour décrire au mieux la dynamique du processus, il peut être important de prendre en compte cette diversité dans le potentiel infectieux. Pour cela, on peut

stratifier par un nombre fini de classes d'âges [Sofonea et al., 2021; Reyné et al., 2022] ou de manière continue à travers un formalisme EDP [Li, 1988; Ferrière and Tran, 2009].

1.3.2 Analyse de survie

L'analyse de survie est une branche des statistiques qui s'intéresse au temps avant la réalisation d'un événement d'intérêt. Utilisés d'abord en économie [Böhmer, 1912], ces outils sont pour la première fois employés en biologie en 1950 pour estimer le taux de survie de personnes atteintes de cancer [Berkson and Gage, 1950]. On peut avoir recours à l'analyse de survie pour analyser diverses séries temporelles, du temps avant la panne d'un appareil à l'efficacité d'un traitement médical [Kleinbaum and Klein, 2012].

L'analyse de survie comprend un ensemble de méthode diverses, dont les buts principaux sont i) d'estimer des fonctions de survie et d'aléa à partir de données brutes, ii) de comparer des fonctions de survie ou d'aléa entre elles, iii) d'évaluer l'effet de variables explicatives sur les temps de survie. Ces travaux s'appuient généralement sur des données issues de suivis longitudinaux, par exemple des essais cliniques évaluant l'efficacité d'un nouveau traitement/vaccin, ou le suivi de personnes infectées par un pathogène.

Un problème classique en analyse de survie est la censure des données. Celle-ci émerge lorsque l'on connaît le statut du patient, mais qu'on ne connaît pas précisément l'instant d'occurrence de l'événement. Un exemple courant dans les suivis longitudinaux est la censure à droite. Ceci se produit pour une observation dont on connaît le début, mais dont la réalisation de l'événement d'intérêt est inconnue. Plusieurs causes intrinsèquement liées à la conception des études cliniques peuvent être à l'origine de censure à droite, notamment car les protocoles d'inclusion et de suivis des études cliniques sont établis en amont d'une étude clinique et le nombre de visites par participant·e est bien souvent fixé à l'avance. De la censure à droite peut également apparaître lorsque des participant·es quittent une étude clinique prématurément. Cela peut avoir lieu lorsque l'on perd de vue un·e participant·e de l'étude pour une cause autre que la fin du suivi. Dans certains cas, on peut également être confronté à des observations censurées à gauche, observations dont l'initiation nous est inconnue mais dont on connaît la réalisation. Enfin, il peut arriver qu'on ne connaisse précisément le moment de réalisation de l'événement mais qu'on sache qu'il a eu lieu entre deux moments (par exemple deux visites sur site). Ce type de donnée est qualifié de censurée par intervalle. Ce genre de censure arrive fréquemment durant les suivis longitudinaux où le changement d'état du·de la patient·e est attesté entre deux visites consécutives sans que l'on sache précisément le moment du changement [Zhang and Sun, 2010].

Il est possible d'interpréter les données censurées à droite (resp. à

gauche) comme des données censurées par intervalle dont la borne supérieure (resp. inférieure) vaut $+\infty$ (resp. 0). Incorporer les données censurées aux analyses est important, car bien qu'incomplètes, elles fournissent de l'information. Cependant, dans la majorité des analyses, il convient de vérifier trois hypothèses concernant le mode de censure. Celle-ci doit être i) indépendante, ii) aléatoire et iii) non-informative [Kleinbaum and Klein, 2012]. Pour le premier cas, cela revient à s'assurer que dans tous les sous-groupes (par exemple les personnes vaccinées) les observations censurées au temps t sont représentatives de toutes les observations dans ce sous-groupe qui sont encore à risque au temps t . Mathématiquement, cela implique une relation d'égalité entre les deux fonctions d'aléas d'un même sous-groupe (censurés du groupe A *vs.* non-censurés du groupe A), concept que nous évoquerons ultérieurement. Le concept de censure aléatoire est plus contraignant que le concept d'indépendance. Il s'agit de vérifier que les sujets censurés au temps t sont représentatifs de tous les sujets à risque au temps t . Ainsi, la censure est indépendante si, pour chaque sous-groupe d'intérêt, elle est aléatoire. On parle de censure non-informative lorsque la distribution des temps de survie ne renseigne pas sur la distribution des temps de censure, et inversement. En particulier, la censure devient informative lorsque l'on retire de l'étude des participant-es pour raisons de santé liées au protocole suivi. Par exemple, dans une étude visant à comparer l'efficacité de deux traitements sur un cancer avec comme comparaison un groupe placebo, il est attendu que les individus du groupe témoin soient plus à risque de développer des complications et qu'il soit nécessaire de les sortir de l'étude car leur vie est en jeu.

Les outils classiques d'analyse de survie (e.g. estimateur de Kaplan-Meier, modèle de Cox ...) reposent sur ces hypothèses de censure indépendante, aléatoire et non-informative. Des méthodes existent pour analyser des données qui ne respectent pas ces trois hypothèses sur la censure, pour plus de détails on peut se référer à Kleinbaum and Klein [2012].

Définition des outils

On note par T , la variable aléatoire correspondant au temps de survenue de l'événement d'intérêt. La fonction de survie, notée $S(t)$, désigne la probabilité que l'événement d'intérêt dure plus que le temps t . Mathématiquement, on a

$$S(t) = \mathbb{P}[T > t] = 1 - \mathbb{P}[T \leq t] = 1 - F(t) \quad (1.25)$$

où $F(t)$ est la fonction de répartition de T . De fait, on peut montrer que $S(t) \in [0, 1]$ et qu'elle est décroissante sur \mathbb{R}_+ . Une autre fonction centrale de l'analyse de survie est la fonction d'aléa (*hazard function* en anglais), notée classiquement $\lambda(t)$ ou $h(t)$. Cette fonction correspond au risque instantané de réalisation de l'événement d'intérêt au temps t , sachant que celui-ci ne s'est pas encore réalisé avant t . Mathématiquement,

on définit la fonction d'aléa comme

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}[t \leq T < t + \Delta t | T \geq t]}{\Delta t} \quad (1.26)$$

Cette fonction d'aléa est une fonction de \mathbb{R}_+ et, contrairement à la fonction de survie, n'est pas bornée à 1. La fonction d'aléa et la fonction de survie sont deux faces d'une même pièce et, mathématiquement, on peut montrer que

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}[t \leq T < t + \Delta t | T \geq t]}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}[t \leq T < t + \Delta t]}{\Delta t S(t)} \\ &= \frac{f(t)}{S(t)} \\ &= -\frac{S'(t)}{S(t)} \end{aligned}$$

avec $f(t)$ la densité de probabilité de T , dérivée de la fonction de répartition $F(t)$. En intégrant entre 0 et t , on peut réécrire la fonction de survie comme

$$S(t) = e^{-\int_0^t h(\tau) d\tau} = e^{-H(t)} \quad (1.27)$$

où $H(t)$ est la fonction d'aléa cumulative, définie comme l'intégrale de la fonction d'aléa.

Supposons maintenant qu'entre deux groupes A et B, la fonction de survie du groupe A soit toujours supérieure à celle du groupe B. On peut en déduire que le groupe A a un meilleur pronostic de survie que le groupe B. On peut également dire que le groupe B est plus à risque que le groupe A, et donc que la fonction d'aléa du groupe B est supérieure à celle du groupe A.

Avant d'évoquer les approches non-paramétriques pour estimer la fonction de survie, évoquons quelques lois classiquement utilisées pour décrire le temps de survenue d'un événement d'intérêt. La loi la plus simple est la loi exponentielle de paramètre λ dont les fonctions de survie et d'aléa sont définies telles que

$$S(t) = e^{-\lambda t} \Leftrightarrow h(t) = \lambda \quad (1.28)$$

Comme évoqué précédemment, la loi exponentielle est une loi sans mémoire. Par conséquent, sa fonction d'aléa est constante puisque le risque que l'événement ait lieu est indépendant du temps de survenue de celui-ci.

Une autre loi couramment utilisée est la loi de Weibull. Elle peut être paramétrisée de multiples manières et nous utiliserons ici les paramètres

de forme k et d'échelle λ pour définir ses fonctions de survie et d'aléa telles que

$$S(t) = e^{-(t/\lambda)^k} \Leftrightarrow h(t) = \frac{k}{\lambda} \left(\frac{t}{\lambda}\right)^{k-1} \quad (1.29)$$

Si $k = 1$, on retrouve la loi exponentielle. Une propriété intéressante de la loi de Weibull est sa flexibilité, en particulier de forme de densité de probabilité, qui s'accompagne d'une facilité de manipulation puisque sa fonction d'aléa est strictement monotone. On peut évaluer rapidement (et graphiquement) la pertinence de l'utilisation d'une loi de Weibull pour l'analyse de données en représentant sur un même graphique la transformée $\ln(-\ln(S(t)))$ comme fonction de $\ln(t)$. En effet, pour une loi de Weibull, on a

$$\begin{aligned} \ln(-\ln(S(t))) &= \ln \left[\left(\frac{t}{\lambda}\right)^k \right] \\ &= k [\ln(t) - \ln(\lambda)] \end{aligned}$$

Autrement dit, si les données sont distribuées selon une loi de Weibull, la transformée $\ln(-\ln(S(t)))$ est linéairement distribuée en fonction de $\ln(t)$ (voir par exemple la figure A.4). Néanmoins, en pratique, l'information de la fonction de survie est indisponible et il faut l'estimer à l'aide des données accessibles. Pour cela, il existe plusieurs estimateurs non-paramétriques de la fonction de survie. Le plus connu est sans doute l'estimateur de Kaplan-Meier [Kaplan and Meier, 1958], noté ici \hat{S}_{KM} et défini tel que

$$\hat{S}_{KM}(t) = \prod_{i: t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \quad (1.30)$$

où t_i correspond aux temps, ordonnés par ordre croissant, pour lesquels au moins un événement a eu lieu, d_i au nombre de réalisations au temps t_i et n_i au nombre d'individus encore à risque au temps t_i^- , avant la i -ème réalisations. La quantité n_i correspond au nombre total d'événements moins le nombre d'événements réalisés avant t_i et le nombre d'individus censurés avant t_i .

Un autre estimateur classiquement utilisé pour la fonction de survie est celui de Nelson-Aalen [Nelson, 1969, 1972; Aalen, 1978]. À proprement parler, c'est un estimateur de la fonction d'aléa cumulative, c'est-à-dire l'intégrale de la fonction d'aléa. En reprenant les mêmes notations que précédemment et notant l'estimateur de la fonction d'aléa cumulative de Nelson-Aalen comme $\hat{H}_{NA}(t)$, on a

$$\hat{H}_{NA}(t) = \sum_{ti \leq t} \frac{d_i}{n_i} \quad (1.31)$$

Les estimateurs de Kaplan-Meier et Nelson-Aalen maximisent tous deux la vraisemblance empirique et donnent des résultats similaires en pratique [Zhou, 2015]. Les fonctions de survie issues de ces deux estimateurs sont continues à droite avec une limite à gauche (cadlåg), en forme d'escalier, contrairement aux fonctions de survie.

Test du *log-rank*

Pour comparer les deux courbes de survies de deux groupes A et B, on utilise classiquement le test du *log-rank* [Mantel, 1966; Peto and Peto, 1972]. L'hypothèse nulle de ce test est l'absence de différence entre les fonctions d'aléas des deux groupes. Mathématiquement, on peut écrire $H_0 : h_A(t) = h_B(t)$. Autrement dit, sous hypothèse nulle, on s'attend à ce que les nombres de réalisations de l'événement d'intérêt soient proportionnels entre les deux groupes A et B. Ainsi, si on note par $d_{A,i}$ (resp. $d_{B,i}$) le nombre de réalisations de l'événement dans le groupe A (resp. groupe B) au i -ème temps de réalisation (t_i) et $n_{A,i}$ (resp. $n_{B,i}$) le nombre d'événements encore susceptibles de se réaliser au temps t_i pour le groupe A (resp. groupe B), alors, sous hypothèse nulle, on devrait observer en moyenne pour le groupe $X \in \{A, B\}$ un nombre de réalisations $e_{X,i}$ tel que

$$e_{X,i} = \left(\frac{n_{X,i}}{n_{A,i} + n_{B,i}} \right) \times (d_{A,i} + d_{B,i}) \quad (1.32)$$

On peut ensuite comparer, pour tous les temps d'occurrence d'au moins un événement, la différence entre les réalisations observées ($d_{X,i}$) et les valeurs attendues ($e_{X,i}$), soit

$$O_X - E_X = \sum_i (d_{X,i} - e_{X,i}) \quad (1.33)$$

Notons que $O_A - E_A = -(O_B - E_B)$. Sous hypothèse nulle, on s'attend à ce que la statistique de test du *log-rank* suive une loi du χ^2 avec 1 degré de liberté, donc

$$\frac{(O_X - E_X)^2}{\text{var}(O_X - E_X)} \sim \chi^2 \quad (1.34)$$

avec

$$\text{var}(O_X - E_X) = \sum_i \frac{n_{A,i} n_{B,i} (d_{A,i} + d_{B,i}) (n_{A,i} + n_{B,i} - d_{A,i} - d_{B,i})}{(n_{A,i} + n_{B,i})^2 (n_{A,i} + n_{B,i} - 1)} \quad (1.35)$$

Il est possible de généraliser ce calcul à des situations de plus de deux groupes. Le raisonnement est similaire, mais le calcul de la statistique de résumé est plus sophistiqué [Kleinbaum and Klein, 2012]. Si on a G groupes à comparer, sous hypothèse nulle, la statistique de résumé suit une loi du χ^2 avec $G - 1$ degrés de liberté.

Régression de Cox à risque proportionnel

Bien que la fonction de survie $S(t)$ soit plus intuitive, il est plus facile de travailler avec la fonction d'aléa pour évaluer l'effet des covariables sur le temps avant la réalisation de l'événement d'intérêt. Bien souvent, il n'est pas judicieux d'utiliser des méthodes paramétriques, car la distribution sous-jacente aux données est inconnue et ne semble pas correspondre à une distribution classique. La force de la régression de Cox [1992] est la non-spécification de la distribution sous-jacente des temps de survie. Soit $\mathbf{X} = (X_1, X_2, \dots, X_p)$ un vecteur de variables explicatives. On peut alors écrire la fonction d'aléa telle que

$$h(t, \mathbf{X}) = h_0(t) \exp \left(\sum_{i=1}^p \beta_i X_i \right) \quad (1.36)$$

La fonction $h_0(t)$ correspond au risque de base, c'est-à-dire le risque lorsque toutes les covariables sont nulles. On parle ici d'un modèle à risque proportionnel car l'effet des covariables est indépendant du temps. Supposons que l'on ait deux individus ne différant que pour une covariable binaire (par exemple homme = 0 vs femme = 1), le rapport des fonctions d'aléa entre ces deux individus s'écrit

$$\frac{h(t, F)}{h(t, M)} = \frac{h_0(t) \exp(\beta_1 X_1, \dots, \beta_{\text{sexe}} \times 1, \dots, \beta_p X_p)}{h_0(t) \exp(\beta_1 X_1, \dots, \beta_{\text{sexe}} \times 0, \dots, \beta_p X_p)} = e^{\beta_{\text{sexe}}} \quad (1.37)$$

Le risque de base est une fonction quelconque. La régression de Cox est donc un modèle semi-paramétrique, à l'inverse des modèles paramétriques qui reprennent les fonctions d'aléas (par exemple les équations 1.28 et 1.29). Un gros avantage du modèle de Cox est sa robustesse. Si les données sont distribuées selon une loi de Weibull ou exponentielle, il donne des résultats comparables aux approches paramétriques.

Ainsi, lorsque la distribution des données n'est pas clairement établie, la régression de Cox permet d'extraire de l'information sans spécification du risque de base, et ce par maximisation de la vraisemblance partielle. On parle de vraisemblance partielle car on restreint le calcul aux données pour lesquelles la réalisation de l'événement est observée, ignorant donc les données censurées. Dans le cas le plus simple où à chaque temps t_i , une seule réalisation observée a lieu, la vraisemblance partielle pour l'événement i s'écrit

$$L_i(\beta) = \frac{h(t_i, \mathbf{X}^i)}{\sum_{j:t_j \geq t_i} h(t_j, \mathbf{X}^j)} = \frac{e^{\beta \mathbf{X}^i}}{\sum_{j:t_j \geq t_i} e^{\beta \mathbf{X}^j}} \quad (1.38)$$

avec \mathbf{X}^i les caractéristiques de l'individu dont l'événement se réalise au temps t_i . Lorsque plusieurs événements se réalisent à un même temps t_i , on utilise une formule plus compliquée de la vraisemblance qui pondère ces occurrences simultanées [Efron, 1977]. Le calcul de la vraisemblance

partielle s'obtient en multipliant les contributions pour les différents événements dont la réalisation n'est pas censurée :

$$L(\beta) = \prod_{k:C_k=1} L_k(\beta) \quad (1.39)$$

En pratique, on utilise rarement la valeur du maximum de vraisemblance β_k^* dans l'interprétation des résultats et on lui préfère la notion de *hazard ratio* $\exp(\beta_k^*)$. En effet, si l'on reprend l'exemple précédent 1.37, on dira que les femmes ont un risque instantané de décès $\exp(\beta_{\text{sexe}})$ fois celui des hommes. Un *hazard ratio* plus petit que 1 indique que la survie du groupe de comparaison est meilleure que celle du groupe de référence alors qu'un *hazard ratio* plus grand que 1 indique que la survie du groupe de comparaison est moins bonne que celle du groupe de référence.

On utilise les résidus de Schoenfeld pour évaluer l'hypothèse des risques proportionnels [Schoenfeld, 1982]. Pour chaque temps t_i où un événement a lieu, on calcule le résidu r_{ij} pour la variable j

$$r_{ij} = X_{ij} - \mathbb{E}[X_{ij}|R_i] \quad (1.40)$$

avec R_i les indices des individus encore à risques au temps t_i . Ces résidus correspondent à la différence entre les caractéristiques de l'individu dont l'évènement se réalise et une moyenne pondérée des individus encore à risques au temps t_i . Sous hypothèse de risques proportionnels, ces résidus sont centrés autour de 0 et distribués de la même manière en fonction du temps.

Dans certaines situations, il arrive que l'hypothèse de risques proportionnels soit incorrecte. Cela peut survenir si des covariables sont dépendantes du temps (par exemple la protection vaccinale peut décroître au cours du temps [Ferdinands et al.]). Des extensions de la régression classique de Cox ont été développées pour prendre en compte ces dépendances temporelles [Kleinbaum and Klein, 2012]. Mais la déviation de l'hypothèse de risque proportionnel peut aussi avoir lieu si l'on omet des variables importantes dans le modèle, potentiellement inconnues. Dans ces circonstances, l'utilisation de fragilités univariées (effets aléatoires) dans le modèle permet également de conduire des analyses avec un modèle de Cox à effets mixtes. De telles méthodes font l'hypothèse que le modèle de Cox complet, avec les variables connues et inconnues, vérifie l'hypothèse de risque proportionnel. Néanmoins, en pratique, sans information a priori, il n'est pas toujours facile de trancher sur les causes d'une déviation de l'hypothèse de risque proportionnel entre une dépendance temporelle et l'oubli de covariables cruciales. Les modèles à fragilités peuvent également être utilisés pour prendre en compte les dépendances dans les observations d'une même sous-population (individus d'une même famille, échantillonnage provenant d'un même centre) ou pour des événements récurrents. On utilise alors des modèles à fragilités partagées

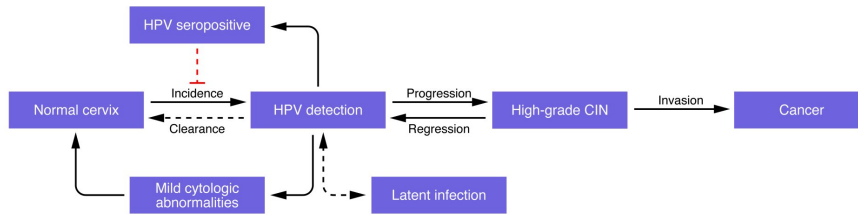


FIGURE 1.4 – **Diagramme de l’histoire naturelle des infections HPV et des cancers du col de l’utérus.** Tous les crédits pour la production du graphique reviennent à [Gravitt \[2011\]](#).

où les individus/observations d’un même cluster partageront les mêmes fragilités entre eux, mais différeront des individus des autres clusters.

1.3.3 Modèle multi-états et modèle de Markov caché

Dans le cadre d’infections chroniques, le virus persiste chez l’hôte et peut causer nombre de complications. Le suivi régulier des patients atteints de telles infections est essentiel pour évaluer la progression de la maladie et intervenir à temps en cas de complication.

Ces états de progression peuvent être définis en fonction de divers critères, relevant de la biologie médicale ou de la génétique par exemple. En particulier, dans le cadre des infections HPV et du cancer du col de l’utérus, plusieurs échelles de progression ont été développées, se concentrant sur des approches histologiques ou cytologiques (voir par exemple la figure 1.4). Les différents schémas partagent un paradigme d’une progression d’une infection aiguë vers un état chronique dont l’exposition prolongée peut induire la formation de lésions pré-cancéreuses voir de cancers [[Schiffman and Solomon, 2013](#)].

Du point de vue mathématique, on peut décrire un telle situation à l’aide d’une chaîne de Markov dont les états représentent les étapes de progression de la maladie. Comme décrit précédemment, les chaînes de Markov peuvent être décrites en temps discret ou continu. En pratique, on ne dispose que de quelques observations par patient qui renseignent sur l’état de la maladie au cours du temps. Les modèles multi-états sont des outils statistiques qui permettent d’estimer les transitions d’une chaîne de Markov à l’aide d’un nombre limité d’observations. Ils sont donc à l’interface entre les processus stochastiques et l’analyse de survie.

Les données ainsi collectées lors de suivis longitudinaux sont discrètes et en général assez limitées (rarement plus de 10 par patient·e dans le cas des suivis HPV). Une première approche peut être de considérer la progression à l’aide d’une chaîne de Markov discrète. Toutefois, cela pose rapidement des problèmes d’interprétation lorsque les données sont échantillonnées à intervalles de temps irréguliers. Nous verrons par la suite comment inférer la dynamique d’une chaîne de Markov en temps continu à l’aide de données discrètes, notamment grâce à la propriété de

Markov.

Néanmoins, en pratique, il peut s'avérer que l'état de la progression de la maladie ne soit pas accessible directement. Notamment dans le cas des infections HPV où il n'est pas envisageable de réaliser une biopsie des couches basales de l'épithélium cervical pour y évaluer la présence du virus. L'information nous est alors accessible via des proxys qui nous renseignent plus ou moins fidèlement sur le processus sous-jacent. Les tests de détection de génomes viraux en sont un exemple. En général, ce n'est pas tant la dynamique du proxy qui est intéressante biologiquement que celle du processus sous-jacent. Nous détaillerons les méthodes d'inférence des modèles de Markov cachés qui permettent de reconstruire l'information du processus sous-jacent à l'aide des observations indirectes du phénomène.

Inférence statistique d'une chaîne de Markov : le modèle multi-états

Lors de suivis longitudinaux, l'information sur l'état du processus ne nous est accessible qu'à travers un nombre fini de visites. La propriété de Markov est particulièrement intéressante dans ce contexte car elle permet de considérer chaque transition entre deux visites comme indépendante des autres transitions et indépendante du passé du participant (c'est-à-dire avant la date d'inclusion), information dont on ne dispose généralement pas. Dans les suivis longitudinaux, il arrive fréquemment que les données d'intérêt (progression de la maladie, statut infectieux) collectées soient discrètes. Une première approche peut être de décrire notre modèle de progression comme une chaîne de Markov discrète.

L'estimation des valeurs de la matrice P pour une chaîne de Markov discrète est assez directe à partir de suivis longitudinaux car cela revient à estimer des proportions pour des données de comptage. Soit $\Gamma = (\gamma_{i,j})_{i,j}$ l'estimateur de la matrice P . On peut calculer $\gamma_{i,j}$ comme le nombre de transitions de l'état i vers l'état j divisé par le nombre total de transitions observées depuis l'état i (les transitions de l'état i vers l'état i entre deux pas de temps sont incluses) [MacDonald, 2009]. Néanmoins, cette démarche repose sur l'hypothèse implicite qu'une seule transition a lieu entre deux observations, sans possibilité de transition non-observée. Ainsi, même si en pratique entre deux visites, le processus est passé de l'état i vers un état k , $k \neq i, j$, puis de l'état k vers l'état j , on englobe ces différentes transitions comme la seule transition de l'état i vers j [Wan et al., 2016]. De plus, il arrive couramment que la durée entre deux visites varie lors d'un suivi, limitant ainsi la pertinence d'un formalisme discret, les itérations n'ayant alors plus la même signification temporelle.

En pratique, l'emploi d'un formalisme continu est généralement plus approprié car les transitions entre états (par exemple l'acquisition d'une nouvelle mutation dans un modèle de progression cancéreuse [Armitage,

1985]) peuvent survenir à n'importe quel instant de la vie de l'individu. Comme évoqué précédemment, une chaîne de Markov en temps continu présente également une matrice de transition $P(t)$ (voir 1.10). Bien que généralement de moindre intérêt que le générateur Q en termes de modélisation d'un processus stochastique, la matrice $P(t)$ est cependant la seule information dont on dispose grâce aux suivis longitudinaux. Avant d'estimer les valeurs de $p_{i,j}(t)$ par maximum de vraisemblance, il est nécessaire de spécifier une structure à notre processus à travers une matrice Q donnée. En effet, comme évoqué au paragraphe précédent, l'information dont on dispose n'est qu'une collection d'instantanés à différentes visites. Or, entre ces deux visites, le processus a pu passer par des états intermédiaires non observés. Ainsi deux modèles aux structures différentes pourront exhiber une même collection d'instantanés, avec des implications biologiques différentes.

En particulier, de la relation 1.15 on peut montrer que

$$P(t) = \text{Exp}(Qt) \quad (1.41)$$

avec ici Exp la fonction exponentielle matricielle définie par développement en série entière de la fonction exponentielle scalaire mais appliquée aux matrices. Ainsi, soit X une matrice, alors $\text{Exp}(X) = 1 + X + X^2/2 + \dots$

Par exemple, considérons un modèle à trois états {cellule souche, cellule A , cellule B } où cellule A et cellule B sont deux types de cellules différentes. On cherche à décrire le temps de transition d'une cellule souche en l'une ou l'autre lignée. Dans un premier modèle, considérons qu'une fois que la cellule souche est passée sous la forme A ou B , elle ne peut plus se différencier en cellule souche. Les deux états A et B sont donc absorbants et on peut écrire la matrice de transition des taux comme :

$$Q = \begin{pmatrix} -(a+b) & a & b \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (1.42)$$

Par contre, si on fait l'hypothèse qu'à l'état A , la cellule peut se différencier en cellule B au taux c , mais que l'état B reste absorbant, alors on écrit la matrice Q comme

$$Q = \begin{pmatrix} -(a+b) & a & b \\ 0 & -c & c \\ 0 & 0 & 0 \end{pmatrix}, \quad (1.43)$$

On peut montrer que dans le premier cas $P_{2,3}(t) = 0$ alors que dans le second cas on a $P_{2,3}(t) = 1 - e^{-ct}$.

De manière générale, soit une matrice Q donnée, pour un individu donné et une paire d'observations successives entre deux états S_t et $S_{t'}$, la contribution à la vraisemblance de cet événement correspond à

$$L = p_{S_t, S_{t'}}(t' - t) \quad (1.44)$$

Dans le cas d'observations exactes d'événements (par exemple la mort d'un patient) ou de transitions (c'est-à-dire sans possibilité de transitions intermédiaires) la formule de la contribution de la vraisemblance peut être spécifiée plus précisément [Jackson, 2011a]. De la même manière que pour le modèle de Cox, il est possible d'évaluer l'effet de covariables sur les taux de transitions. Par analogie avec la fonction d'aléa, on peut écrire les éléments de la matrice de transition taux $Q = (q_{ij})$ comme :

$$q_{ij} = q_{ij}^0 \exp(\beta_{ij}^t \mathbf{X}) \quad (1.45)$$

avec \mathbf{X} les covariables d'intérêt [Marshall and Jones, 1995] et β^t la transposée du vecteur β .

Observation indirecte du processus et modèle de Markov caché

Toutefois, il arrive que l'information sur notre variable d'intérêt ne soit pas directement accessible du fait de limitations technologiques, pratiques, voire éthiques. Dans certains cas, on peut obtenir de l'information de manière indirecte par l'observation d'un processus intermédiaire qui dépend du processus caché. Les modèles permettant l'étude d'une variable dite cachée par l'intermédiaire d'observations émises depuis celle-ci sont les modèles de Markov cachés.

Formellement, un modèle de Markov caché consiste en une paire de variables aléatoires que l'on peut noter $(X_t, O_t)_{t \geq 0}$. Le processus sous-jacent, ou non-observé, ici noté $(X_t)_{t \geq 0}$, satisfait la propriété de Markov. Ainsi pour tous $t, s \geq 0$, la loi conditionnelle de X_{t+s} sachant la filtration $(\mathcal{F}_t^X)_{t \geq 0}$ est égale à la loi conditionnelle de X_{t+s} sachant X_t . Le processus $(O_t)_{t \geq 0}$ quant à lui ne satisfait pas a priori la propriété de Markov mais sa loi est dépendante de l'état du processus $(X_t)_{t \geq 0}$. Plus précisément, pour tout $t \geq 0$, la loi de O_t dépend que de la valeur de X_t [MacDonald, 2009].

La loi du processus $(O_t)_{t \geq 0}$ sachant $(X_t)_{t \geq 0}$ est généralement appelée loi d'émission du processus sous-jacent. Les deux processus ne sont pas forcément définis sur le même espace d'états (en particulier $(O_t)_{t \geq 0}$ peut être défini sur \mathbb{R}) et la loi d'émission du processus peut être quelconque. En particulier, soit $(X_t)_{t \geq 0}$ une chaîne de Markov à m états, il découle des relations énoncées que :

$$\mathbb{P}[O_t = j] = \sum_{i=1}^m \mathbb{P}[X_t = i] \mathbb{P}[O_t = j | X_t = i] \quad (1.46)$$

Ainsi on peut écrire la contribution de l'individu i à la vraisemblance comme :

$$\begin{aligned} L_i &= \mathbb{P}[O_{t_1}^i, \dots, O_{t_j}^i] \\ &= \mathbb{P}[O_{t_1}^i, \dots, O_{t_j}^i | X_{t_1}^i, \dots, X_{t_j}^i] \mathbb{P}[X_{t_1}^i, \dots, X_{t_j}^i] \end{aligned} \quad (1.47)$$

En reprenant la propriété de Markov sur le processus $(X_t)_{t \geq 0}$ et grâce à 1.46 [MacDonald and Zucchini, 1997] on a :

$$\begin{aligned}
L_i = & \left(\sum_{x_1=1}^m \mathbb{P}[X_{t_1}^i = x_1] \mathbb{P}[O_{t_1}^i | X_{t_1}^i = x_1] \right) \times \\
& \left(\sum_{x_2=1}^m \mathbb{P}[X_{t_2}^i = x_2 | X_{t_1}^i = x_1] \mathbb{P}[O_{t_2}^i | X_{t_2}^i = x_2] \right) \times \dots \quad (1.48) \\
& \times \left(\sum_{x_j=1}^m \mathbb{P}[X_{t_j}^i = x_j | X_{t_{j-1}}^i = x_{j-1}] \mathbb{P}[O_{t_j}^i | X_{t_j}^i = x_j] \right)
\end{aligned}$$

Le calcul du maximum de vraisemblance d'un tel modèle est ardu à réaliser directement. Dans de nombreux cas, des algorithmes d'espérance-maximisation sont employés pour estimer le maximum de vraisemblance [Dempster et al., 1977].

1.4 Objectifs et aperçu de la thèse

La modélisation mathématique offre des outils précieux pour la description et l'analyse de processus biologiques, la confrontation d'hypothèses concurrentes, voire pour proposer de nouveaux scénarios explicatifs. Sa richesse provient du cadre mathématique qui permet d'obtenir des résultats quantitatifs, mais également de la diversité des approches et des outils employés. Un modèle reste une mise en équation d'une version simplifiée, néanmoins la plus fidèle possible, d'une réalité biologique. Toutefois, si certaines hypothèses sont très commodes pour le formalisme mathématique, elles sont parfois assez loin de la réalité biologique. L'élaboration d'un modèle nécessite donc des compromis entre précision, réalisme et généralisation. De plus, les systèmes biologiques, tant à l'échelle microscopique que macroscopique, ne peuvent être décrits dans leur intégralité par un seul système mathématique. Combiner diverses approches permet d'apporter une vision d'ensemble du système, voire de confronter des pistes d'explications concurrentes.

Les trois premiers chapitres de ce manuscrit de thèse sont consacrés aux infections transitoires par les papillomavirus humains. En particulier, nous nous concentrons sur la dynamique de ces infections chez les jeunes femmes, avec une attention spécifique portée aux questions de latence et leurs implications sur la dynamique du système.

Le premier chapitre porte plus particulièrement sur la description de l'analyse de survie des infections HPV chez des jeunes femmes de l'étude clinique PAPCLEAR. Cette cohorte a suivi des femmes entre 18 et 25 ans de la région de Montpellier et se distingue des autres études cliniques sur HPV par un suivi sur le long terme et des échantillonnages réguliers. Combinée à un kit de détection simultané de 25 génotypes HPV,

cette étude clinique fournit d'amples données sur l'histoire naturelle des infections HPV. Dans ce premier chapitre, nous décrivons en détail les durées des phases de détection du virus et les durées entre ces phases à l'aide des outils de l'analyse de survie.

Dans le second chapitre sur les HPV, nous nous penchons plus en détails sur les implications de l'existence d'une phase latente à l'aide de modèles de Markov cachés. En particulier, il est possible que lorsque l'infection entre en phase latente, cela diminue la sensibilité des tests de détection, conduisant à une sous-détection de l'infection. Pour ce travail, nous avons combiné des jeux de données en ajoutant à l'étude clinique PAPCLEAR, deux cohortes vaccinales menées par GlaxoSmithKline sur l'innocuité du vaccin Cervarix™, pour ajuster nos chaînes de Markov cachées dans plusieurs scénarios biologiques.

Le dernier chapitre sur les HPV, évalue la possibilité de clairance des infections par la seule organisation du cycle de vie du virus. À travers un modèle combinant deux processus de branchement imbriqués, l'un décrivant la dynamique de renouvellement du tissu épithélial et l'autre la répartition des copies virales lors des divisions cellulaires, nous décrivons les pressions non-immunitaires que pose le cycle de vie du virus sur la persistance au long terme. Ce modèle permet de replacer le cycle de vie du virus dans un cadre évolutif où les mécanismes d'évasions immunitaires via une persistance en faible nombre dans les cellules basales sont contrebalancés par une susceptibilité aux fluctuations aléatoires.

Dans une moindre mesure, j'ai participé avec les membres de l'équipe à la description de la pandémie de SARS-CoV-2 lors des premiers mois de confinement. Cela a été l'occasion de changer de perspective pour travailler à l'échelle inter-hôte. Nous avons proposé un modèle permettant la prise en compte de fluctuations stochastiques dans les premiers et derniers stades, phases durant laquelle la dynamique épidémique est le fruit de quelques individus infectieux. En particulier, les hypothèses déterministes des modèles classiquement utilisés en épidémiologie (SIR et dérivés) sont en général trop simplistes pour capter l'essence de la dynamique infectieuse dans ces périodes de fortes pressions stochastiques.

Le chapitre final qui clôturé la section principale de ce manuscrit offre une occasion de contextualiser les travaux de recherche menés durant ma thèse par rapport aux préoccupations de leurs communautés scientifiques respectives et propose plus généralement une réflexion sur le rôle du modélisateur dans la prise de décision en santé publique, en particulier dans un contexte épidémique majeur comme la pandémie de COVID-19.

CHAPITRE

2

Caractérisation des infections transitoires HPV chez les jeunes femmes



2.1 Aperçu de l'article

Les infections HPV sont très courantes, et l'exposition au virus est maximale au début de la vie sexuelle [Schiffman and Solomon, 2013]. Toutefois, la majorité des infections acquises jeunes ne persiste pas et on ne détecte plus leur présence au bout de trois ans. Pour HPV16 en particulier, il n'est plus détecté à un an dans 41,3 % des cas, 72,3 % à deux ans et 93,1 % à trois ans [Insinga et al., 2007]. Classiquement, on suppose que la perte de détection est associée à l'élimination du virus de l'hôte et la détection d'un génotype à une acquisition récente du virus. Cette vision binaire a été nuancée par la possibilité de persistance des infections HPV à des niveaux non-détectables, aussi appelés état latent [Gravitt, 2012; Gravitt and Winer, 2017; Doorbar, 2023]. Cette phase latente a été mise en avant dans un modèle animal, mais les preuves manquent chez l'humain. Les indices les plus convaincants concernent des populations spécifiques (femmes vivant avec le VIH, femmes abstinentes) mais pas la population générale des jeunes adultes. Si cette latence est mentionnée dans ce chapitre, le suivant contient une analyse plus fournie. Dans ce premier travail, nous chercherons à mieux comprendre les motifs de détection chez des jeunes femmes infectées, sans essayer d'attribuer une origine aux nouvelles détections.

Comprendre les mécanismes qui conduisent une infection à persister est crucial en termes de santé publique [Doorbar, 2023]. Selon le contexte actuel, l'établissement d'une infection chronique HPV est un pré-requis nécessaire à la survenue d'un cancer du col de l'utérus. Toutefois, il reste encore des zones d'ombres sur la différence de potentiel oncogénique entre les infections chroniques et les infections latentes. De manière générale, une meilleure compréhension des infections transitoires et des motifs de détection HPV chez les jeunes femmes est capitale pour l'aiguillage des politiques de prévention et l'identification de populations plus à risques de développer des cancers du col de l'utérus.

L'étude clinique PAPCLEAR a été conduite entre 2016 et 2020 au CHU de Montpellier et a enrôlé 190 femmes dont 149 ont été suivies longitudinalement. Elle a été conçue pour maximiser le suivi des infections depuis leur initiation jusqu'à leur clairance via des échantillonnages fréquents. Le protocole est détaillé ailleurs [Murall et al., 2019b]. Brièvement, le dépistage des HPV commence par une étape d'amplification par PCR d'un segment de 65 nucléotides du gène L1. La présence d'alphapapillomavirus est détectée par hybridation DEIA. Les échantillons positifs sont ensuite génotypés par hybridation inverse LiPA, qui permet la détection simultanée de 25 génotypes [Kleter et al., 1998a, 1999; Quint et al., 2001; van Doorn et al., 2002]. Dans l'éventualité où le kit SPF10-LiPA25 retourne un résultat négatif malgré un DEIA positif, une amplification PCR PGMY et un séquençage Sanger [Coutlée et al., 2002] sont employés pour tenter de déterminer le génotype HPV. Ce kit de dépistage est plus sensible que la majorité de ceux utilisés en routine dans

les tests de dépistage [Geraets et al., 2015b].

Nos connaissances de l'infection naturelle ont beaucoup progressé grâce à l'analyse des bras contrôle des études vaccinales. Ces études portaient sur l'étude de la dynamique au long terme des infections chroniques et de l'effet protecteur du vaccin contre ces infections. Cela se traduit dans le design expérimental par une fréquence d'échantillonnage généralement comprise entre trois et six mois [Insinga et al., 2007; Ramanakumar et al., 2016; Paul et al., 2021; Malagón et al., 2023]. Un des avantages de l'étude clinique PAPCLEAR est le suivi régulier et fréquent des participantes (voir Figure A.1). Certaines études font exception et se sont penchées sur la dynamique temporelle fine des infections avec des prélèvements bi-hebdomadaires, mais la visée principale de ces cohortes était l'étude du microbiote vaginal [Brotman et al., 2014; Liu et al., 2020a]. De plus, dans ce dernier cas, le nombre d'individus est relativement restreint (moins de 50 participantes incluses dans le suivi) et la durée de suivi individuel inférieure à celle de PAPCLEAR.

Nous avons analysé ces suivis bimestriels à l'aide d'outils classiques de l'analyse de survie (estimateur de Nelson-Aalen, régression de Cox), la donnée d'observation étant ici à l'échelle du génotype par patiente. Nous avons pris en compte les potentielles corrélations entre génotypes détectés chez un même individu par inclusion de fragilité à l'échelle du patient dans le modèle de Cox [Wienke, 2010].

En premier lieu, nous avons analysé le temps de détection des génotypes HPV, c'est-à-dire la durée des épisodes positifs à un génotype donné. Nous avons également étudié le temps avant la survenue d'une infection incidente pour un génotype donné, ainsi que la durée d'absence de détection entre deux épisodes positifs consécutifs.

Dans notre étude, chaque visite clinique est associée à une information binaire concernant la détection ou non du génotype chez l'individu. La date de début (et de fin) des événements d'intérêt est inconnue, mais on connaît l'intervalle durant lequel l'événement commence (et termine). Ainsi, formellement, nos quantités d'intérêt sont des variables à intervalles doublement censurés. De telles données sont difficiles à manier et nous les avons simplifiées en données uniquement censurées à droites. Pour ce faire, nous avons défini les instants de début et fin de positivité aux points médians entre la visite négative et positive. Le biais introduit par cette imputation reste toutefois limité [Law and Brookmeyer, 1992].

Nos résultats suggèrent que les jeunes femmes participant à l'étude PAPCLEAR expérimentent de nombreux épisodes de positivité, avec des co-détectations fréquentes de plusieurs génotypes, généralement courts. Deux facteurs apparaissent positivement corrélés à la durée des épisodes : les difficultés financières et le fait de se déclarer uniquement comme « caucasienne » dans le questionnaire d'inclusion.

Les participantes déclarant à la visite d'inclusion avoir eu plus de trois partenaires sexuel·le·s au cours de leur vie ont des temps entre épisodes consécutifs plus courts que celles qui ont signalé n'en avoir eu qu'un

ou deux. Ceci s'explique peut-être par le fait que les participantes déclarant un plus grand nombre de partenaires sexuel·e·s sont également celles ayant déclaré le plus de nouveaux partenaires sexuel·le·s durant le suivi. De manière générale, on observe fréquemment des épisodes de re-détection d'un même génotype au cours du suivi. La vaccination est associée à une diminution du risque de détection des génotypes HPV, mais nous manquons de puissance statistique pour établir si cette différence est significative ou non.

Nos résultats sur les dynamiques entre les génotypes à haut risque et ceux à bas risque sont cohérents avec la littérature : les infections sont détectables plus longtemps pour les génotypes à haut risque [Goodman et al., 2008] et la durée entre deux épisodes est plus courte pour les génotypes à haut risque [Ramanakumar et al., 2016] comparativement aux génotypes à bas risque.

Enfin, nos résultats suggèrent que les infections incidentes observées (que ce soit des premières détections ou des re-détections) ont pour origine probable une nouvelle acquisition plutôt que la ré-émergence d'infections latentes. L'analyse des auto-prélèvements hebdomadaires réalisés par les participantes à la maison, en particulier entre deux épisodes consécutifs séparés par une seule visite négative, pourrait aider à approfondir ce point.

2.2 Article

HPV detection patterns in young women from the PAPCLEAR longitudinal study : implications for HPV screening policies.

Auteurs: *Thomas BENETEAU*, Soraya Groc, Carmen Lia Murall, Vanina Boué, Baptiste Elie, Nicolas Tessandier, Claire Bernat, Marine Bonneau, Vincent Foulongne, Christelle Graf, Sophie Grasset, Massilva Rahmoun, Michel Segondy, Vincent Tribout, Jacques Reynes, Christian Selinger, Nathalie Boulle, Ignacio G. Bravo, Mircea T. Sofonea, Samuel Alizon

DOI: <https://doi.org/10.1101/2023.09.30.23296382>

Abstract

Objectives. HPV infections are ubiquitous. For most infections, we lose track of the presence of the virus in host in less than three years after the start of infection. The mechanisms regulating the persistence of HPV infection are still partially understood. In this work, we focus on incident HPV detection in young women and we characterise the dynamics of these infections and evaluate the effect of genotype and host socio-economic factors on the duration of HPV detection and time between detection.

Methods. We investigated human papillomavirus (HPV) genotype detection patterns in 182 young women in Montpellier, France. We relied on SPF₁₀-LiPA25 screening assay for the simultaneous detection of 25 HPV genotypes. We used survival analysis tools with frailty effects to investigate the contribution of viral and host factors to variations in the time of HPV detectability, time of first incident detection, and time before re-detection.

Results. Women of the PAPCLEAR cohort experienced numerous positive HPV events, including frequent redetection of the same genotype. We retrieve classical results that HR-genotypes are detected for longer duration than LR-genotypes. HR-genotypes were also more likely to be detected than LR-genotypes during the follow-up. The number of lifetime sexual partner was strongly associated with increased risk of new positive detection while vaccination was related to a lower risk of displaying incident infections. Covariates related to socio-economic difficulties were associated with longer duration of HPV positivity.

Conclusions. Young women display numerous event of HPV detection, with frequent codetections of multiple genotypes at the same time and redetection of the same type after periods of no detection. These new detection are almost certainly the result of new acquisition from sexual

partners, with little evidence of re-emergence of latent infections. A better characterisation of transient infections might help unveil doubts and misconception on HPV physiopathology, favouring adherence to preventive policies.

Introduction

Human Papillomaviruses (HPV) are the most oncogenic viruses known to infect humans, accounting for more than 600,000 deaths worldwide each year [Plummer et al., 2011]. They are also one of the most common sexually transmitted infections, with estimates suggesting that by 45yo, more than 80% of the people are or have already been infected by an HPV [Chesson et al., 2014b]. It is generally accepted that the initial HPV infection is acquired during the first sexual exposures, with the prevalence peaking after sexual debut, and that the risk of contracting a new HPV infection increases with the number of sexual partners [Bruni et al., 2010]. HPV presence generally goes undetected within the first three years, an event generally referred as HPV clearance [Schiffman et al., 2011b]. This clearance, however, may not necessarily imply true immune clearance. The interpretation of re-detection events is delicate as they might originate from various sources : true new infection, transient deposition from a sexual partner, or detection of latent infection [Gravitt and Winer, 2017]. Most certainly, HPV detection is a combination of these different pathways. Deciphering the cause of new HPV detection is still a major challenge. Answering this question is crucial in the optimisation of future public health policies, to evaluate the effectiveness of catch-up vaccination or organise the age stratification of vaccination policies. Longitudinal studies are valuable data both in terms of density and length of follow-up. Extracting full potential of such raw material is challenging and require the use of rigorous statistical tools. In this work, we used the PAPCLEAR cohort which of samples collected every 8 weeks in 132 young women, aged 18-25, for which we test the presence of 25 HPV genotypes using the SPF10-LiPA25 technique [Geraets et al., 2015a]. In particular, we evaluated viral genotypes and host factors involved with attention to frailty effects at the patient level and accounting for the censoring in the data to maximize the quality of the analysis.

Materials and methods

Study design and participants

The PAPCLEAR cohort has been detailed elsewhere [Murall et al., 2019b]. In short, this monocentric longitudinal study included 189 women, who were between 18 and 25 years old at inclusion, lived in the area of Montpellier (France) and reported having at least one new sexual partner over the last 12 months. Women with a history of HPV-associated pathology were excluded from the study. Pregnant women or women who were planning a pregnancy within the first year of inclusion were also excluded from the study. A graphic summarising the inclusion protocol can be found in the Supplementary materials A.4. A total of 150 women were followed longitudinally for up to 2 years between 2016 and 2020. The additional 39 participants were part of a cross-sectional analysis, that was prematurely suspended due to the pandemic. On-site visits of infected participants took place every 8 weeks with a gynaecologist or a midwife, who performed a cervical smear. Except for inclusion, participants were told to avoid sexual contact the day before the visit took place to avoid unwanted transient sexual deposition from the partner. At inclusion, the participants self-completed an extensive questionnaire related to demographic, socioeconomic, and behavioural risk factors. For the next visits, the participants also filled in follow-up questionnaires to notify changes in their habits. All participants provided written informed consent.

Genotyping

We first tested for the presence of alpha papillomaviruses in the cervical smears using the DEIA assay [Kleter et al., 1998b]. DEIA-positive samples were genotyped using the LiPA25 assay, which was chosen for its sensitivity and can detect up to 25 different HPV genotypes [Geraets et al., 2015a]. Among these, we refer to high-risk (HR) genotypes for HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68 [IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2012] and to low-risk (LR) for the remaining 12 (HPV6, 11, 34, 40, 42, 43, 44, 53, 54, 70, 74). If the LiPA25 test was negative, the genotype was determined using the PGMY PCR amplification [Coutlée et al., 2002] and Sanger sequencing of the PCR product. If the sequencing did not yield a clear result, samples were labelled as 'non-typable'.

Statistical analyses

All statistical analyses were conducted using R 4.2.2 with additional packages listed in Supplementary materials A.1.

We excluded all the women with less than three visits, meaning all the cross-sectional group and 18 participants from the longitudinal group (1 participant quit the study, 4 were seen once, and 13 were seen only

twice). All analyses were genotype-specific, with the unit of observation being the HPV genotype. Therefore, each participant could contribute to multiple observations. Following earlier studies, we assumed the dynamics of each genotype to be independent at the participant level and between participants [Dillner et al., 2011]. If not specified, the results were pooled across all genotypes.

As used in previous works [Malagón et al., 2023], we defined an HPV genotype as ‘prevalent’ if detected at inclusion. We also defined a genotype as ‘incident’ if detected at posterior visits but not at inclusion. Patterns of positive detection separated by only one negative visit, sometimes referred to as ‘intermittent detection’ [Malagón et al., 2023], are handled differently between studies and there does not appear to be consensus on the way to deal with such data. In the main analysis, we considered the two episodes as separated but we also conducted analysis with merged intermittent patterns (Supplementary materials A.1).

Model selection was performed using the corrected Akaike Information Criterion (AICc) as a metric for the penalised goodness of fit [Hurvich and Tsai, 1993]. We evaluated the goodness of fit for all sub-models of the maximum model (i.e. with all the covariates) and estimated the hazard ratios of the Cox regression using a full averaging procedure on the best models. Thorough details are available in the Supplementary materials A.1. To test for differences between the two populations, we used Fisher’s exact test for qualitative variables and Wilcoxon-Mann-Whitney’s test for quantitative variables. We displayed the raw p -values in the results Tables and Figures. In the following, a p -value < 0.05 is considered significant.

Survival analyses

For each genotype and each participant, we analysed the duration of HPV detectability, the duration between HPV-positive episodes, and the time to incident HPV infection. Survival functions for these quantities were computed using the non-parametric Nelson-Aalen estimator [Nelson, 1972; Aalen, 1978]. For the time until the first incident detection, we fitted a Weibull distribution to the survival function to predict the cumulative risk of incident detection at 5 years since inclusion (see the Supplementary materials A.1 for more details).

For a given episode, we defined the time of HPV detectability as the duration between the midpoint at the start of the episode and the midpoint at the end of the episode. We included all incident episodes, even the shortest episodes that were only detected during one visit, elsewhere referred to ‘transient’ infections, but hereafter called ‘singletons’, and the right-censored observations. The latter corresponds to patients who tested positive for HPV at their last scheduled visit. We also included prevalent episodes whose start is unknown and for which the duration of HPV detectability is right-censored. Similarly, the time between conse-

cutive positive episodes was computed as the duration between the midpoint at the end of the expired episodes and the midpoint at the start of the new episodes. The time until the first incident infection here corresponds to the time from inclusion to the midpoint at the start of the first incident detection. Extensive information can be found in Supplementary Methods A.1.

We checked for differences in HPV detectability and time to first incident detection between HR-genotypes and LR-genotypes using log-rank tests [Mantel, 1966; Peto and Peto, 1972]. To evaluate the effect of non-viral variables, we used Cox proportional hazards models [Cox, 1992]. We stratified the Cox regression with different baseline hazard functions for genotypes not detected, first detection, first redetection, and second redetection. We assumed no interaction between the strata. For all Cox regression analyses, we checked the validity of the proportional hazards (PH) assumption using Schoenfeld’s residuals [Schoenfeld, 1982]. The covariates included in the analysis are the number of lifetime sexual partners (LTSP), the BMI at baseline, the self-declared ethnic origin (Caucasian vs. non-caucasian or mixed-origin), the HPV vaccination status, the sexual affinity, the use of condom or contraceptive pills, an indicator of financial difficulties (participant had to decline medical care because of financial reasons), the number of years between inclusion and first menstruation, the number of years between inclusion and first sexual intercourse and the smoking stats (past, current or never). The numbers for each category can be found in Table 2.1.

We considered here that the unit of observation was the HPV genotype at the patient level. Thus as one participant would experience codetection of multiple genotypes, this could induce some correlations between the observations. To account for correlations between observations of the same cluster (i.e. a participant), we add shared frailty effects at the patient level in the Cox regression [Wienke, 2010]. We tested for the relevance of adding the frailty at the patient level using a likelihood ratio test with one degree of freedom between the two models (with and without the frailty effect).

Results

Descriptive analysis

The 150 participants of the PAPCLEAR cohort came for 6 visits on average (Poisson 95% CI : 5.63 - 6.43). For the 132 women included in the analysis, the follow-up duration was on average 311 days (IQR : 182 - 431), this accounted for a total of 1543 months of follow-up. The PAPCLEAR participants included in the analysis were on average aged 21.3yo (IQR : 20 - 23) at inclusion and around half of them (62/132) were vaccinated against HPV (56 with Gardasil and 6 with Cervarix). Baseline characteristics for the 132 women included can be found in Table 2.1.

In about 74% of the women included in the analysis (98/132), we detected at least one episode during their follow-up and 47% (62/132) experienced codetections (i.e. the simultaneous detections by more than one genotype). Overall, we detected 342 distinct episodes, 186 incident (including 137 first detections, 44 redetections and 5 second redetections) and 156 prevalent, including 211 from HR types and 114 from LR types. For 17 episodes, we could not determine the genotype detected. The three most frequent detected types in descending order were HPV51, HPV53, and HPV66, in agreement with previous results on the PAP-CLEAR cohort [Murall et al., 2019b]. A total of 83 (62.9%) participants were positively detected for at least one HPV genotype.

An average of 2.6 (Poisson 95% CI : 2.32-2.88) episodes per woman were detected during the whole follow-up, which yielded an average attack rate of 2.99 HPV episodes detected per person-year.

Detection of first incident HPV episode

Overall, we detected 137 first incident detection for all detectable genotypes and all participants. At one year, we estimated a 4.80% (log-log 95% CI : 3.99 - 5.77) cumulative proportion of first incident HPV detection pooled across all genotypes. After two years of follow-up, the proportion of first incident infection detection increased to 7.26% (log-log 95% CI : 5.84 - 9.02). To assess the variation after 5 years, we used the Weibull fit and predicted the proportion to reach 16.56% (95% CI : 7.37 - 29.73). Information regarding the Weibull fit can be found in the Supplementary materials A.1.

We assessed the rate of incident detection by oncogenic risk. We found that HR-HPV were more likely to be detected than LR-HPV over the whole follow-up (log-rank p-value < 0.01). We, however, lacked statistical power to verify that the difference in survival functions between HR genotypes and LR genotypes was consistent for redetection or second redetection.

Risk factors for the time between consecutive detection

In addition to the 137 observed first incident detection (3271 right-censored), we detected 156 prevalent episodes, 44 observed redetection (213 right-censored) and 5 second redetection (36 right-censored). Among the redetection, 33 consecutive episodes were only separated by one negative visit, pattern elsewhere referred to as ‘intermittent’ detection [Malagón et al., 2023]. Compared to participants reporting 1 or 2 LTSP, women who reported 3-10 LTSP had increased risk of experiencing new detection (hazard ratio : 2.40; 95% CI : 1.07 - 5.39), first incident or redetection. We observed a similar trend for participants reporting more than 10 LTSP, compared to the reference of 1 or 2 LTSP. However, we lacked

TABLE 2.1 – **Baseline description of the PAPCLEAR cohort for participants included in the analysis (> 2 visits).** Except for the vaccine used not included in the analysis, the first level display for all categorical variables is the reference level used in the Cox analysis. Missing observations were removed from the analysis.

Covariates	States/mean [IQR]	# women	% women
Age at inclusion	21.27 [20 - 24]	132	100
Age of first menstruation	12.85 [12 - 14]	131	99.24
Years between 1st menstruation and inclusion	8.42 [7 - 10]	131	99.24
Age of first sexual intercourse	16.39 [15 - 17]	132	100
Years between 1st intercourse and inclusion	4.89 [3 - 7]	132	100
Vaccination	Unvaccinated	66	50.00
	Vaccinated	66	50.00
Vaccine used	Cervarix	6	9.09
	Gardasil	56	84.18
	Non-specified	4	6.06
Self-declared ethnicity	Mixed or other	27	20.45
	Caucasian	105	79.55
Self-declared sexual affinity	Bi-/Homosexual	12	9.09
	Heterosexual	120	90.90
Smoking	Never	61	46.21
	Past	19	14.39
	Current	52	39.39
Contraceptive pills †	Not using	64	48.48
	User	68	51.51
Male condoms	Not using	56	42.42
	User	76	57.57
Number of lifetime sexual partner	1 ;2	20	15.15
	3 ;10	63	47.72
	11+	48	36.36
	missing	1	0.76
Financial difficulties *	No experience	117	88.64
	Experienced in the last 12mo	15	11.36

† emergency pills not included

* defined as a participant who declined medical care because of financial reasons

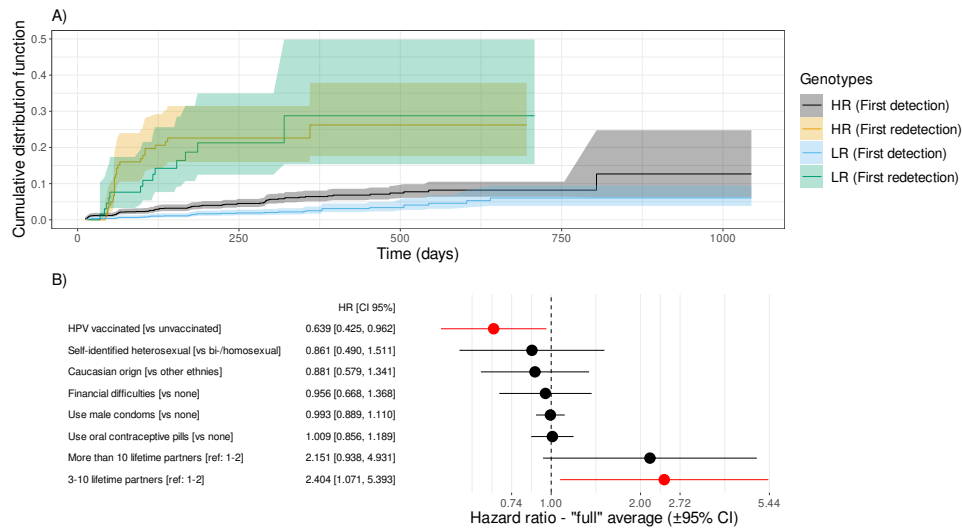


FIGURE 2.1 – Cumulative distribution functions for the time to first incident detection and time to first genotype redetection stratified by HR/LR genotypes and effects of host covariates on these estimates. A) Cumulative distribution function (CDF) of the time to first incident HPV detection since inclusion and the time to the first redetection, stratified by HPV genotype status (HR and LR). B) Hazard ratio for the best models selected by Cox regression with frailty at the patient level. Significant covariates are in red and hazard ratios greater than 1 indicate the covariate is associated with an increased risk of detection, hence lower duration between consecutive episodes.

statistical power to report significant association (hazard ratio : 2.15; 95% CI :0.94 - 4.93). Merging intermittent patterns (Figure A.2) yielded similar results, this time with a significant association for the group reporting more than 10 LTSP. We also found that vaccinated participants were less likely to display new incident detection or redetection compared to unvaccinated participants (hazard ratio : 0.64; 95% CI : 0.43-0.96). Thorough results of the Cox analysis for the time between consecutive episodes are displayed in Figure 2.1.

Loss of HPV detection

On the total of 342 detected episodes, 156 were prevalent HPV detection and for 40 episodes we did not observe the loss of detectability. For 17 episodes, the participants entered positive to a genotype and left the follow-up still positive for that same genotype, without any negative visit in-between. The majority of the episodes were positive for only one visit (198/342; 57.9%), but a significant proportion of them are censored observations (86/198; 43.4%) and thus potentially just a glimpse of a longer event. We estimated a median period of HPV detectability of 113 days (log-log 95% CI : 92.5 - 124). Our results suggest that around 23.5%

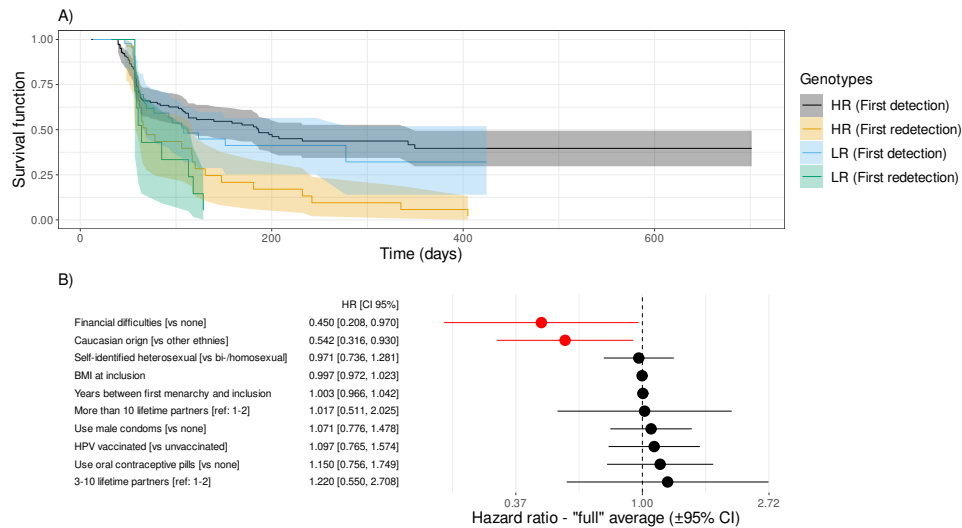


FIGURE 2.2 – Survival function for the time of HPV detectability stratified by HR/LR genotypes and effects of host covariates on this estimate. A) Survival functions stratified by the genotype (HR/LR) for the time to loss of HPV detection. B) Hazard ratios for the host factors. Significant factors are in red and a hazard ratio lower than one indicates that the trait is associated with a decreased rate of loss of HPV DNA detection, hence longer survival functions. The reference level is indicated in the bracket for the qualitative variables (see Methods for details).

(log-log 95% CI : 16.5 - 31.4) of the HPV episodes were still detectable after 700 day of follow-up. We found that HR-HPV types were detected significantly longer than LR-HPV infections (log-rank p-value < 0.05), the survival functions are displayed in Figure 2.2. The median duration of detectability was 130 days (log -log 95% CI 106 - 186) for HR genotypes and 96 days (log-log 95% CI 67.5 - 113) for LR genotypes.

Finally, we investigated the effect of key host factors (listed in Table 2.1) on the duration of HPV detectability using LiPA assays. Our analysis showed that HPV was detected for a significantly longer duration in participants who experienced financial difficulties (defined as a participant who declined medical care because of financial reasons) in the last 12mo before inclusion compared to participants who did not experience it (Hazard ratio : 0.45; 95% CI : 0.21 - 0.97). Infections were also detected longer in participants who identified themselves solely as 'Caucasians' (Hazard ratio : 0.54; 95% CI : 0.32 - 0.96) compared to participants who identified themselves as non-Caucasian or mixed-origin (Figure 2.2). These results were similar when merging intermittent patterns (Figure A.3).

Discussion

In this work, we analysed the HPV detection patterns of 132 young women from the PAPCLEAR cohort. We estimated a cumulative probability of first incident HPV detection since baseline, pooled across all genotypes, of 4.80% (log-log 95% CI : 3.99 - 5.77) at one year and predicted it to reach 16.56% (95% CI : 7.37 - 29.73) at five year. While we used convenient denominations such as ‘first incident detection’ or ‘first redetection’ following previous works, these expressions surely not bare the same relevance in terms of natural history. It is very unlikely for most of the investigated participants that a first detection during the follow-up corresponds to a true first exposure, only 11.36% participants declared having had their first sexual partner in the last 12 months. Besides, these definitions are dependent of the sampling frequency, as the probability of documenting a short period between redetections increases with the sampling rate. Thus, it is not straightforward to compare our results with other studies. Our results for first detection are very similar to results for first redetection but quite discordant with estimates for first incident detection from other studies [Malagón et al., 2023]. Those differences might first come from the difference in the cohort design, as the sampling rate for the Ludwig-McGill cohort was about every 6 months. Besides, the two populations were quite different. Participants here are all between 18-25 years old at inclusion. In the latter, there is a wider age diversity among the women, with around 80% of the women being older than 25 years old. Sexual activity is negatively correlated with age after sexual debut, thus our population might be more exposed to HPV due to increased sexual activity [Herbenick et al., 2010]. Additionally, the Ludwig-McGill cohort was sampled from low-income families from Brazil, while we included women without income criteria.

We found that the participants experienced numerous detected episodes (2.99 per women-year), most of them being positively detected for only one visit (198/374 singletons, 86/198 censored). In a quarter of the participants included in the analysis (34/132), no alphapapillomaviruses was detected, while for 18 women we detected more than 5 episodes during the follow-up. Redetection, were not uncommon as in about a third of women displaying at least one HPV positive visit we detected redetections of the same genotype. The frequency of codetections is similarly high (47%) and is consistent with previous studies [Ramanakumar et al., 2016].

To date, except for some specific populations (e.g. abstinent women [Paul et al., 2021]), it is quite difficult to settle on the origin of new HPV detection in humans. In our cohort, the number of lifetime partners was negatively associated with the duration between episodes. We found that in participants reporting a number of lifetime partners higher than 3, the time between episodes was significantly shorter than for women with 1 or 2 reported LTSP at inclusion, with little difference between participants

reporting 3 to 10 LTSP and those reporting more than 11 LTSP. The number of LTSP was not associated with the number of years between first intercourse and inclusion. Besides, women reporting a higher number of LTSP at inclusion were also more likely to report intercourse with new partner during the follow-up. Taken together, our results suggest that new detection and redetection observed here are more likely to be new acquisition than re-emergence of latent infection. Additionally, we recall that only women who declared having sex with a new partner in the last 12 months were included in the cohort [Mural et al., 2019b]. It was noted elsewhere that in the setting of new sexual partners, true incident infection was the preferential explanation [Trottier et al., 2010; Moscicki et al., 2013]. Compared to unvaccinated participants were less at risk of displaying incident infections. These results are consistent with observations in a dozen country [Brotherton, 2019].

We found that HR-HPV types were more likely to be detected for a longer period than LR-HPV types, which corroborates earlier studies [Goodman et al., 2008]. Conversely, we found that the time between HPV positive events was shorter for HR-HPV types compared to LR-HPV types, consistent with other work [Ramanakumar et al., 2016].

The mean duration of HPV detectability was globally lower than previous studies [Goodman et al., 2008; Ramanakumar et al., 2016; Paul et al., 2021; Malagón et al., 2023]. This can be partially explained by the difference in sampling rates with compared to studies (8 weeks compared to 6 months) and the inclusion of all positive episodes, including the singletons, sometimes excluded from analysis [Paul et al., 2021]. Participants that experienced financial difficulties (defined as a participant who declined medical care because of financial reasons) in past 12 months prior to inclusion displayed longer periods of HPV detectability compared to participants that did not declared suffering from financial struggles. People in situations of poverty generally tend to live in areas with low medical coverage, thus also limiting their access to medical care [Vallée et al., 2010]. While there was no significant difference in vaccination uptake between participants who faced financial difficulties and those who did not, taken together, our results suggest that people with financial difficulties might be less prone to seek medical guidance, especially since specialists are not fully reimbursed in France, thus putting their more at risk of genital infections and complications. Our results also suggest that self-declared Caucasian participants experienced longer periods of HPV detection. However, we lacked information to assess if that trend originated from genetic origin or socio-demographic/behavioural differences between the two groups. Besides, our population is relatively limited in number (132 women, 27 mixed origin or non-Caucasian origin, 105 Caucasian origin) and restricted to a specific region in France. It does not reflect the French population diversity, and thus might just be the results of selection bias.

Clarifying the dynamic of HPV infection, especially regarding the

distinction between re-detection and new acquisition is decisive to inform public health policies. Efficient screening policies and prevention have been implemented to limit progression towards cancer with good compliance to these measures [Jansen et al., 2020]. While most HPV infections are generally benign, testing positive during HPV screening can cause psychological stress and anxiety [McBride et al., 2021], especially if self-sampling becomes widespread [Nishimura et al., 2021].

We believe a better characterisation of HPV infections, especially regarding the link between infection status and detection data, will help unveil doubts and misconception on HPV physiopathology, favouring adherence to preventive measures [Thompson et al., 2016].

Acknowledgements

The authors thank all the participants of the PAPCLEAR study and the clinical staff and nurses for their help.

Financial support

TB is funded by la *Ligue contre le Cancer* (grant No TAKX21133). This work was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 648963 to SA). The sponsor had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

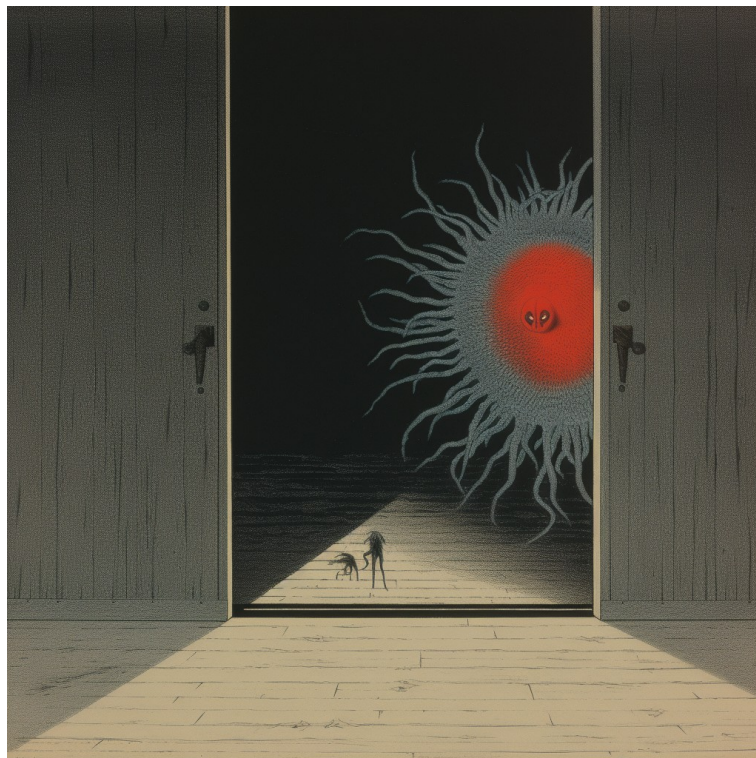
Declaration of Competing Interest

J. R. reports personal fees from Gilead (consulting and payment or honoraria for lectures, presentations, speaker’s bureaus, manuscript writing, or educational events), Janssen (payment or honoraria for lectures, presentations, speaker’s bureaus, manuscript writing, or educational events), Merck (payment or honoraria for lectures, presentations, speaker’s bureaus, manuscript writing, or educational events), Thera-technologies (payment or honoraria for lectures, presentations, speaker’s bureaus, manuscript writing, or educational events), and ViiV Healthcare (consulting and payment or honoraria for lectures, presentations, speaker’s bureaus, manuscript writing, or educational events) and support for attending meetings and/or travel from Gilead and Pfizer, outside of the submitted work. All the other authors do not report any conflict of interest or personal relationships that could have appeared to influence the work reported in this paper.

CHAPITRE

3

Inférence de la dynamique infectieuse HPV par chaînes de Markov cachées



3.1 Introduction

Le cancer du col de l’utérus est le quatrième cancer le plus diagnostiqués chez les femmes et également le quatrième plus meurtrier avec un fardeau de près de 342 000 décès en 2020 [Sung et al., 2021]. La quasi-totalité de ces cancers est causée par une infection HPV à haut-risque [Boshart et al., 1984; zur Hausen, 1996].

Les suivis réguliers des infections HPV montrent qu’elles sont très courantes. Dans la majorité des cas, leur détection s’accompagne d’une cytologie saine [Doorbar, 2023]. On perd généralement la détection du virus en moins de trois ans [Insinga et al., 2007] même s’il est courant de re-détecter le même génotype après une période de non-détection [Gravitt, 2012].

Plusieurs pistes d’explications peuvent être évoqués quant à la cause de ces re-détections : nouvelles acquisitions, ré-émergence d’infections latentes, dépôt transitoire d’acides nucléiques par le·la partenaire sexuel·le. De la même manière, la perte de détection d’une infection HPV peut être due à une élimination du virus chez l’hôte (avec ou sans séropositivité) ou à la progression de l’infection vers des niveaux en dessous des seuils de détection [Gravitt and Winer, 2017]. Dans ce dernier cas, on parle d’infection latente [Gravitt and Winer, 2017; Doorbar, 2023]. Les phénomènes de latence ont été attestés expérimentalement pour les papillomavirus infectant le lapin [Amella et al., 1994; Zhang et al., 1999; Maglennon et al., 2011, 2014]. Chez l’humain, les résultats sont plus contrastés [Gravitt and Winer, 2017]. Des épisodes de re-détections de certains génotypes sont observés chez des femmes adultes abstinentes, séropositives au VIH ou non [Strickler et al., 2005; Theiler et al., 2010; Rositch et al., 2012; Paul et al., 2021]. Cependant, chez les femmes sexuellement actives, il est difficile de trancher quant à la causalité des nouvelles détections HPV [Gravitt and Winer, 2017]. En particulier, chez les jeunes femmes, il est généralement admis que la majorité des détections provient de nouvelles acquisitions [Trottier et al., 2010; Moscicki et al., 2013]. Une étude évoque la possibilité de latence chez les adolescentes [Shew et al., 2015] mais leur analyse statistique est peu satisfaisante. Ils basent leurs conclusions sur la comparaison entre les *hazard ratios* de deux variables qualitatives, le nombre de partenaires sexuels dans les trois derniers mois et la présence de nouveaux partenaires parmi eux. Ils concluent à la latence comme explications majoritaires des motifs de détection car le *hazard ratio* de la première co-variable est significativement supérieure à 1 mais celui de la seconde n’est pas significatif. Néanmoins, en regardant plus en détails, les deux *hazard ratios* sont très proches et aucune donnée n’est publiée sur la relation entre les deux variables, qui pourraient être corrélées.

Les frottis cervicaux permettent de détecter la présence du virus dans les couches superficielles des tissus du col de l’utérus à l’aide de techniques PCR [Kamal, 2022]. On n’a cependant pas accès directement au réservoir de l’infection HPV que sont les cellules basales de l’épithélium.

Ces données binaires (présence/absence de détection du génotype viral) peuvent être considérées comme des témoins imparfaits de la présence du virus dans l'hôte. Si on ne considère que deux états possibles (présence ou absence du virus), les données de détection semblent cohérentes avec l'état de l'infection. Néanmoins, lors des phases latentes, on peut avoir une différence entre la biologie de l'infection et les données collectées. Du point de vue biologique, le virus est toujours présent chez l'hôte, mais les capacités de détection du virus sont limitées par la technologie employée (sensibilité et spécificité des tests employés) et le type du tissu échantillonné.

Dans ce travail, nous utilisons des modèles de Markov cachés pour décrire la dynamique des infections HPV, dont l'information n'est pas accessible, à partir de données issues de tests de détection du virus qui reflètent partiellement la présence du virus dans le tissu selon l'état de l'infection sous-jacente. On applique ce formalisme à trois cohortes distinctes. Deux jeux de données proviennent d'études interventionnelles menées par GlaxoSmithKline pour évaluer l'efficacité du vaccin Cervarix™ contre les infections HPV et le troisième jeu de données est issu de l'étude observationnelle PAPCLEAR. Les données utilisées sont donc des captures, à des instants donnés, d'un processus dont la dynamique varie en continu. Le formalisme des chaînes de Markov cachées a déjà été utilisé dans le contexte des cancers HPV [Bureau et al., 2003; Soper et al., 2020; Ikesu et al., 2022] mais il s'agit ici de la première application de ces outils pour l'évaluation de la dynamique des infections transitoires HPV. Il s'agit à notre connaissance du premier modèle mécaniste de la dynamique HPV à inclure un compartiment latent.

Nous avons inféré la dynamique du processus pour cinq scénarios différents, un premier où nous n'avons considéré que deux états pour l'infection HPV (absence ou présence du virus) et quatre cas avec un compartiment latent supplémentaire. Nous calculons en particulier le temps passé dans l'état non-infecté, pour estimer le temps entre des infections consécutives, et le temps passé infecté par un génotype. En particulier, notre travail fournit une estimation du temps passé dans les phases latentes ou les phases productives de l'infection. Nous montrons en particulier la différence de temporalité entre les données de détection et la dynamique infectieuse intra-hôte. Nous avons également évalué l'effet de la vaccination sur les diverses transitions d'un état à l'autre du système. Nos résultats montrent un effet protecteur de la vaccination contre l'acquisition de nouvelles infections et suggèrent un effet bénéfique dans l'élimination du virus dans les phases latentes.

3.2 Matériel et méthodes

3.2.1 Études cliniques et participantes

Nos analyses intègrent des jeux de données provenant de trois études cliniques. En premier lieu, nous utilisons les données de l'étude clinique PAPCLEAR (NCT02946346), déjà exploitées dans le chapitre 2. Il s'agit d'une étude clinique observationnelle débutée en 2016 visant une meilleure compréhension de la dynamique intra-hôte des infections HPV chez les jeunes adultes [Murall et al., 2019b]. Au total, sur les 150 femmes initialement incluses dans l'étude longitudinale, 132 femmes ont été suivies suffisamment de fois pour que leurs données de suivies soient pertinentes à analyser (voir chapitre 2). Les participantes étaient généralement suivies toutes les huit semaines par une gynécologue ou sage-femme et des prélèvements biologiques étaient effectués, notamment des frottis cervicaux [Murall et al., 2019b]. La durée de suivi maximale d'une participante infectée était de deux ans. Le protocole de l'étude clinique PAPCLEAR est détaillé ailleurs [Murall et al., 2019b] et un graphique résumant la structure de l'étude est accessible en annexe (figure A.5). Le protocole d'inclusion (voir ci-dessous et Murall et al. [2019b]) ne prévoyait pas de processus de sélection particulier en fonction du statut vaccinal et sur les 132 femmes analysées ici, 62 étaient vaccinées, dont 6 avec le vaccin bivalent Cervarix™ et 56 avec le vaccin quadrivalent Gardasil™.

Nous disposons également de données issues de deux cohortes menées par l'entreprise GlaxoSmithKline (GSK) pour évaluer de l'efficacité et de la sécurité de leur vaccin, le Cervarix™. Le premier jeu de données est issu de l'étude clinique NCT00518336¹, également appelé HPV-023, et correspond à une deuxième session de suivi, sept ans après l'inclusion des participantes dans l'étude initiale (HPV-001, NCT00689741). Le recrutement s'est effectué dans trois pays : États-Unis, Canada et Brésil. La première session de suivi prolongé (HPV-007, NCT00120848) a étendu l'étude sur les trois pays pour une durée de trois ans. L'étude HPV-023 s'est quant à elle restreinte aux participantes brésiliennes. La première étude HPV-001, lancée en 2001, est une étude clinique de phase IIb en double aveugle randomisée et les participantes n'ont pas été informées de leur statut vaccinal avant la fin de l'étude HPV-023. Les participantes étaient séparées en deux groupes de proportion équivalente : un groupe ayant été vacciné au Cervarix™ et l'autre groupe ayant reçu 500 μg d'hydroxyde d'aluminium (AlOH_3) comme placebo. Il a été proposé aux participantes ayant reçu le placebo une vaccination au Cervarix™ à la fin de l'étude HPV-023. Sur les 1.113 participantes incluses dans l'étude HPV-001, notre analyse se limite aux 433 participantes de l'étude HPV-023. Pour l'étude HPV-023, il était prévu que les participantes soient suivies

1. Pour les études cliniques GSK, il est possible d'accéder aux détails de l'étude clinique via <https://clinicaltrials.gov/study/NCT00XXXXXX> avec NCT00XXXXXX son identifiant.

sept fois au total à une fréquence d'une visite tous les six mois ; visites durant lesquelles des frottis cervicaux étaient réalisés. Le détail des visites est présenté sur la Figure A.6.

Le dernier jeu de données dont nous disposons provient d'une cohorte japonaise initiée en 2006 et qui avait également pour but d'évaluer l'efficacité et la sécurité du vaccin Cervarix™. Elle s'organise en deux études cliniques : le suivi initial (HPV-032, NCT00316693) et l'étude de suivi prolongé (HPV-063, NCT00929526) dans la continuité de HPV-032 deux ans plus tard. La première étude s'étale sur deux ans et est une étude clinique de phase IIb en double aveugle randomisé. Au total, 1.040 participantes ont été incluses dans le suivi dont 519 ont reçu les trois doses du vaccin et 521 ont reçu trois doses d'Aimmugen™, un vaccin contre l'hépatite A également commercialisé par GSK. Durant cette phase initiale, les participantes étaient suivies tous les six mois avec des prélèvements biologiques. Sur les 1.040 participantes initiales, 752 ont choisi de prendre part au suivi prolongé. L'étude HPV-063 et HPV-032 diffèrent sur plusieurs points. Bien que l'étude clinique dure également deux ans, les participantes ne sont suivies que deux fois au total, aux mois M36 et M48 depuis la première injection vaccinale. De plus, il a été indiqué au début d'HPV-063 aux personnes chargées de les suivre ainsi qu'aux participantes à quel groupe elles appartenaient. Seuls les laboratoires responsables des tests de dépistage sont restés aveugles tout du long de l'étude.

Critères d'inclusion et populations cibles

Pour l'étude PAPCLEAR, les participantes ont été recrutées dans la région de Montpellier et étaient pour la plupart étudiantes. Elles devaient avoir entre 18 et 25 ans et être sexuellement active avec un nouveau partenaire dans les douze derniers mois précédant l'inclusion. Les candidates présentant un historique de pathologie liée à HPV (verruge génitale ou lésions cervicales), les femmes enceintes ou prévoyant de l'être dans l'année à venir, les personnes infectées par le VIH, les personnes sous traitement médical intensif ou celles ayant prévu de quitter la région de Montpellier dans les 18 mois après l'inclusion ont été exclues [Murall et al. \[2019b\]](#).

Les critères d'inclusion et d'exclusion des études HPV-001, HPV-007 et HPV-023 ont été évoqués dans des travaux antérieurs [[Harper et al., 2004, 2006](#); [Roteli-Martins et al., 2012](#); [Naud et al., 2014](#)], qui sont complétées de manière approfondie sur les sites de référencement des études cliniques correspondants². Brièvement, pour être incluse dans HPV-001 les participantes devaient être âgées entre 15 et 25 ans au moment du premier test de dépistage. Elles devaient être en bonne santé, présenter un utérus intact et ne pas être enceinte, ni prévoir de l'être jusqu'à deux mois après la fin du schéma vaccinal. Des frottis cervicaux étaient effectués à l'inclusion pour évaluer l'état du col de l'utérus des candi-

2. <https://clinicaltrials.gov>

dates et seules les femmes présentant une analyse cytologique normale étaient incluses. Des tests de détection d'infections antérieures aux HPV étaient aussi effectués : les femmes séropositives par ELISA aux HPV16 ou HPV18 étaient exclues, ainsi que celles positives à l'ADN de n'importe quel génotype HPV à haut-risque par PCR. Enfin, seules les femmes déclarant avoir eu six ou moins de partenaires sexuel-les avant inclusion étaient admises. Pour participer aux suivis HPV-007 puis HPV-023, les participantes devaient avoir complété l'étude clinique précédente à son terme. À l'inverse de l'étude HPV-063, toute personne ayant été informée de son statut vaccinal a été exclue des extensions de suivi.

Enfin, les études cliniques HPV-032 et HPV-063 ont été détaillées dans des publications antérieures [Konno et al., 2010, 2014]. L'étude initiale a inclus des femmes entre 20 et 25 ans. Elles devaient être en bonne santé et présenter un utérus intact. Les candidates enceintes ou prévoyant de l'être avant le huitième mois après la première visite ont été exclues. Les femmes incluses dans l'étude HPV-032 enceintes trois mois avant le début de l'extension HPV-063 ont été exclues de cette dernière. Les participantes ayant été diagnostiquées avec une lésion de haut grade à l'inclusion de HPV-032 ont également été exclues de HPV-063.

Pour chacune des études, un consentement écrit et éclairé a été signé par les participantes.

Dépistage et génotypage HPV

Pour l'identification de génotypes HPV à haut risque, toutes les études cliniques analysées ici ont utilisé des méthodes d'amplification par PCR d'un segment de 65 nucléotides du gène L1. La présence d'alphapapillomavirus a été détectée par hybridation DEIA. Les échantillons positifs ont ensuite été génotypés par hybridation inverse LiPA, permettant la détection simultanée de 25 génotypes dont 14 génotypes à haut-risque³ et 11 génotypes à bas risque⁴ [Kleter et al., 1998a, 1999; Quint et al., 2001; van Doorn et al., 2002; Geraets et al., 2015b]. Pour l'étude clinique HPV-023, un test PCR supplémentaire spécifique a été réalisé par amplification d'un segment de 92 nucléotides du gène E6/E7 pour GPV16 et amplification d'un segment de 126 nucléotides du gène L1 pour HPV18 [van Doorn et al., 2006]. Pour l'étude PAPCLEAR, dans l'éventualité où le kit SPF10-LiPA25 retournait un résultat négatif malgré un DEIA positif, une amplification PCR PGMY et un séquençage Sanger [Coutlée et al., 2002] ont été réalisés pour tenter de déterminer le génotype HPV.

3.2.2 Analyse statistique

Nous modélisons la dynamique intra-hôte des infections HPV à l'aide d'un modèle de Markov caché (HMM) $(X_t, O_t)_{t \in \mathbb{R}_+}$, avec X_t la valeur du

3. Les génotypes HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68

4. Les génotypes HPV6, 11, 34, 40, 42, 43, 44, 53, 54, 68, 70, 74

processus caché au temps t et O_t la valeur prise par le processus observé au temps t . Seules les participantes vues au moins deux fois ont été incluses dans l’analyse. On fait l’hypothèse que la dynamique de chaque génotype est indépendante des dynamiques des autres génotypes, même s’ils sont présents chez la même participante au même moment [Dillner et al., 2011]. Comme classiquement fait dans la littérature HPV [Shew et al., 2015; Ramanakumar et al., 2016; Paul et al., 2021; Malagón et al., 2023], nous avons réalisé les analyses de manière génotype-spécifique à l’échelle du patient : on peut voir chaque paire participante-génotype comme un individu statistique. L’inférence des modèles est regroupée pour les 25 génotypes. Les estimations des paramètres peuvent être vues comme la dynamique infectieuse d’un génotype générique, moyenné sur l’ensemble des génotypes détectables.

Toutes les analyses statistiques ont été réalisées en R 3.6.3 à l’aide du paquet `msm` [Jackson, 2011b] et sont stratifiées selon le statut vaccinal des participantes.

Chaîne de Markov sous-jacente. On modélise le processus $(X_t)_{t \in \mathbb{R}_+}$, qui est une chaîne de Markov pouvant prendre un nombre fini d’états. Dans notre cas, $(X_t)_{t \in \mathbb{R}_+}$ décrit l’état de l’infection intra-hôte. On considère deux cas de figure. Dans le premier, on suppose l’absence de compartiment latent. Dans ce cas, l’hôte est soit infecté par le virus, soit non-infecté. Dans le second cas, on considère que l’infection peut être séparée en deux catégories : une phase productive et une phase latente. On note l’état « sain », où le virus est absent de l’hôte, par S et les états infectieux par P , pour une infection productive, et L , pour une infection latente.

Le processus $(X_t)_{t \in \mathbb{R}_+}$ satisfait la propriété de Markov dans le sens où la loi des états futurs du processus ne dépend du passé que son état présent. Mathématiquement, soit $(\mathcal{F}_t^X)_{t \geq 0}$ une filtration engendrée par un processus X , $(\mathcal{F}_t^X)_{t \geq 0} := \sigma(\{X_s, s \leq t\})$, on a

$$\mathbb{P}[X_{t+s} | (\mathcal{F}_t^X)_{t \geq 0}] = \mathbb{P}[X_{t+s} | X_t]. \quad (3.1)$$

Loi d’émission. Les seules données d’observation de la dynamique infectieuses accessibles sont les résultats des tests de détection des génomes HPV par la méthode PCR-DEIA/SPF₁₀-LiPA₂₅. Ces résultats se présentent sous forme binaire, la valeur 1 indiquant la détection du génotype et la valeur 0 sa non-détection. On suppose ici que la loi d’émission, c’est-à-dire la distribution aléatoire qui lie le processus caché au processus observé, suit une loi de Bernoulli de paramètre p , où p représente la probabilité que le matériel génétique du virus soit à des niveaux détectables par les techniques de laboratoire. Comme toutes nos données reposent sur le même kit de détection, on fait l’hypothèse que cette probabilité ne varie pas selon les études cliniques.

On suppose également que la loi d'émission du processus ne dépend que de l'état du processus sous-jacent à l'instant donné :

$$\mathbb{P}[O_t | \mathcal{F}_t^X, \mathcal{F}_t^O] = \mathbb{P}[O_t | X_t]. \quad (3.2)$$

Dans la suite du document, on notera par $p_X \in [0, 1]$ la probabilité de détecter la présence du virus lorsque l'infection est dans l'état $X \in \{S, P, L\}$. L'ensemble des p_X définit ainsi la loi d'émission du processus. A priori, on s'attend à ce qu'en l'absence de virus chez l'hôte, la probabilité de détecter de l'ADN viral soit quasiment nulle ($p_S = 0$) et qu'elle soit proche de 1 lors d'une infection productive ($p_P = 1$).

Scénarios envisagés. Pour un processus de Markov à $n \in \mathbb{N} \setminus \{0\}$ états, la matrice du générateur comprend au plus $n^2 - n$ degrés de liberté. Chacun de ces termes traduisant une potentielle transition d'un état donné vers un autre état différent, cela représente un total de $2^{n^2 - n}$ modèles possibles. Toutefois, dans notre situation, une partie significative de ces modèles n'est pas pertinente du point de vue biologique. En premier lieu, on peut éliminer les scénarios présentant un état absorbant. Pour le modèle à deux états, cela ne nous laisse qu'un seul modèle à considérer, que l'on note scénario Δ . Pour le modèle à trois états, on se restreint à l'exploration des modèles dont les transitions entre les états S et P sont non nulles, laissant au plus 16 modèles à analyser. Enfin, on fait l'hypothèse que l'état L ne peut pas précéder l'état P en sortie directe du compartiment S , autrement dit, une infection ne peut pas d'abord être latente puis devenir productive sans clairance de l'infection entre les deux. Cela nous laisse donc avec cinq scénarios possibles (Figure 3.1). Dans la suite des analyses, on met de côté le scénario E car il distingue les états latents et productifs en deux compartiments qui ne communiquent pas, cas de figure incompatible avec les connaissances actuelles [Gravitt and Winer, 2017; Doorbar, 2023].

L'inférence des taux de transitions instantanés de nos modèles est réalisée par maximisation de la vraisemblance (pour plus de détails, se référer à la section 1.3.3 et à l'équation 1.44). On modélise l'effet des covariables sur le temps de transition en faisant l'hypothèse de risque proportionnel [Marshall and Jones, 1995], comme dans une analyse de survie classique (voir l'équation 1.36), pour le générateur $Q = (q_{ij})$:

$$q_{ij}(\mathbf{X}) = q_{ij}^{(0)} \exp(\beta^t \mathbf{X}). \quad (3.3)$$

avec β^t la transposée du vecteur β . Nous avons adopté une approche fréquentiste pour l'inférence [Jackson, 2011b] et, pour faciliter la convergence de l'algorithme, nous avons initialisé les paramètres de loi d'émission avec $p_S = 10^{-5}$, $p_P = 0.5$ et $p_L = 0.5$ et les valeurs non nulles du générateur à 10^{-4} . La reconstruction des états cachés à partir des états observés a été calculée à l'aide de l'algorithme de Viterbi [Viterbi, 1967]. Pour le modèle à deux états, nous avons également comparé les inférences

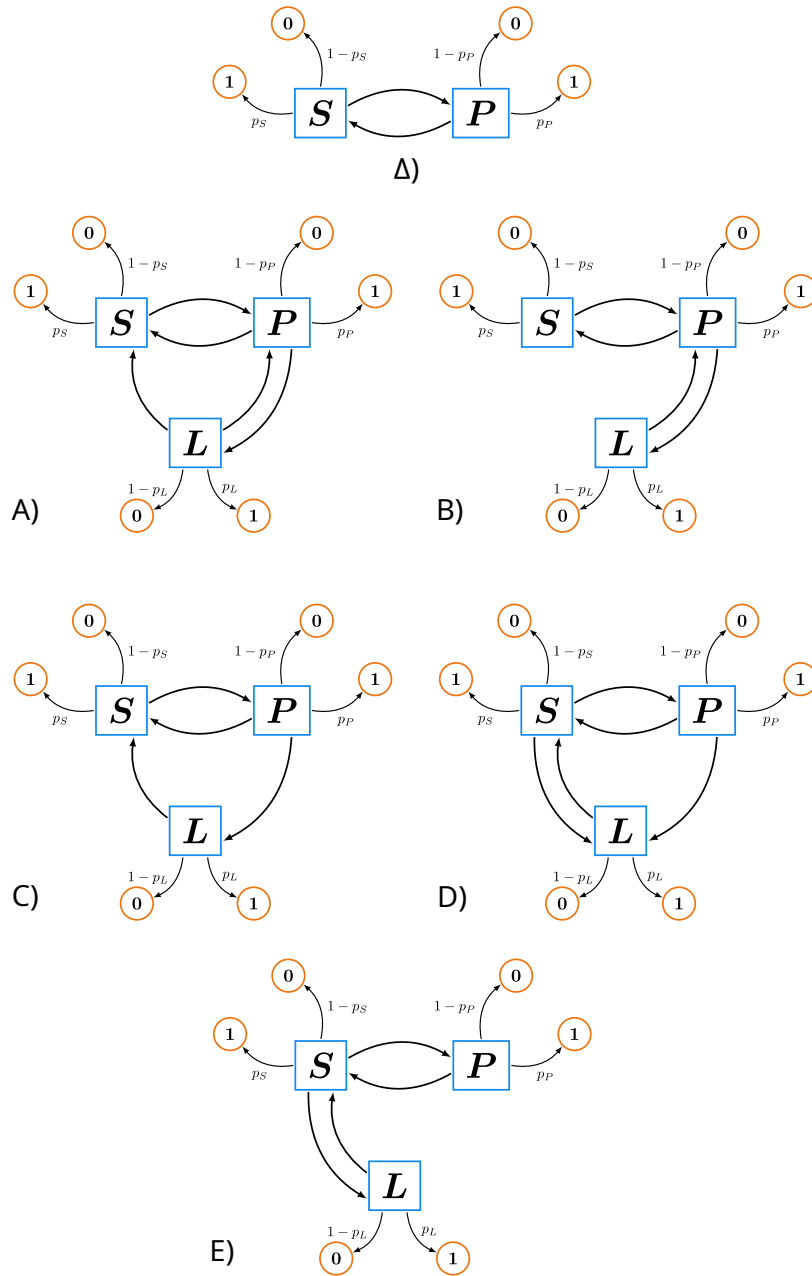


FIGURE 3.1 – Cinq variantes d’une chaîne de Markov à trois états cachés et à loi d’émission binomiale. Ces cinq modèles supposent l’absence d’un état absorbant, la communication réciproque entre les compartiments S et P et l’impossibilité pour le compartiment L de précéder l’état P sans passage intermédiaire par le compartiment S . Les compartiments dans les rectangles bleus sont les états cachés du processus et les cercles oranges correspondent aux états observés.

sur la matrice de transition entre un modèle de Markov caché et un modèle multi-état (MSM) pour lequel $p_S = 0$ et $p_P = 1$. Ce modèle décrit plus les durées entre épisodes détectées et s’approche plus des analyses de survies du chapitre précédent, avec cette fois-ci un ajustement des

données à une loi exponentielle.

3.3 Résultats préliminaires

3.3.1 Analyse descriptive

Le détail des résultats est récapitulé dans le Tableau 3.1. Brièvement, seules les participantes qui ont été suivies sur au moins de 2 visites ont été incluses dans l'analyse, ce qui représente 132 femmes de la cohorte PAPCLEAR, 957 des cohortes HPV-032 et HPV-063 et 431 de la cohorte HPV-023, pour un total de 1520 participantes. En moyenne, nous avons considéré environ 90 % des participantes incluses dans l'analyse et les femmes exclues de l'analyse ne sont pas significativement différentes de celle incluses.

Toutes les cohortes incluent des jeunes femmes, la cohorte PAPCLEAR regroupant les participantes les plus jeunes, avec 21,3 ans en moyenne, et la cohorte brésilienne les plus âgées, avec 26,5 ans en moyenne. Dans chacune des trois études, les participantes sont réparties équitablement entre statut vacciné et non-vacciné pour HPV.

Les durées totales de suivis varient grandement selon les études. Pour les cohortes interventionnelles (HPV-007, HPV-032 et HPV-063), la durée moyenne de suivi est cohérente avec le plan expérimental : 24 mois pour les participantes de l'étude HPV-032 avec une prolongation jusqu'à 48 mois pour celles qui ont aussi été suivies dans HPV-063 et environ 30-36 mois pour les participantes de l'étude HPV-007. Pour l'étude PAPCLEAR, la durée de total de suivie est plus faible et environ d'un an, mais pour chaque participante, on dispose en moyenne du même nombre d'observations avec six visites par participantes au cours du suivi.

Sur les trois études, on compte 56,0 % participantes (851/1520) qui ont présenté au moins une fois au cours du suivi une détection positive à un génotype HPV. Cette proportion est la plus forte chez les participantes de l'étude PAPCLEAR avec 74,2 % des participantes incluses chez qui on a détecté au moins un génotype, contre 58,9 % (564/957) pour les cohortes japonaises et 43,9 % (189/431) pour la cohorte brésilienne. Ces chiffres sont cohérents avec la littérature [Chaturvedi et al., 2011]. Les informations sur les génotypes détectés entre chaque étude sont détaillées dans le tableau A.2 en annexe. On observe également des épisodes de co-détectations (deux génotypes HPV ou plus détectés au cours d'une même visite) chez un peu plus de la moitié (467/851) des participantes positives au moins une fois à HPV. Cela varie selon les études cliniques considérées. Les proportions de co-détectations sont un peu plus importantes que dans d'autres travaux [Chaturvedi et al., 2011].

TABLEAU 3.1 – Récapitulatif des participantes incluses dans l'analyse.

Cohortes	Cohorte brésilienne	Cohorte japonaise	PAPCLEAR	Global
# de participantes (% de la pop. totale)	431 (99,31%)	957 (91,49%)	132 (88,00%)	1520 (93,25%)
âge moyen* (IQR)	26,46 (24 - 29)	22,49 (21 - 24)	21,27 (20 - 24)	23,51 (21 - 25)
# moyen de visites (IQR)	5,60 (5 - 6)	6,25 (6 - 7)	6,57 (5 - 8)	6,10 (5 - 7)
Durée moyenne de suivi (IQR)	847,9 (847 - 883)	1216,6 (1064 - 1449)	347,8 (203 - 468)	203,41 (165 - 203)
# jours entre 2 visites (IQR)	184,1 (161 - 192)	231,61 (175 - 295)	62,46 (52 - 66)	1036,6 (760 - 1407)
# participantes vaccinées	220 (51,0%)	474 (49,5%)	66 (50,0%)	760 (50,0%)
participantes avec ≥ 1 détection HPV	189 (43,9%)	564 (58,9%)	98 (74,2%)	851 (56,0%)
participantes avec ≥ 1 codétection HPV	83 (19,3%)	322 (33,6%)	62 (47,0%)	467 (30,7%)

* au début de l'étude

3.3.2 Description des scénarios

Une première valeur d'intérêt à extraire des modèles est le temps de séjour moyen passé dans les différents états cachés. Mathématiquement, ces temps de séjour suivent une loi exponentielle de paramètre q_X , $X \in \{S, P, L\}$, l'espérance du temps de séjour valant $1/q_X$ et correspondant au temps moyen passé dans le compartiment X avant de le quitter pour l'un ou l'autre des autres compartiments $Y \in \{S, P, L\} \setminus \{X\}$. Les temps de séjour pour chaque scénario sont récapitulés dans le Tableau 3.2. Le temps de séjour dans le compartiment S varie autour de 30×10^3 jours pour l'ensemble des inférences des chaînes de Markov cachée et est réduit de moitié pour le modèle multi-états. Ce temps traduit la durée moyenne passée par un individu sans infection (ou sans détection positive pour le MSM), pour le génotype moyen. Si l'on suppose 25 génotypes aux caractéristiques semblables à celles du génotype moyen, la probabilité de ne pas être infecté par un seul génotype au bout d'un an est proche de 0,74 et celle après deux ans légèrement inférieure à 0,54 (avec une loi exponentielle de paramètre $25 \times q_S$) pour les chaînes de Markov cachée. Pour les modèles à trois compartiments, le temps moyen de la phase productive est inférieure à celui de la phase latente et est généralement de l'ordre de la centaine de jours.

Le détail de l'ensemble des taux de transitions est récapitulé dans la Figure A.8 en annexe. À partir du générateur, on peut calculer la probabilité de passer de l'état $X \in \{S, P, L\}$ à l'état $Y \neq X$, au moment d'un changement d'état par $\mathbb{P}[X \rightarrow Y] = Q_{X,Y}/q_X$. En particulier, on remarque que pour les scénarios A et D, la probabilité de passage par l'état latent depuis l'état productif est un événement très courant ($\mathbb{P}_A[P \rightarrow L] = 0.993$; $\mathbb{P}_D[P \rightarrow L] = 0.988$) mais est plus rare dans les scénarios B et C ($\mathbb{P}_B[P \rightarrow L] = 0.260$; $\mathbb{P}_C[P \rightarrow L] = 0.471$). Dans le scénario D, la probabilité de transition de l'état S vers L est faible ($\mathbb{P}_D[S \rightarrow L] = 0.026$).

Concernant les distributions d'émission, pour les différents modèles, la probabilité de détecter le virus chez un individu sain est très faible (de l'ordre de 8×10^{-3}) et la probabilité de détecter la présence du virus en phase productive est proche de 1 (Figures 3.2 et A.9). Logiquement, la valeur de p_P étant la plus faible pour le HMM à deux états ($p_P = 0,91$). Lorsque l'infection est latente, la probabilité de détection du virus varie selon les scénarios, mais est généralement inférieure à p_P . Cette probabilité reste globalement supérieure à 0,5 sauf pour le scénario B et le scénario C où la borne inférieure de l'intervalle de confiance passe en dessous de 0,5.

3.3.3 Temps de retour à l'état sain

Une autre quantité d'intérêt est la durée d'une infection. Pour les modèles à deux états, la distribution du temps infecté est directe et suit une loi exponentielle de paramètre q_P . Pour les modèles à trois états,

TABLEAU 3.2 – Temps de séjour dans les états cachés. L'unité de temps est en nombre de jour.

État caché	S	P	L
Scénario Δ (MSM)	$14,9 \times 10^3$ [$14,2 \times 10^3 - 15,7 \times 10^3$]	252,5 [241,1 - 264,5]	–
Scénario Δ (HMM)	$28,6 \times 10^3$ [$26,5 \times 10^3 - 31,0 \times 10^3$]	609,0 [550,6 - 673,7]	–
Scénario A	$32,4 \times 10^3$ [$29,9 \times 10^3 - 35,0 \times 10^3$]	154,7 [84,1 - 284,6]	322,1 [95,1 - 1090,5]
Scénario B	$28,9 \times 10^3$ [$26,6 \times 10^3 - 31,1 \times 10^3$]	403,5 [302,7 - 537,8]	564,1 [167,2 - 1902,5]
Scénario C	$29,6 \times 10^3$ [$26,9 \times 10^3 - 32,5 \times 10^3$]	335,3 [181,8 - 618,5]	807,6 [357,2 - 1825,7]
Scénario D	$32,3 \times 10^3$ [$29,8 \times 10^3 - 35,0 \times 10^3$]	120,9 [84,5 - 172,9]	570,0 [505,7 - 642,3]

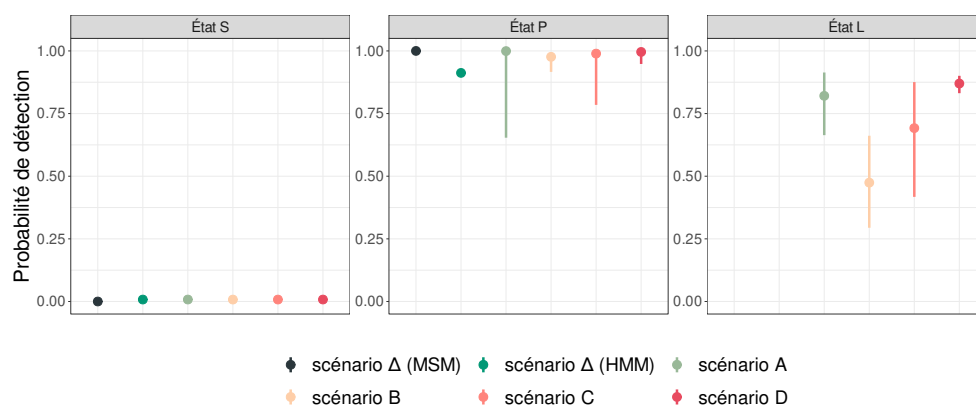


FIGURE 3.2 – Probabilité de détection du virus selon l'état caché. Les probabilités de détection depuis l'état S sont écrasées par la perspective mais un zoom sur les valeurs de p_S selon les scénarios est disponible en annexe (Figure A.9).

les choses sont moins évidentes. Comme nous avons deux compartiments infectieux, on définit le temps d'une infection comme le temps entre l'instant où l'individu quitte le compartiment S et l'instant de premier retour au compartiment S , temps que l'on note également temps de retour à l'état sain $T_S \in \mathbb{R}_+^*$. Celui-ci diffère du temps de séjour évoqué précédemment et qui nous renseigne sur le temps moyen passé dans un compartiment donné avant de le quitter, que ce soit vers un compartiment infecté ou sain. Dans la suite, on note $\forall t \in \mathbb{R}_+$

$$\psi_k(t) = \lim_{\tau \rightarrow 0} \frac{\mathbb{P}_k [T_S \in [t, t + \tau]]}{\tau} \quad (3.4)$$

la densité au temps t de la loi de probabilité de T_S dans le scénario $k \in \{\Delta, A, B, C, D\}$.

La fonction de densité des scénarios A et B peut être définie comme une série infinie de fonctions définies par récurrence. Toutefois, la résolution de cette série est non-triviale. Pour les scénarios C et D, on dispose d'une formule explicite de la loi de $\psi(t)$

$$\psi_\Delta(t) = q_P e^{-q_P t} \quad (3.5)$$

$$\psi_A(t) = \sum_{k=0}^{+\infty} \delta_k(t) + \lambda_k(t) \quad (3.6)$$

$$\psi_B(t) = \sum_{k=0}^{+\infty} \delta_k(t) \quad (3.7)$$

$$\psi_C(t) = Q_{P,S} e^{-q_P t} + \frac{Q_{P,L} q_L}{q_P - q_L} (e^{-q_L t} - e^{-q_P t}) \quad (3.8)$$

$$\psi_D(t) = \frac{Q_{S,P}}{q_S} \psi_C(t) + \frac{Q_{S,L}}{q_S} q_L e^{-q_L t} \quad (3.9)$$

avec $(Q_{X,Y})_{(X,Y) \in \{S,P,L\}}$ la valeur du taux de transition de l'état X vers l'état Y , $q_X = -Q_{X,X}$ et les fonctions $\delta_k(t)$ et $\lambda_k(t)$ définies par récurrence

$$\delta_k(t) = \int_0^t Q_{P,L} q_L e^{-(q_P + q_L)u} \delta_{k-1}(t-u) du \quad ; \quad \delta_0(t) = Q_{P,S} e^{-q_P t} \quad (3.10)$$

$$\lambda_k(t) = \int_0^t Q_{P,L} q_L e^{-(q_P + q_L)u} \lambda_{k-1}(t-u) du \quad ; \quad \lambda_0(t) = Q_{P,L} Q_{L,S} e^{-(q_P + q_L)t} \quad (3.11)$$

Dans tous les cas, on peut estimer numériquement ces distributions par méthode de Monte-Carlo. À partir d'une collection de temps de retour indépendants et de même loi, on estime la fonction de survie $(\int_k^\infty \psi(u) du)$

TABLEAU 3.3 – **Temps de retour à l'état sain.** L'unité de temps est en nombre de jours. Pour les scénarios A et B, 100.000 trajectoires ont été générées pour estimer la moyenne et les quantiles.

	Moyenne	Quantiles			
		10%	25%	75%	90%
scénario Δ (MSM)	252,52	26,61	72,64	350,07	581,45
scénario Δ (HMM)	609,05	64,17	175,21	844,32	1402,39
scénario A	673,12 \pm 3, 51	144,24	274,49	902,56	1406,97
scénario B	741,67 \pm 5, 47	58,40	165,00	978,95	1841,02
scénario C	715,77	67,46	274,29	900,13	1695,05
scénario D	680,79	137,88	269,91	915,57	1438,14

à l'aide de l'estimateur de Nelson-Aalen [Nelson, 1969, 1972; Aalen, 1978]. Les résultats pour chaque scénario sont détaillés sur la Figure A.10 et un graphique avec le logarithme de la fonction de survie est disponible en annexe pour mieux visualiser la proximité dans les distributions des différents scénarios (Figure A.11).

Pour les deux scénarios où l'on dispose de la densité de probabilité, on peut calculer l'espérance du temps de retour en S avec la formule

$$\mathbb{E}[T_S] = \int_0^{+\infty} u\psi_k(u)du \quad (3.12)$$

Les quantiles du temps de retour à 10 %, 25 % 75 % et 90 % depuis le début de l'infection sont obtenus à l'aide des fonctions de répartitions $\Psi_k(t) = \int_0^t \psi_k(u)du$:

$$\Psi_{\Delta}(t) = 1 - e^{-q_P t} \quad (3.13)$$

$$\begin{aligned} \Psi_C(t) &= \frac{Q_{P,L}}{q_P - q_L} (1 - e^{-q_L t}) \\ &+ \frac{q_P Q_{P,S} - q_L (Q_{P,S} + Q_{P,L})}{q_P (q_P - q_L)} (1 - e^{-q_P t}) \end{aligned} \quad (3.14)$$

$$\Psi_D(t) = \frac{Q_{S,P}}{q_S} \Psi_C(t) + \frac{Q_{S,L}}{q_S} (1 - e^{-q_L t}) \quad (3.15)$$

Pour les scénarios A et B, on estime numériquement ces quantités à l'aide des méthodes de Monte-Carlo. Les résultats pour chaque scénario sont récapitulés dans le Tableau 3.3

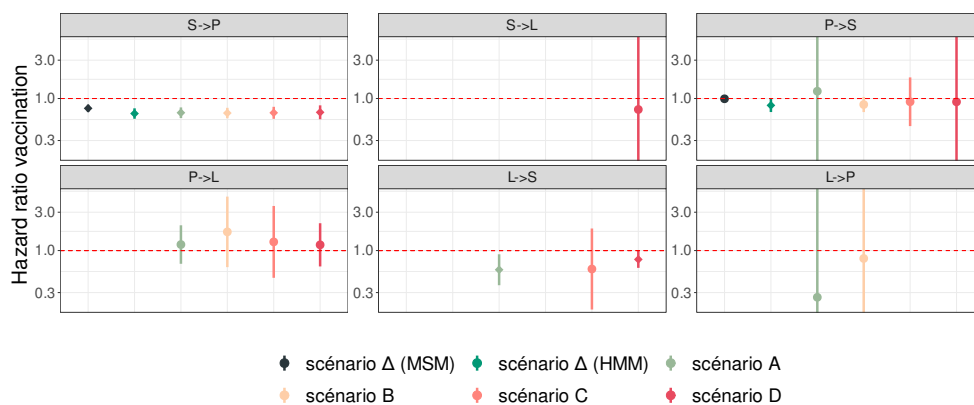


FIGURE 3.3 – **Effet de la vaccination sur les taux de transitions entre états cachés.** Pour toutes les analyses, la référence est la non-vaccination. Un *hazard ratio* inférieur à 1 (sous la ligne tiretée rouge) indique que le temps de transition est plus long pour les personnes vaccinées comparativement aux personnes non-vaccinées, et inversement si le *hazard ratio* est supérieur à 1. Les points en diamants soulignent les *hazard ratios* significatifs et les barres indiquent les intervalles de confiance à 95 %. Par soucis de lisibilité, l’axe des ordonnées a été restreint à l’intervalle [0,12 - 7,5].

3.3.4 Effet de la vaccination

Les *hazard ratios* de la vaccination sur les transitions entre états cachés sont récapitulés dans la Figure 3.3. On retrouve l’effet protecteur de la vaccination sur l’acquisition de nouvelles infections, avec un *hazard ratio* significativement inférieur à 1 pour les femmes vaccinées comparativement à celles non-vaccinées pour la transition de l’état S vers P sur l’ensemble des cinq scénarios. Cela indique que le temps de transition de l’état S vers l’état P est plus long chez les femmes vaccinées que chez celles non-vaccinées.

On ne détecte aucune autre tendance majeure de l’effet de la vaccination sur les taux de transition entre états cachés.

3.4 Discussion

Dans ce travail, nous avons modélisé la dynamique infectieuse des papillomavirus humains à l’aide d’un modèle de Markov caché. L’utilisation de ce modèle est motivée par la difficile obtention de l’information sur la présence du virus directement dans les couches basales. On dispose d’indices indirects par le biais de dépistages de l’ADN viral à la surface de l’utérus. Ces tests de détection présentent généralement une bonne sensibilité au virus [Geraets et al., 2015b] mais il a été suggéré que le virus pouvait entrer en phase latente durant laquelle il est maintenu sous contrôle par le système immunitaire à de faibles niveaux [Gravitt and

Winer, 2017; Doorbar, 2023]. Les modèles de Markov cachés permettent d’inférer la dynamique d’un processus caché par la seule observation d’un autre processus dont la loi d’émission dépend du système sous-jacent au moment de l’observation.

Notre modèle représente la première estimation de la dynamique infectieuse HPV à l’aide d’un modèle de chaîne de Markov caché. À l’inverse des approches classiques d’analyse de survie, ce travail permet de prendre en compte la pluralité de la dynamique infectieuse en distinguant une phase productive, assurément détectable, et une phase latente, moins fréquemment détectée lors de tests PCR. Cela se manifeste notamment dans la différence de dynamique entre les modèles de Markov cachés et le modèle multi-états. Les temps de séjour dans les compartiments sont beaucoup plus courts pour ce dernier.

Nous proposons ici cinq scénarios possibles d’histoire naturelle des infections HPV. Les cinq approches par chaînes de Markov cachées aboutissent à des estimations proches de la durée moyenne de présence du virus chez l’hôte, mais avec des implications biologiques distinctes. La présence du compartiment latent entraîne une augmentation des quantiles durées infectieuses dans tous les scénarios comparativement au scénario à deux états. Dans le scénario Δ HMM, on retrouve une sensibilité du test de détection ($p_P = 0,91$) similaire aux résultats de la littérature [Geraets et al., 2015b]. Les résultats des modèles à trois états suggèrent que la sensibilité des tests de détection serait excellente dans les phases productives de l’infection, mais médiocre lorsque l’infection entre en phase latente, phase durant laquelle le virus persiste à des niveaux faibles [Gravitt and Winer, 2017; Doorbar, 2023]. Là où dans les scénarios A et D les infections transitent dans la majorité de cas ($> 95\%$) par le compartiment latent avant de s’éteindre, pour les scénarios B et C, la transition par le compartiment latent a lieu dans moins de la moitié des cas. Dans le scénario D, pour lequel l’infection peut entrer en état latent sans passer par l’état productif, cette transition n’a lieu que très rarement ($\mathbb{P}_D[S \rightarrow L] = 0.026$). Cela suggère que même dans ce cas de figure, les infections latentes sont plutôt la progression d’une infection productive sous contrôle que deux types distincts d’infection. Dans chaque scénario, on retrouve des temps d’acquisition de nouvelles infections assez similaires. Si on se place à l’échelle d’un seul génotype, on estime un temps moyen d’acquisition du virus autour de 80-90 ans. Rapporté à l’ensemble des 25 génotypes, le risque d’acquisition d’au moins un génotype après un an est proche de 25 % et atteint près de 45 % après deux ans. Ces résultats sont sensiblement plus importants que pour d’autres cohortes ou travaux ne spécifiant pas de phase latente [Bruni et al., 2010; Paul et al., 2021; Malagón et al., 2023]. Cela peut également provenir de la différence de populations entre les cohortes analysées. En particulier, nous nous sommes concentrées sur des populations relativement jeunes. Or la prévalence est généralement maximale lors des premiers rapports sexuels et diminue avec l’âge [Bruni et al., 2010] Néanmoins, nos estimations

restent cohérentes avec d'autres travaux analysant la cohorte brésilienne [Ramanakumar et al., 2016].

Nos estimations du temps d'infection sont supérieures à certaines valeurs estimées précédemment [Insinga et al., 2007; Ramanakumar et al., 2016]. Néanmoins, ces études ne prenaient pas en compte explicitement la présence d'un compartiment latent et se basaient directement sur les données de détection pour inférer la durée infectieuse. Cela a pu engendrer une sous-estimation de cette dernière.

On retrouve l'effet protecteur de la vaccination avec un risque d'acquisition plus faible chez les participantes vaccinées comparativement aux participantes non-vaccinées. Cela est cohérent avec la littérature [Brotherton, 2019]. Nos résultats suggèrent également un effet bénéfique de la vaccination dans la clairance des infections en phase latente, cohérent avec la vision de la latence comme étant sous contrôle par le système immunitaire [Doorbar, 2023].

Une limitation de notre approche provient de l'utilisation de la propriété de Markov. Cette limitation provient en partie des contraintes d'accès aux données GSK qui s'effectue par l'intermédiaire d'un bureau virtuel dont la version du logiciel R est obsolète et nous restreint à l'utilisation de certains paquets dont MSM. La propriété de Markov sur la chaîne latente induit une dynamique suivant une loi exponentielle. Or, les distributions exponentielles ne sont pas forcément les mieux adaptées pour décrire les phénomènes de persistance intra-hôte, surtout si les processus immunitaires entrent en jeu. On peut par exemple relâcher la contrainte imposée par la propriété de Markov en faisant l'hypothèse que la chaîne sous-jacente est un processus semi-markovien. Dans ce cas, la distribution des états futurs dépend à la fois de l'état présent et du temps passé depuis le dernier saut. De tels modèles permettent d'incorporer des distributions aux fonctions d'aléa dépendantes du temps, comme la loi de Weibull (qui peut être vue comme une loi avec vieillissement), plus réalistes que la loi exponentielle qui fait l'hypothèse implicite que la probabilité instantanée de la réalisation d'un événement est constante à tout instant. Néanmoins, l'inclusion de données de multiples études cliniques aux fréquences d'échantillonnage différentes permet d'appréhender la dynamique infectieuse sur plusieurs temporalités (dynamique au court terme avec PAPCLEAR et dynamique au long terme avec les études GSK), en particulier si l'échantillonnage n'est pas régulier. La propriété de Markov nous permet d'envisager chaque observation comme un instantané de la dynamique infectieuse, dépassant les contraintes de censure discutées au chapitre précédent.

Ce travail fournit une vérification logique (*proof of concept* [Servidio et al., 2014]) qui soutient l'hypothèse de latence par compatibilité avec les observations de ces cohortes. Nous nous sommes concentrés ici sur une population assez jeune, assez exposées aux nouvelles acquisitions HPV [Bruni et al., 2010]. Il serait intéressant d'évaluer les dynamiques infectieuses dans des populations plus âgées ou susceptibles de présen-

ter préférentiellement des persistances en phase latente (femmes abstinentes, séropositives au VIH) pour comparer les dynamiques et identifier les déterminants de l'entrée en phase latente plutôt que la clairance de l'infection.

Remerciements

Nous remercions toutes les participantes de l'étude clinique PAP-CLEAR ainsi que l'ensemble du personnel médical pour leur aide. Nous remercions également GlaxoSmithKline de nous avoir fourni un accès aux données des études vaccinales et l'université de Montpellier pour avoir pris en charge les démarches administratives et légales pour le partage de données entre l'entreprise et nous.

CHAPITRE

4

Modélisation du cycle de vie HPV : rôle de l'aléatoire dans la clairance des infections



4.1 Aperçu de l'article

Malgré l'enjeu de santé publique que représentent les infections HPV, relativement peu d'attention a été portée à la modélisation mathématique de ces infections, la majorité des travaux sur le sujet relevant surtout de la médico-économie des politiques de prévention [Ryser et al., 2017]. Les approches de modélisation mécaniste appliquées aux HPV sont, quant à elles, relativement limitées, en particulier sur le sujet des infections transitoires. Pourtant, elles sont particulièrement utiles dans la génération et l'analyse de nouvelles hypothèses explicatives [Keeling and Rohani, 2008] et complémentaires des approches classiquement utilisées pour l'étude des infections virales [Djidjou-Demasse et al., 2020].

Les papillomavirus infectent les cellules souches des tissus épithéliaux. Ces tissus présentent une stratification tridimensionnelle, ce qui influe grandement sur la dynamique infectieuse, avec une couche basale permettant le renouvellement du tissu et un flux de cellules unidirectionnel des couches basales vers les couches apicales [Murall et al., 2019a]. Ce renouvellement s'effectue par des cellules souches qui se divisent principalement de manière asymétrique en donnant une cellule souche fille, qui reste dans le tissu basal, et une cellule souche parabasale, qui migre vers la surface. Toutefois, de rares événements de divisions symétriques peuvent avoir lieu [Averette et al., 1970; Clayton et al., 2007] conduisant à deux cellules filles basales ou parabasales. Les HPV ne se transmettent pas de cellule à cellule dans le tissu et la libération de particules virales ne s'effectue que dans les parties les plus apicales de l'épithélium. Ainsi, la persistance de l'infection est inféodée au maintien de virus dans les couches basales. Dans les cellules basales, le virus persiste majoritairement sous forme épisomale, c'est-à-dire sous forme d'ADN circulaire [Doorbar et al., 2012]. Cette forme représente le support intra-cellulaire du génome viral et permet une persistance dans le noyau sans intégration dans les chromosomes humains (sauf pendant les cancers). Notre modèle s'intéresse tout particulièrement à la dynamique des épisomes viraux infectant les cellules souches des épithéliums cervicaux.

Ryser et al. [2015] ont été les premiers à mettre en avant le rôle de l'aléatoire dans la dynamique des infections HPV. Dans ce travail, ils étudient l'impact de l'homéostasie, c'est-à-dire le maintien par l'organisme de paramètres biologiques à des niveaux optimaux, des cellules basales du col de l'utérus sur la probabilité de clairance des infections HPV. Ils suggèrent notamment la survenue aléatoire de rares divisions symétriques pourrait contribuer à expliquer 80 % des clairances HPV. Ce travail est limité car restreint aux seules situations homéostatiques, où la probabilité de division symétrique en deux cellules souches est égale à la probabilité de division symétrique en deux cellules différenciées. Notre approche permet d'étudier la dynamique infectieuse dans les contextes de sorties d'homéostasie, notamment dans le contexte des réponses de guérison des micro-lésions.

Une composante originale de notre modèle est l'intégration additionnelle d'une composante intra-cellulaire qui décrit l'attribution aléatoire des épisomes viraux lors de division cellulaire. Cet aspect est particulièrement crucial dans la dynamique infectieuse car le virus se maintient en faibles copies dans la cellule hôte sous forme d'épisomes [Doorbar, 2006; Moody, 2017]. Comme le virus n'est pas lytique, il n'y a pas de transmission de cellules à cellules dans le tissu, hormis lors des mitoses pendant lesquelles les copies virales sont transmises aux cellules filles. Cette transmission lors des mitoses est aléatoire même si certaines protéines virales pourraient contribuer à favoriser la transmission vers les cellules filles souches [McPhillips et al., 2005; Egawa et al., 2012]. Cependant, comme le nombre de copies virales par cellule infectée est limité, surtout en début d'infection, les attributions aléatoires des épisomes dans les cellules filles ont un rôle majeur dans la poursuite d'une infection.

Les couches basales sont également les zones de l'épithélium où la pression immunitaire est la plus forte [Pudney et al., 2005; Doorbar et al., 2012; Stanley, 2021]. La persistance du virus en faible nombre dans les cellules souches avec une activité métabolique réduite pourrait donc permettre de limiter les risques d'interaction avec les agents du système immunitaire [Doorbar, 2006; Stanley, 2006]. Néanmoins, une telle stratégie s'accompagne d'une exposition accrue aux fluctuations stochastiques, qui peuvent compromettre la persistance du virus dans le tissu.

À l'échelle moléculaire, les phénomènes de clairance sont partiellement compris, même si le rôle des gènes E1 [Egawa et al., 2012] et E6/7 a été mis en avant [Tomaić, 2016; Mirabello et al., 2017]. La piste immunitaire, bien qu'intuitive, n'est également pas entièrement suffisante [Carter et al., 2000; Gravitt, 2012; Stanley, 2021]. Comme suggéré par Ryser et al. [2015], d'autres mécanismes biologiques peuvent être invoqués pour expliquer ces dynamiques infectieuses.

Les données actuelles sur ce sujet sont limitées et ne permettent pas de trancher définitivement sur la question. La modélisation offre dans ce contexte l'opportunité d'explorer ces nouvelles pistes grâce à l'élaboration d'un modèle théorique approprié. Par une approche théorique combinée à une exploration numérique de l'ensemble des scénarios possibles, nous avons étudié l'hypothèse selon laquelle la clairance des infections HPV pourrait ne résulter que de phénomènes aléatoires inhérents au cycle de vie du virus. Contrairement aux chapitres précédents où la confrontation aux données biologiques était centrale, nous proposons ici une approche plus théorique à valeur de *preuve de concept* [Servedio et al., 2014].

Plus généralement, notre modèle permet de replacer la dynamique du virus dans un contexte évolutif autour du compromis entre évasion immunitaire et persistance intra-hôte. Sans pour autant exclure totalement la possibilité d'infection latente, nos travaux suggèrent qu'en l'absence d'autres processus spécifiques non inclus dans le modèle, la persistance d'infection sous forme latente peut être fragilisée sur de longues durées si le réservoir de cellules souches infectées est limité.

4.2 Article

Episome partitioning and symmetric cell divisions : quantifying the role of random events in the persistence of HPV infections

Auteurs: *Thomas BENETEAU*, Christian SELINGER, Mircea T. SOFONEA, Samuel ALIZON

DOI: [10.1371/journal.pcbi.1009352](https://doi.org/10.1371/journal.pcbi.1009352)

Abstract

Human Papillomaviruses (HPV) are one of the most prevalent sexually transmitted infections (STI) and the most oncogenic viruses known to humans. The vast majority of HPV infections clear in less than 3 years, but the underlying mechanisms, especially the involvement of the immune response, are still poorly known. Building on earlier work stressing the importance of randomness in the type of cell divisions in the clearance of HPV infection, we develop a stochastic mathematical model of HPV dynamics that combines the previous aspect with an explicit description of the intracellular level. We show that the random partitioning of virus episomes upon stem cell division and the occurrence of symmetric divisions dramatically affect viral persistence. These results call for more detailed within-host studies to better understand the relative importance of stochasticity and immunity in HPV infection clearance.

Author summary

Every year, infections by Human Papillomaviruses (HPV) are responsible for a large share of infectious cancers. The prevalence of HPVs is very high, which makes it a major public health issue. Fortunately, most HPV infections (80 to 90%) are cleared naturally within three years. Among the few that persist into chronic infections, the majority (88%) also regress. Hence for a given HPV infection, the risk of progression towards cancerous status is low. The immune response is often invoked to explain HPV clearance in non-persisting infections, but many uncertainties remain. Besides immunity, randomness was also suggested to play an important role. Here, we examine how random events occurring during the life cycle of the virus could alter the persistence of the virus inside the host. We develop a mechanistic model that explicitly follows the dynamic of viral copies inside host cells, as well as the dynamics of the epithelium. In our model, infection extinction occurs when all viral copies end up in differentiated cells and migrate towards the surface. This can happen upon cell division during the random allocation of the episomes (e.g. independent circular DNA copies of the viral genome) or when a stem

cell divides symmetrically to generate two differentiated cells. We find that the combination of these random events drastically affects infection persistence. More generally, the importance of these random fluctuations could match that of immunity and call for further studies at the within-host and the epidemiological level.

Introduction

Human Papillomaviruses (HPV) are the most oncogenic viruses known to humans [Plummer et al., 2016a]. Up to 8.6% of all cancers occurring worldwide in women are caused by HPV infections, versus 0.8% for cancers in men [de Martel et al., 2017]. Among the various types of cancers attributable to HPV, cervical cancers stand out since they are the third-most diagnosed form of cancer in women [Schiffman and Solomon, 2013].

The burden HPVs impose on human populations largely originates from their high prevalence in the general population. For example, it was estimated that the average lifetime probability (from sexual debut to age 70) of acquiring a genital HPV infection in the US was 84.6% for women and 91.3% for men [Chesson et al., 2014a]. But since up to 90% of these infections remain subclinical and are cleared in less than 3 years [Schiffman et al., 2011a], viral persistence, in particular for specific genotypes such as HPV-16, determines potential oncogenicity [Insinga et al., 2007].

Despite its ubiquity, the mechanisms underlying HPV infection clearance and persistence are still poorly understood. Although the innate immune response (e.g. natural killer cells) is known to be involved in chronic infections [Van Hede et al., 2014], few data exist for non-persistent infections. Adaptive immune responses (e.g. the role of T helper cells) have been studied in the context of vaccine trials, but the results remain inconsistent [Scott et al., 2001]. It has been noticed that people with recurrent genital warts were more likely to be HPV seropositive than those who do not have a history of genital warts [Carter et al., 1995]. However, more generally, the breadth and magnitude of adaptive immune responses following HPV infection are believed to be limited [Stanley, 2008] due to innate immune evasion mechanisms and a lack of inflammatory responses elicited by infected epithelial cells. Furthermore, in natural infections, there is no viremia (*i.e.* presence of viruses in the blood), and the production of detectable antibodies against the virus is low [Stanley, 2008]. Another longitudinal study investigated HPV-16 seroconversion statuses among various groups of infected women but did not reach any conclusive result, potentially by lack of statistical power [Farhat et al., 2009].

Finally, the inherently stochastic nature of the episomal life cycle of HPV, *i.e.* replication of the viral genome as independent circular DNA copies, could be an additional factor contributing to the clearance of HPV

infections. HPVs are non-lytic viruses that persist in the basal stem cells by replicating with host cell chromosomes [Moody, 2017]. In the basal layer, most cell divisions are asymmetric : one basal stem cell gives birth to one daughter stem cell, and one daughter differentiated cell upon cell division. Occasionally, especially in the context of a perturbation such as wounding [Doupe et al., 2012; Roshan et al., 2016], symmetric cell divisions occur that either yield two stem cells or two differentiated cells. Infection persistence in the epithelial tissue occurs when viral copies are transmitted to the daughter stem cell during cell divisions. Conversely, transmission to new hosts occurs via the viral copies in the daughter differentiated cell. These copies migrate and mature with the cell before being released to the surface upon shedding.

Symmetric divisions of infected basal stem cells into two daughter stem cells or two daughters differentiated cells [Clayton et al., 2007; Doupe et al., 2012] strongly impact the spread and persistence of the virus in the basal layer of the tissue. Since HPV infections begin with few infected cells [Doorbar, 2006], randomness in the type of cell division plays a crucial role in the clearance of HPV infections [Ryser et al., 2015]. A second source of stochasticity occurs within the nucleus of infected cells. During stem cell divisions, the number of episomes that are passed on to each daughter cell is random. Since cells are infected with a small number (potentially unique) of viral copies [Doorbar et al., 2012], stochasticity in episome distribution upon division could impact the persistence of HPV infection. The partition of viral plasmids in human cells is uncommon but not specific to HPV. This phenomenon is still not clearly understood despite its importance in the maintenance of infections and progression towards cancers [Chiu and Sugden, 2018]. To date, this aspect of HPV infections has not been taken into account. In a recent review [Ryser et al., 2017] on the use of mechanistic mathematical models applied to HPV infection, the authors globally call for more within-host modeling studies. While health-economic models have been extensively used to evaluate the cost-benefit of public interventions in the prevention of HPV infection or cervical cancer, mechanistic models have only been used to some specific topics (e.g. viral evolution induced by vaccine use [Murall et al., 2014, 2015] or impact of cell division in HPV natural infection [Ryser et al., 2015]).

Building on the work of Ryser [Ryser et al., 2015] that emphasized the contribution of symmetric cell divisions on HPV clearance, we develop an original model of HPV viral copies dynamic in the basal layer.

Taken separately, the intra- and inter-cellular stochastic processes fail to capture part of the life cycle of HPV infections. The intracellular approach does not account for the spread of the infection inside the epithelial tissue, whereas the inter-cellular approach ignores the necessity of viral persistence in proliferating cells. Thus, we hypothesize that combining the two aspects can provide additional insights into the dynamics of HPV infection.

We investigate the role of stochasticity in the clearance of HPV infections and quantify the role of both symmetric divisions and episomal partitioning in this process. We provide further evidence for the crucial effect of randomness in the clearance of HPV infections.

Methods

Our model focuses on the basal layer of a stratified squamous epithelium (SSE), but it can easily be adapted to other tissues as the HPV replication pattern is essentially the same in most sites [Rautava and Syrjänen, 2012; Stanley, 2012]. We only follow the dynamics of episomes in infected stem cells because SSE is in perpetual renewal and the turnover interval of the non-dividing cells lasts 3 to 6 weeks [Averette et al., 1970; Stanley, 2006, 2010]. Therefore, in the absence of cell-to-cell transmission, it is improbable that an infection could persist only in the differentiated layers of SSE. We also neglect potential re-infection of the SSE from virions released by desquamating cells because these are thought to be unlikely given the importance of micro-wounding in the establishment of new infections [Moody, 2017].

The life-cycle is divided into two steps following the non-lytic properties of HPV : 1) during the division of an infected cell, HPV episomes are distributed randomly between the two daughter cells, and 2) infected stem cells can divide symmetrically, thereby ending or, conversely, further spreading the infection. We graphically summarize each aspect of the model in the Figure A.16. Both the intra- and inter-cellular mechanisms are modelled as discrete-time branching processes [Harris, 1948; Athreya and Ney, 1972].

More formally, let $X_n, n \in \mathbb{N}$ be the total number of HPV episomes in all infected stem cells, n being the number of stem cell generations since the beginning of the infection. Unless stated otherwise, a time step corresponds to one cell generation. X_n is a random variable, the distribution of which is non-trivial, for it is constructed following several random events. We assume the random events between cell lineages, *i.e.* descending cells by asymmetric divisions, to be independent and identically distributed. Within a cell lineage, we assume other stochastic events to be identically distributed (but not independent). Independence is verified under certain specific conditions, for instance if we assume a cell lineage only divides asymmetrically.

Intracellular dynamics : viral distribution during cell division

HPVs are non-lytic viruses that follow the life cycle of their host cells. In basal (stem) cells, the number of viral copies is usually considered to be limited to several dozens of episomes per cell : 10 to 200 according to

Doorbar [Doorbar, 2006], and 50 to 100 according to more recent work [Moody, 2017]. We denote by C the maximum number of episomes per infected stem cell. We model intracellular dynamics as the superposition of two distinct random steps :

- (i) A division, where each episome in the dividing cell is distributed to the two daughter cells according to a Bernoulli distribution $Bernoulli(p)$. For asymmetric divisions, parameter p stands for the probability for an episome to be relocated in the daughter stem cell. Episomes distributed to the differentiated daughter cell are considered withdrawn from the system as they migrate to the surface without contributing to the infection of stem cells.
- (ii) An amplification, which occurs once the division over. For this random phenomenon we either assume a Dirac distribution $\delta(\lambda)$ or a Poisson distribution $Poisson(\lambda)$, with $\lambda \in \mathbb{N}$. We refer to λ as the episome replication factor.

Without prior knowledge of the biology of the amplification phase, the Dirac distribution appears to be the most parsimonious as it implicitly assumes that all viral copies exhibit the same average behavior. However, this assumption could be biased at the beginning of the infection as viruses remain in small numbers. In this case, the number of copies amplified could diverge from the average behavior due to individual variance. As little information is available in the literature, we choose a Poisson distribution as it only requires the same parameter as for the Dirac case.

The order of the two steps can be reversed without affecting the intrinsic characteristics of this branching process (see Appendix A.3 for more details). Since HPV genomes are present in low copy numbers and HPV gene expression is set at a minimum level in basal stem cells [Stanley, 2012], we assume that the number of HPV viral copies does not affect the dynamics of the host stem cells.

The infection can spread in the basal layer following symmetric divisions in two stem cells, a phenomenon governed by strong stochastic effects and detailed in the next section. To distinguish between daughter stem cells, we use the notation described in Appendix A.3 and define p as the probability for an episome to be distributed in the left (**L**) daughter stem cell. We assume episome distribution bias towards one of the daughter cells does not apply when the two daughter cells are stem cells, hence we assume distribution to be even ($p = 0.5$) during symmetric division.

Intra-host episome dynamics

In what follows, we group all the layers apical to the basal layer in a general population of differentiated cells. Let $(\Gamma_n)_{n \in \mathbb{N}}$ be the number of infected stem cells at time n . We denote by s resp. r the probability of a symmetric division to yield two stem cells resp. two differentiated cells.

We assume s and r to be small compared to the probability of asymmetric cell division. Following the work of [Ryser et al. \[2015\]](#), we model infected stem cells dynamics using the following branching process :

$$\Gamma_{n+1} = \sum_{i=1}^{\Gamma_n} \phi_i \quad (4.1)$$

where ϕ_i are independent and identically distributed random variables with the following distribution :

$$\phi_i = \begin{cases} 2 & \text{with probability } s \\ 0 & \text{with probability } r + \psi \\ 1 & \text{with probability } 1 - r - s - \psi \end{cases} \quad (4.2)$$

where ψ is a random variable accounting for intracellular effects leading to the end of the infection of the infected stem cell S . A histological study from 1970 estimated that $r + s \leq 0.08$ [[Averette et al., 1970](#)], and a more recent analysis also estimated 8% of cell divisions to be symmetric [[Clayton et al., 2007](#)].

Numerical simulations and sensitivity analysis

Some analytical insights can be obtained from the previous equations but in-depth investigations require numerical simulations. We estimated the cumulative probability of extinction (p_{ext}) given a set of parameters (see Table 4.1 for further details on the parameters used) by carrying out $\Omega = 10^3$ independent runs parametrized by the same parameter values. We also include the type of intracellular amplification in the analysis (Dirac or Poisson distributed). We displayed a random trajectory of episomal dynamic in the Figure A.16.

Using the Ω different trajectories, we calculate the cumulative probability of extinction $p_{ext}(t)$ as follows :

$$p_{ext}(t, \mathcal{S}) = \frac{1}{\Omega} \sum_{k=1}^{\Omega} \mathbb{1}_{\{X_t^k=0|\mathcal{S}\}} \quad (4.3)$$

where $X_t^k|\mathcal{S}$ is the total number of episomes at time t given parameters set \mathcal{S} for the k -th trajectories. If $\Gamma_0 > 1$, we assume that all infected cells initially harbor N_0 viral copies, hence a total of $X_0 = \Gamma_0 \times N_0$ episomes. p_{ext} is computed for all $t \leq 100$ cell generations. Based on data from the 1970s [[Averette et al., 1970](#)], we bound the rate of divisions between $[0.03, 0.07] \text{ day}^{-1}$, which means the upper bound of 100 cell divisions is sufficient to estimate infection clearance after 3 years. For each $t \in \mathbb{N}$, we estimate p_{ext} using Maximum Likelihood estimation for a Binomial proportion. For each of these proportion we computed the 95% confidence interval using Agresti-Coull method [[Agresti and Coull, 1998](#)].

TABLE 4.1 – **Parameters notation and description along with their biologically estimated ranges.** The parameters p, λ, C and N_0 govern the intracellular process, the three others are involved in the inter-cellular approach. Another parameter (not numerical) can be added to the list : the type of intracellular amplification (Dirac or Poisson distributed) as the two types slightly differ in their construction. Extensive analysis of the impact of each distribution on the probability of extinction has been conducted in Appendix A.3.

Notation	Description	Range	Source
p	Probability for an episome to be distributed to the daughter stem cell	$[0, 1]$	[McPhillips et al., 2005]
λ	Episome replication factor (number of viral copies)	$[2, 10]$	—
s	Probability of symmetric divisions into two stem cells	$[0.01, 0.04]$	[Averette et al., 1970; Clayton et al., 2007]
r	Probability of symmetric divisions into two differentiated cells	$[0.01, 0.04]$	[Averette et al., 1970; Clayton et al., 2007]
C	Viral capacity per infected stem cell	$[10, 200]$	[Doorbar, 2006; Moody, 2017]
N_0	Number of episome in each infected cells at the beginning of the epidemic	$[1, C]$	—
Γ_0	Number of infected stem cells at the beginning of the epidemic	≥ 1	—

To evaluate the role of each parameter on the probability of extinction, we perform a global sensitivity analysis using variance-based me-

thods (also known as Sobol indices). This approach decomposes the variance of the simulation outputs (here p_{ext}) into fractions that can be attributed to each of our parameters or interactions between them [Sobol, 2001; Saltelli et al., 2007]. To do so, we store parameter sets in two different matrices A and B of dimension $n \times 7$, where the columns correspond to the six parameters p, λ, s, r, C , and N_0 (see Table 4.1) and the type of intracellular amplification. We denote by $A_B^{(k)}$ the matrix A whose k -th column has been replaced by the k -th column of matrix B . We generate A and B sets using the Latin Hypercube Sampling (LHS) method implemented in the `lhs` package [Carnell, 2020] in R v4.0.3 [R Core Team, 2020b]. Each matrix contains $n = 500$ different parameter sets (more details can be found in Appendix A.3). We execute the global sensitivity analysis using `multisensi` package [Bidot et al., 2018]. We computed model outputs (i.e. extinction probabilities) for all sets $A, B, A_B^{(k)}, k \in \llbracket 1, 7 \rrbracket$ to obtain the following Monte-Carlo estimator for the variance of the extinction probability up to time t [Saltelli et al., 2010] :

$$\text{Var}_i(t) = \frac{1}{\Omega} \sum_{k=1}^{\Omega} p_{\text{ext}}(t, B)_k \left(p_{\text{ext}}(t, A_B^{(k)}) - p_{\text{ext}}(t, (A)_k) \right)$$

We do not include the initial number of infected stem cells (Γ_0) in this analysis, since the sensitivity of p_{ext} with respect to Γ_0 follows from the branching process property :

$$\mathbb{P}[X_t = 0 | \Gamma_0 = \nu] = \mathbb{P}[X_t = 0 | \Gamma_0 = 1]^\nu \quad (4.4)$$

Results

Our model contains two nested branching processes, which makes it difficult to obtain analytical results. However, analytical expressions can be derived if we neglect either of the two stochastic processes (see Appendix A.3 for detailed calculus and Figures A.14 and A.15 for numerical verification) and we use these results as boundaries to interpret the results of our numerical simulations and to determine how the cumulative probability of extinction (p_{ext}) varies with the different parameters. Simulations also allow us to explicit this direction of variation in p_{ext} with each parameter. Figure 4.1 displays the difference between probabilities of extinction for two parameter sets that only differ by the value of one parameter, we subtract the probability of extinction for the lowest of the two varying parameters to the probability of extinction of the highest of the two.

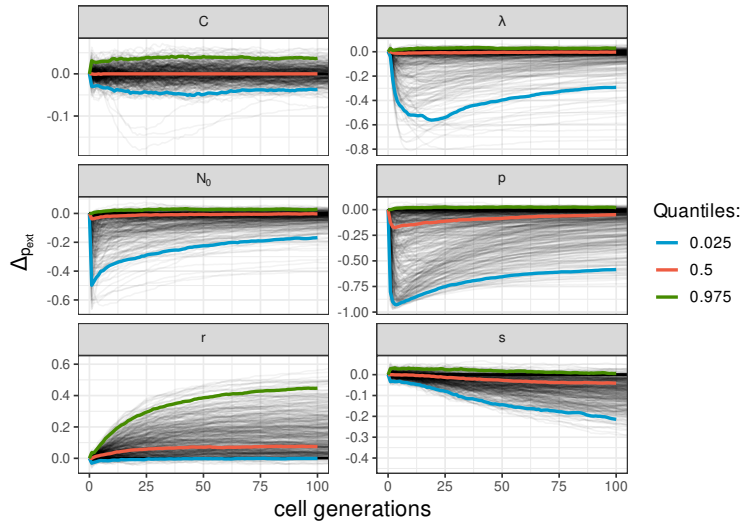


FIGURE 4.1 – **Difference in the cumulative probability of extinction for two sets of parameters that only differ by one value.** On each panel, the title indicates which parameter varies between the two sets. In each of this panel, we subtract the probability of extinction for the lowest of the two varying parameters to the probability of extinction of the highest of the two. We find consistent results as theory prediction : the probability of extinction decreases with λ, p, s, N_0 and increase with r . The viral capacity in cell (C), has little to no effect on the probability of infection. Line colors show the median (in red) and the first and third quartiles (in blue and green).

Intracellular stochasticity

If stochasticity only acts at the within-cell level, *i.e.* symmetric divisions are assumed not to occur, we retrieve classical results from Bienaymé-Galton-Watson (BGW) branching process theory [Athreya and Ney, 1972] : the extinction probability p_{ext} then decreases with the episome replication factor, λ , and with the probability for an episome to end up in the daughter stem cell, p (see Figure 4.2 for the Poisson case and Figure A.12 for the Dirac distribution). As λ increases, we observe that p_{ext} converges toward $1 - p$. Biologically, this means that the main event governing the probability of extinction is the first cell division and the distribution of the original episome into a differentiated or a stem cell.

Numerical simulations allow us to investigate the effect of intracellular processes while accounting for symmetric divisions. In particular, our sensitivity analysis highlights the effect of λ and p (Figure 4.3). At the beginning of the infection, intracellular phenomenon explains around 70% of the variance, mostly carried by the probability of allocation to daughter stem cell p . Indeed, in the early phase of the infection, the total number of episomes is usually low and, since symmetric divisions are rare ($r + s \leq 0.08$ [Averette et al., 1970]), the early dynamic of episomes is mostly

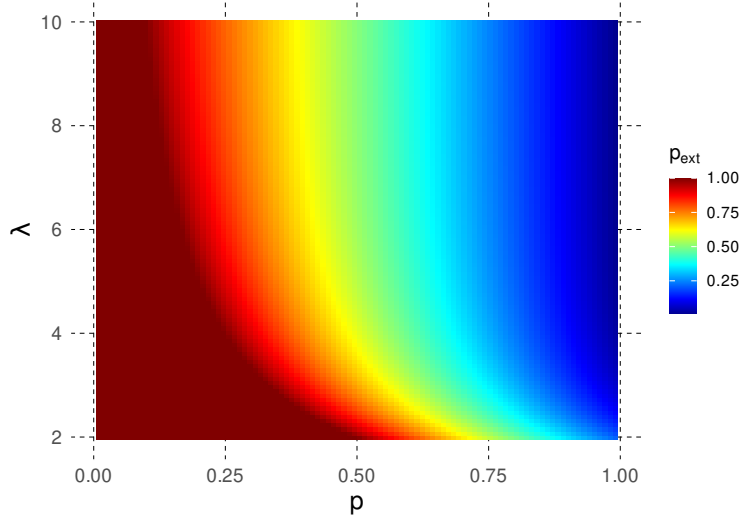


FIGURE 4.2 – **Effect of episode replication (λ) and asymmetry in episode partitioning (p) on the cumulative probability of extinction (p_{ext}).** We here assume a Poisson scenario. p_{ext} decreases with both λ and p , and reaches a threshold $p_{\text{ext}} = 1 - p$ for a given p when λ increases. This is consistent with the fact that if λ is sufficiently high, the main source of extinction is the first cell division of a stem cell containing 1 episode.

governed by intracellular process. We also retrieve the results from the analytical model, although the decrease in p_{ext} with increasing λ is more limited (Figure 4.1).

As explained in the methods, the numerical simulations slightly deviate from the analytical framework because the number of episodes per cell is assumed to be limited to C . However, we expect most of these results to hold in the general framework because we focus on a maximum time scale of 100 cell generations. We observe more deviations on that time interval when C is small and when the intracellular regime is slightly supercritical (further details can be found in Appendix A.3 and Figure A.12).

Inter-cellular stochasticity

When omitting intracellular stochasticity, theory allows us to predict that the branching process will go extinct with probability $p_{\text{ext}} = \min(r/s, 1)$, which we confirm with numerical simulations, see Figure A.15. Besides, we also find consistent results with theory as the estimated p_{ext} decreases (resp. increases) with s (resp. r), see Figure 4.1.

As intracellular stochastic processes only increase the risk for the infection to go extinct, we can lower bound the cumulative probability of extinction of our general process by the cumulative probability of extinction for the sole inter-cellular process, the proof is detailed in Appendix

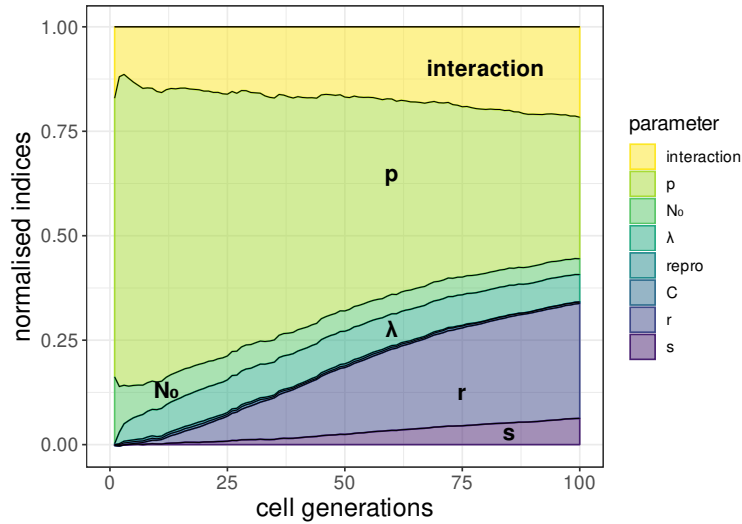


FIGURE 4.3 – **Variance-based sensitivity analysis for the cumulative probability of extinction (p_{ext})**. The normalized stacked first-order sensitivity indices ($\text{Var}_i/\text{Var}_{p_{\text{ext}}}$, with $\text{Var}_{p_{\text{ext}}}$ the total variance of p_{ext}) are displayed for all time steps and all parameters. The intracellular aspect dominates early in the infection. The importance of inter-cellular aspect increases over time, mostly through the probability of symmetric divisions in two differentiated cells. Intracellular viral capacity C and type of reproduction (e.g. Dirac or Poisson distributed) are barely visible due to their small effect on the variability of the outputs.

A.3. Generally speaking, no matter how rapidly the virus replicates at the intracellular level, its spread is bounded by the cellular dynamics of the host tissue. Therefore, we have

$$\mathbb{P}[X_\infty = 0 | \Gamma_0 = 1] \geq \min\left(\frac{r}{s}, 1\right) \quad (4.5)$$

Globally, these results suggest that symmetric divisions can favor the persistence of the virus depending on the local context of the tissue. In particular, when $s > r$, e.g. when the epithelium switches from a maintenance behavior to a repair behavior [Doupe et al., 2012; Roshan et al., 2016], symmetric divisions appear to favor the spread of the virus, i.e. its persistence in the tissue. Additionally we can notice in Equation 4.2 that if the inter-cellular process is critical (i.e. $r = s$), then the general process is subcritical. Indeed, under such circumstances we have

$$m := \mathbb{E}[\phi] = 1 - \psi \leq 1 \Rightarrow p_{\text{ext}} = 1$$

In Figure 4.3, the proportion of the variance explained by symmetric divisions increases over time. Indeed, on average the first symmetric division occurs after $(1/(r + s))$ cell generations. Hence, the early phase of the infection is mostly governed by intracellular aspect. But later in the infection, if the time between two symmetric events is sufficiently large,

the intracellular dynamic has almost surely converged to one of the two states $\{0, C\}$. Under such circumstances, only the inter-cellular is acting on the dynamic of the infection. We observe a significant difference in the proportion of variance explained by the parameter r compared with s , the latter having a smaller slope. This difference is directly induced by the intracellular stochasticity that can terminate infections prematurely.

Initial conditions

We then assess the effect on the probability of extinction of the initial number of episomes and their distribution in one or more stem cells. Let us assume the branching process to start with a total of $X_0 = \nu \times \omega$ episomes, $(\nu, \omega) \in \mathbb{N}^2, \nu > \omega$. We then look for the optimal distribution of these episomes from the virus' perspective. For this, we evaluate the probability of extinction under two extreme scenarios : i) ν cells each containing ω episomes, or ii) ω cells each containing ν viral copies. From branching theory, we know that for the inter-cellular level, the probability of extinction when starting with $\alpha \in \mathbb{N}$ identical and independent infected cells is equal to the probability of extinction of one cell raised to the power α .

Therefore, we only need to compute the probabilities of extinction for two sets of parameters that are identical, except for the number of episomes and number of infected stem cells at time $t = 0$, that are switched. We use numerical simulations to evaluate those probabilities, Figure A.3 shows the estimations of p_{ext} for six nearly-identical sets of parameters that only vary with respect to N_0 and Γ_0 .

To determine which scenario favors virus persistence we compare the cumulative probabilities of extinction for each of pairs of parameters that only differ by the initial arrangement of the viral copies. Figure 4.4 shows the quartiles for the two extreme scenarios. This indicates a benefit for the virus to spread its copies to the maximum number of stem cells at the start of the infection, the bigger the difference between ν and ω the more significant this benefit is.

Discussion

As immune mechanisms failed to answer completely the problem of spontaneous clearance in natural infections, Ryser [Ryser et al., 2015] has suggested random events in the type of cell divisions could induce extinction and thus complement explanation based on immunity. Following this approach, we refine the model initially exposed by Ryser [Ryser et al., 2015] and developed a model that also accounts for the risk of stochastic extinction due to virus partitioning. We hypothesized that an HPV infection could face extinction due to the sole effects of random processes occurring both at the intra- and inter-cellular level. Using a stochastic

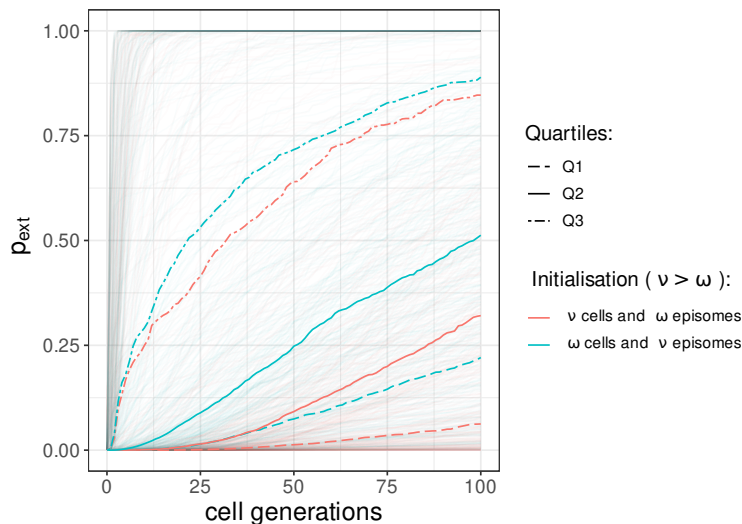


FIGURE 4.4 – **Estimation of the cumulative probabilities of extinction and their quartiles for the two extreme scenarios.** In red are displayed the estimations when the $N_0 \times \Gamma_0$ initial viral copies are allocated in ν cells, each containing ω episomes whereas in blue, the episomes are distributed in ω cells, each containing ν copies ($\nu > \omega$). The three quartiles are lower for the first scenario, thus indicating a lower risk of extinction when the copies are split in the maximum number of cells possible.

model comprised of two nested branching processes and numerical simulations, we find that random allocation of episomes during cell divisions drastically limit the persistence in the early phase of the epidemic. Later, if episomes persisted in some infected cells, randomness in the type of cell divisions can also lead to extinction, even though depending on the local context, symmetric divisions can also favor the spread of the virus in the tissue (when $s > r$). Notably, it was shown that the same cells can exhibit different proliferative patterns, switching from a balanced mode to an expanding mode if the tissue is wounded, and back to balanced mode once confluence is reached back [Doupe et al., 2012; Roshan et al., 2016]. While the precise molecular mechanisms underlying those shifts are still unknown, studies suggest that actin signaling could play a key role in the fate of the keratinocytes [Roshan et al., 2016]. The importance of each aspect varies over throughout the infection, and more generally as HPV viral copies increase in number, the role of stochasticity in the explanation of extinction fades. Our results show the first stages of the infection matter a lot for the persistence of the virus, which is particularly important for viruses like HPV that have a low number of copies initially [Doorbar et al., 2012].

Another outcome of our model is that many infections by HPV might clear naturally in a matter of weeks/months before even being noticed by the host immune system since episomes remain in small number in the

basal layer. Adding to the pioneering work of Ryser [Ryser et al., 2015] to highlight the role of stochastic processes in the clearance of HPV infections, we show that adding intracellular stochasticity greatly impacts extinction probability, especially in the early stages of the infection. Combining the two processes enables us to reach substantial clearance rates without involving immune response.

These results could explain why the antibody production against the virus is low after natural infections [Stanley, 2008; Doorbar et al., 2012], as the infection might just fade randomly in within the first divisions of the initially infected cells. HPVs might have evolve to limit pressures exerted by randomness during the first cell divisions by amplifying the viral copies in the early phase of the epidemic. Notably in the case of HPV16, the E1 protein seems to play a decisive role in the establishment of a persistent infection until the number of episomes per cell reaches host capacity but the same protein was dispensable for the maintenance of the infection. Our model corroborates these observations : we show that the early phase of the infection can be detrimental to the infection, E1 protein could favor the maintenance of the infection either by acting on the probability to be distributed to the daughter stem cell (p parameter) or promote episome amplification (λ parameter). Upon reaching host cell capacity, even if the intracellular regime is critical, it is very unlikely for the virus to go extinct, hence E1 protein is dispensable during maintenance phase of the infection [Egawa et al., 2012].

Papillomaviruses have evolved over million years infecting various animals' epithelium [Bravo et al., 2010]. From this long coevolution between the virus and its host, a large diversity of evolutionary strategies have emerged. Notably, HPV genotypes have developed specific host- and tissue-tropism that manifest through distinct pathologies [Mistry et al., 2008; Egawa et al., 2015; Doorbar et al., 2015]. HPV research mostly focuses on the Alpha genus since it contains all high-risk HPVs (HR-HPVs) that are susceptible to induce various forms of cancer in the genital areas and upper aerodigestive area [Gillison, 2000; Crow, 2012; Pytynia et al., 2014; de Martel et al., 2017]. The Alpha genus also contains genotypes that cause benign infections and are classified as low-risk HPVs (LR-HPVs). Although benign, LR-HPV infections can cause hyperproliferative lesions such as genital warts. These HPVs have been less studied, compared to HR-HPV, despite accounting for a bigger number of types [Egawa and Doorbar, 2017]. The distinct evolution of LR and HR-HPVs, which illustrate the persistence - virion production trade-off, has been well investigated in the literature [Schiffman et al., 2005; ?]. The role of the oncogenic proteins E6 and E7 have been put forward to explain the capacity of the HR-HPV to induce cellular immortalization and cellular progression towards cancer development [Tomaić, 2016; Mirabello et al., 2017]. However, it is unclear how these differences could impact the early phases of the infection and interfere with the stochastic pressures acting on the virus.

Globally our results suggest that remaining with a low copy number is a way to evade immunity [Doorbar, 2013], but such a strategy is risky for the virus is under strong stochastic forces that plummet the persistence of the infection. Viruses committed to such strategy might alleviate these pressures exerted by randomness using proteins in the very early phase of the infection, as the E1 protein for HPV16 [Egawa et al., 2012]. These proteins could as such, be a viable candidate for potential treatment. Yet, these proteins might not be the most adequate targets for antiviral drugs as papers showed that the role of these elements is crucial in the very early steps of the infection but lose importance once host capacity is reached [Egawa et al., 2012]. We might not be able to detect HPV sufficiently early to inhibit E1 protein to prevent the persistence of infections.

Ryser [Ryser et al., 2015] assume a balance between the two types of symmetric divisions, thus restraining the inter-cellular dynamic to the critical regime. We allow the possibility for the tissue to deviate from that assumptions. Notably, we allow the probability of symmetric divisions in two stem cells to exceed the probability of symmetric divisions in two differentiated cells as keratinocytes can alternate between two modes of proliferation depending on local context. If local confluence is removed by scratch injury, cells switch to the expanding mode, characterized by an excess of cycling cells production, until confluence is reached again, after which they switch back to a balanced mode where similar proportions of proliferating and differentiating cells are produced, thus generating population asymmetry maintaining epidermal homeostasis [Roshan et al., 2016]. The two modes of proliferation could differ by the rates of cell divisions or the probability of division in two stem cells (parameter s), or both. Our model predicts that HPVs greatly benefit from the expanding mode, which echoes previous empirical work on the link between HPV persistence and wound healing responses [Ledwaba et al., 2004; Doorbar, 2006; Fuchs and Nowak, 2008; Schiller et al., 2010].

To further calibrate the models and delineate the true role of stochasticity, a comparison with longitudinal follow-ups of women recently infected by any type of HPV or presenting chronic latent infections would be of primer interest. Ideal time windows between two measurements should not exceed 1 month. Besides, we should perhaps reconsider how we label infections that cleared in less than 1 month. Many follow-ups would treat them as "carriage", i.e. HPV from the partner and not a "true" infection. Perhaps there are more non-persistent infections than we thought? Overall to decipher this issue, we need accurate infection duration distributions over short time scales, ideally with virus loads measures.

Comparing longitudinal follow-ups data and numerical simulations would require clarifying the estimation of the time between two cell divisions. We rely on old measures [Averette et al., 1970] to set the upper bound of the number of cell generations to compute in our three years simulations. This work estimates the rate of divisions around [0.03, 0.07]

day⁻¹. These results are coherent with more recent work on the kinetics of epidermal maintenance stating an overall division rate of 1.1 per week [Klein et al., 2007]. To fit his model to follow-ups data, Ryser [Ryser et al., 2015] estimated the rate of division using proxies (height of the stratified squamous epithelium expressed in number of cells, renewal time of the tissue, fraction of proliferative cells in the basal layer) and find a much higher division rate [0.29,1] day⁻¹. Updates on the estimation of this variable would be of primary interest to compare theoretical and experimental approaches.

Acknowledgments

The authors acknowledge the IRD itrop HPC (South Green Platform) at IRD Montpellier for providing HPC resources that have contributed to the research results reported within this paper.

CHAPITRE

5

Modélisation de la première vague de COVID-19 en France



5.1 Aperçu de l'article

Ce travail a été initié lors de la première vague épidémique de COVID-19 en France (printemps 2020) et se focalise sur les caractéristiques de ce premier épisode. Nous avons cherché à estimer en premier lieu la date de début de l'épidémie en France et à la comparer avec d'autres estimations se basant sur des méthodes différentes.

Le premier confinement d'une durée de 55 jours (du 17 mars au 11 mai 2020) a permis une réduction importante de la propagation virale en France, mais à sa levée, le virus était toujours en circulation sur le territoire [Roques et al., 2020; Salje et al., 2020; Sofonea et al., 2021; Forien et al., 2021]. Certaines mesures instaurées ont été maintenues à sa sortie (port du masque, distanciation), auxquelles s'est rapidement ajouté un dépistage massif. Ce dernier fait partie de la stratégie plus globale de test-traçage-isolation [dec, 2020; Ashcroft et al., 2022] qui, avec les mesures de distanciation et d'hygiène, permet de freiner l'épidémie [Contreras et al., 2021]. Bien que moins efficace qu'un confinement, ces mesures ont permis de limiter la propagation du virus [Pullano et al., 2021; Bosetti et al., 2021]. Comme le virus continuait de circuler, nous avons également cherché à évaluer la date d'extinction de l'épidémie après la première vague et le risque de rebond épidémique dans un contexte de levée de confinement. Enfin, nous avons examiné le gain qu'aurait apporté la mise en place d'un confinement plus précoce sur le temps nécessaire à l'éradication de l'épidémie.

Chaque individu est différent dans sa réponse à un agent infectieux. Par exemple, l'infectiosité varie d'une personne à l'autre. Lorsque l'épidémie se propage et que le nombre de personnes infectées est grand, on peut faire une approximation par la loi des grands nombres et simplifier le processus en remplaçant l'hétérogénéité individuelle par la valeur moyenne à l'échelle de la population. Dans ces cas-là, on peut donc décrire la dynamique de l'épidémie à l'aide d'approches déterministes. Néanmoins, durant les phases d'émergence et d'extinction d'épidémie, la dynamique épidémique est le fait d'un nombre limité de chaînes de transmission et ces approches perdent de leur pertinence. Il est ainsi plus judicieux de se tourner vers un formalisme stochastique [Lloyd-Smith et al., 2005].

Durant les premiers mois d'épidémie en France, les données sur l'état de l'épidémie étaient limitées. Nous ne disposions que des seules données hospitalières SARS-CoV-2 (admissions tous services, admissions en soins critiques, décès hospitaliers) et peu d'information sur les dynamiques de transmission pour la population française. Nous avons donc modélisé l'incidence de COVID-19 à travers une suite discrète de tirages de Poisson paramétrés par les niveaux d'incidences passés, le taux de contact dans la population et le nombre de reproduction du virus. La loi de Poisson (nommée en référence à Poisson [1837]) peut se voir comme la limite d'un nombre infini de tirages de lois binomiales dont la probabilité de réussite est faible ($p \rightarrow 0$). Cette loi décrit ainsi le nombre de réussites d'un évè-

nement rare (ici la transmission du pathogène) sur un très grand nombre de répétitions (ici les contacts entre deux individus). C’est une loi assez simple avec un seul paramètre, généralement noté λ , dont les moments sont facilement calculables, l’espérance et la variance étant toutes deux égales à λ .

En épidémiologie, le nombre moyen d’infections secondaires causées par un individu est généralement décrit à l’aide du nombre de reproduction de base R_0 . À partir des seules données de séries temporelles, il n’est pas possible de reconstruire les chaînes de transmission effectives pour calculer directement ce R_0 . On utilise alors des estimations sur la distribution de l’intervalle sériel, temps entre les symptômes du donneur et ceux du receveur, comme indicateur du temps de génération (temps entre deux infections) [Wallinga and Lipsitch, 2006] rarement observé chez l’humain pour des raisons éthiques et pratiques.

Dans notre modèle, nous nous sommes basés sur les estimations de Nishiura et al. [2020b] concernant la distribution de l’intervalle sériel et sur les travaux parallèles, à la fois déterministes et non markoviens, de l’équipe pour l’estimation des paramètres épidémiques clefs [Sofonea et al., 2020; Danesh et al., 2020]. Plusieurs méthodes sont possibles pour incorporer le nombre de reproduction dans le modèle. En premier lieu, on peut opter pour une approche globale où l’on fixe le R_0 pour tous les individus de la population (c’est-à-dire en simplifiant par la loi des grands nombres). Cette approche ne permet cependant pas de prendre en compte la diversité des capacités infectieuses entre individus. Une approche alternative consiste à spécifier une distribution pour le nombre de reproduction individuel en choisissant une moyenne égale au R_0 dans la population générale. Diverses lois de probabilités peuvent être pertinentes. Nous avons opté pour une loi Gamma de moyenne R_0 et un paramètre de forme issus d’une épidémie de SARS-CoV [Lloyd-Smith et al., 2005] qui donne des estimations cohérentes avec les premiers travaux sur l’épidémie de COVID-19 [Althouse et al., 2020a; Endo et al., 2020; Liu et al., 2020b; Sun et al., 2020]

Nos résultats suggèrent une survenue de l’épidémie en France autour de la mi-janvier. Pour atteindre l’extinction de l’épidémie après la première vague, nos analyses suggèrent une durée moyenne d’un peu moins de 200 jours après la levée du confinement, dans le scénario le plus restrictif. On estime que la mise en place de mesures de confinement un mois plus tôt aurait permis de réduire ce temps d’extinction à 95 jours, soulignant la nécessité de réagir rapidement dans un contexte d’émergence épidémique (sous l’hypothèse de croissance exponentielle).

Nos estimations sur la survenue de l’épidémie en France sont cohérentes avec celles d’autres travaux de l’équipe [Sofonea et al., 2020; Danesh et al., 2020; Sofonea et al., 2021] et d’autres équipes [Salje et al., 2020; Roques et al., 2020; Bosetti et al., 2021; Forien et al., 2021]. Nos travaux illustrent aussi l’importance de la stochasticité dans la fin d’épidémie, phénomène peu étudiée dans ce contexte. Les projections sur l’ar-

rêt de l'épidémie ont été bousculées par l'arrivée du variant Alpha en octobre 2021. En effet, au moment de la rédaction du manuscrit, on ne pouvait inclure des phénomènes peu soupçonnés comme l'émergence de variants au R_0 plus élevé [Gaymard et al., 2021].

5.2 Article

Estimating dates of origin and end of COVID-19 epidemics

Auteurs: *Thomas BENETEAU*, Baptiste ELIE, Mircea T. SOFONEA, Samuel ALIZON

DOI: [10.24072/pcjournal.63](https://doi.org/10.24072/pcjournal.63)

Abstract

Estimating the date at which an epidemic started in a country and the date at which it can end depending on interventions intensity are important to guide public health responses. Both are potentially shaped by similar factors including stochasticity (due to small population sizes), superspreading events, and ‘memory effects’ (the fact that the occurrence of some events, e.g. recovering from an infection, depend on the past, e.g. the number of days since the infection). Focusing on COVID-19 epidemics, we develop and analyse mathematical models to explore how these three factors may affect early and final epidemic dynamics. Regarding the date of origin, we find limited effects on the mean estimates, but strong effects on their variances. Regarding the date of extinction following lockdown onset, mean values decrease with stochasticity or with the presence of superspreading events. These results underline the importance of accounting for heterogeneity in infection history and transmission patterns to accurately capture early and late epidemic dynamics.

Ce papier a également fait l’objet d’un rapport publié en français et disponible sur HAL [[Bénéteau et al., 2020](#)].

Introduction

The ability to make robust epidemiological inferences or predictions strongly relies on the law of large numbers, which buffers the variability associated with individual processes. Many models of infectious diseases spread are deterministic and therefore assume that the number of infected hosts is large and above what has been termed the ‘outbreak threshold’ [Hartfield and Alizon, 2013]. This assumption is violated at the beginning and end of an epidemic, where stochasticity may have a strong effect [Britton and Scalia Tomba, 2019].

In this study, we tackle two issues. First, we wish to estimate the date of origin of an epidemic in a country, focusing on the case of COVID-19 outside China. This question is important because the infection being imported, some cases may be detected before the reported beginning of an epidemic wave, which is somehow counter-intuitive to an audience not familiar with stochasticity. Furthermore, transmission often takes place before an epidemic wave is detected, as shown in several places using SARS-CoV-2 genomic data, e.g. Washington state in the USA [Bedford et al., 2020] or France [Danesh et al., 2020]. Second, we investigate how many days strict control measures need to last to ensure that the prevalence falls below key thresholds. Despite its public health implications, this latter question has rarely been investigated. There are some exceptions, for instance in the context of poliomyelitis [Eichner and Dietz, 1996], Ebola virus disease [Thompson et al., 2019], and MERS [Nishiura et al., 2016] epidemics. However, these estimates neglect superspreading events and/or do not include non-Markovian effects (i.e. memory effects). Indeed, they often rely on ordinary differential equations, meaning that the probability of an event to occur (e.g. recovering from an infection) does not depend at all on the past (e.g. the number of days since the infection started). Recently, however, it has been shown that incorporating secondary cases heterogeneity can significantly lower the delay until an Ebola virus disease outbreak can be considered to be over [Djaafara et al., 2020].

Maintaining the lockdown so as to reach ‘zero-COVID’ requires extended effort because the incidence might oscillate at a low value due to stochasticity for a long period. However, in practice, and as illustrated by several countries, lockdown measures could be eased after the epidemic reaches a sufficiently low incidence. Indeed, when the number of cases is low enough, stricter contact tracing, as well as local control measures can be sufficient to stop the virus spread. For instance, in Taiwan or South Korea, the epidemic was controlled for months as long as the incidence was kept below 20 new cases per day [Max Roser and Hasell, 2020]. In New Zealand, control measures were lifted only when the incidence reached 2 cases per day. This is why we investigate the time for incidence to reach given thresholds that can be greater than 0.

The COVID-19 pandemic led to an unprecedented publication rate

of mathematical models, several of which involve stochasticity. For instance, [Hellewell et al., 2020] analysed the initial steps of the outbreak to estimate the fraction of the transmission chains that had to be tracked to control the epidemics. Their results depend on the value of the basic reproduction number (denoted R_0), which corresponds to the mean number of secondary infections caused by an infected individual in an otherwise fully susceptible population [Anderson and May, 1991], but also on individual heterogeneity. Indeed, if few individuals tend to cause a large number of secondary infections while the majority tends to cause none, the probability of outbreak emergence is much lower than if all individuals cause the same number of secondary infections [Lloyd-Smith et al., 2005]. Accounting for this property, a study used the early COVID-19 outbreaks incidence data in different countries to estimate the dispersion of the distribution of individual R_0 [Endo et al., 2020]. Finally, [Althouse et al., 2020b] have also used stochastic modelling to explore the role of super-spreading events in the pandemic and its consequences on control measures.

Here, we develop an original discrete stochastic (DS) model, which features some of the known characteristics of the COVID-19 epidemics. The model is non-Markovian, which means that individual histories matter for the dynamics. More specifically, the probability that an event occurs (e.g. infecting another host) depends on the number of days spent in a state (e.g. being infected). Furthermore, following earlier studies [Hellewell et al., 2020], we account for the fact that not all hosts transmit on the same day post-infection. This is captured by assuming a distribution for the generation time, which is the time between infection dates of an ‘infector’ and an infected person. Since the time of infection is complicated to estimate, we approximate the generation time by the serial interval, which is the time between the onset of the symptoms in the ‘infector’ and that in the infected person [Nishiura et al., 2020a; He et al., 2020]. We also allow for heterogeneity in transmission patterns by assuming a negative binomial distribution of the secondary cases. To investigate the importance of stochasticity, we had to use deterministic models in addition to ours. To have memory effects in a deterministic setting, we reanalysed an earlier non-Markovian model [Sofonea et al., 2020] by setting the date of origin of the epidemic as the main free parameter. Finally, to remove both memory effects and stochasticity, we analyse a classical deterministic Markovian model, which is commonly used to analyse COVID-19 epidemics [Grant, 2020].

By comparing the outputs of these models, we explore the importance of stochasticity, individual heterogeneity, and non-Markovian effects on the estimates of the dates of origin and end of a nation-wide COVID-19 epidemic, using France as a test case and mortality data because of its extensive sampling compared to case incidence data.

Methods

The Discrete Stochastic (DS) model

Our model simulates the number of newly infected individuals per day (i.e. the daily incidence) as an iterative sequence following a Poisson distribution. We assume that the average number of secondary cases is equal to R_0 and that the host population is homogeneously mixed (i.e. no spatial structure). These assumptions are relevant if a small fraction of the population is infected [Trapman et al., 2016].

More specifically, each individual is assumed to cause a random number of secondary infections throughout his/her infection, depending on his/her infectiousness (β). Here, infectiousness represents the relative infectious contact rate of an individual. It summarises both biological aspects (efficiency of transmission per contact, susceptibility of the recipients), and the contact rate of the individual, during the whole infectious period. Secondary infections occur randomly several days after contracting the disease. The probability of infecting someone some days after getting the disease is captured by the generation time, which we approximate using the serial interval [Nishiura et al., 2020a].

Let ω_a be a random variable describing the probability of infecting someone a days after contracting the disease. An individual infected since a days infects new individuals at a rate $R_0 \times \beta \times \omega_a$ during that day. Therefore, the number of secondary infections occurring a days after being infected, which is considered as a count of independent events, follows a Poisson distribution parameterized by $R_0 \times \beta \times \omega_a$. From the additive property of the Poisson distribution, we find that the mean number of secondary cases during the entire infectious period is equal to the individual infectiousness. We then repeat this process for all individuals to determine the disease global progression.

Let Y_t be the random variable describing the incidence, i.e. the number of new infections, on day t , t being the number of days since initialisation of the process. The sequence of Y_{t+1} is defined using the Poisson additive property :

$$Y_{t+1} \sim \text{Poisson} \left(R_0 \eta_t \sum_{i=0}^t \omega_{t-i} \sum_{k=1}^{Y_i} \beta_{k,i} \right) \quad (5.1)$$

where η_t is the average normalized contact rate in the population at day t , $\beta_{k,i}$ is the infectiousness of individual k , infected at day i , and ω_{t-i} is the probability of an individual infected at time i to infect someone at time t ($t - i$ is the age of the infection).

We consider two scenarios (a) without and (b) with individual heterogeneity. If we denote by \mathcal{B} the distribution of random variables $\beta_{x,y}$, accounting for the infectiousness of an individual x infected at day y , then, in each scenario we assume that :

- a) \mathcal{B} is a Gamma distribution with shape parameter $k = 0.16$ and mean R_0 , implying that individuals are heterogeneous in infectiousness and/or contact rate, which can lead to ‘superspreading’ events. We use the shape parameter (k) value estimated for a SARS outbreak in 2003 [Lloyd-Smith et al., 2005], which is consistent with estimates for SARS-CoV-2 epidemics [Endo et al., 2020; Althouse et al., 2020b; Liu et al., 2020a; Sun et al., 2020].
- b) \mathcal{B} is a Dirac distribution, noted $\delta(1)$, implying that there is no heterogeneity and individuals have the same infectiousness and infection duration distribution. This is equivalent to $k \rightarrow +\infty$ in the previous scenario. The sequence $(Y_t)_{t \in \mathbb{N}}$ then simplifies into :

$$Y_{t+1} \sim \text{Poisson} \left(R_0 \eta_t \sum_{i=0}^t \omega_{t-i} Y_i \right) \quad (5.2)$$

To model the control intensity over the epidemic at time t such as, for instance, a national lockdown, we vary the contact rate parameter η_t . We assume that η_t is piecewise constant and that its discontinuities capture changes in public health policies (see Figure A.18).

Overall, we define the temporal reproduction number (R_t) at time t such that

$$R_t = \eta_t \mathbb{E}[\mathcal{B}] = \eta_t R_0 \quad (5.3)$$

Beginning of the epidemic wave

To infer the starting date of the epidemic wave, we run our discrete stochastic (DS) algorithm starting from one infected individual until the infection dynamic becomes deterministic, *i.e.* the law of large numbers applies. We set the mortality incidence threshold to 100 daily cases, which was reached on 23 March 2020 in France. Neglecting the delay from infection to death, this would correspond to a daily incidence of more than 11,000 new cases according to the infection fatality ratio ; a value much higher than the outbreak threshold above which a stochastic fade-out is unlikely [Hartfield and Alizon, 2013]. We use independent estimates for the other parameters and perform a sensitivity analysis, shown in the Appendix.

To simulate death events in the DS model, we use the infection fatality ratio p [Verity et al., 2020], *i.e.* the proportion of those infected who will go on to die from that infection. If we write X_t the number of individuals infected at time t who will die :

$$X_t \sim \text{Binomial}(Y_t, p) \quad (5.4)$$

For each of the X_t individuals, the day of death is set by drawing a time from infection to death following θ , *i.e.* a Gamma distribution. θ was previously estimated on French hospital data [Sofonea et al., 2020]

(Table A.3), and its estimate compare very well with other independent estimates made from contact tracing data [Linton et al., 2020].

We repeat the algorithm 10,000 times to obtain a stable distribution of starting dates and discard epidemics that die out before reaching the threshold incidence. To allow for comparison with empirical data, we compute a sliding average of this time series over a 7-days window.

Finally, we assume that the consequences of the lockdown, which was initiated in France on March 17, did not affect the death incidence time series until the very end of March because of the delay between infection and death, which we estimate in France to be more than 11 days for 95% of the cases [Sofonea et al., 2020].

End of the epidemic wave

Here, we estimate how many additional days of lockdown would have been necessary to reach epidemic extinction for various lockdown intensity. Using the case of France as an example, the estimated lockdown contact rate is $\eta_{FR} = 0.243$, and we start our simulation on May 11, when the lockdown measures were partially lifted (*i.e* 55 days of lockdown). To avoid the accumulation of uncertainties, we initialise the model with incidence values obtained from a discrete-time non-Markovian model [Sofonea et al., 2020] for the past 15 days before the start of the simulation, in France. This threshold arises directly from the choice of the serial interval distribution : 99.9% of the transmissions occur within less than 15 days, using the generation time (Table A.3).

We then use a Monte-Carlo procedure to estimate key features of the time series $(Y_t)_t$, such as the mean extinction time or the cumulative extinction probability. This is done by running 10,000 independent and identically distributed simulations of our process for each set of parameters.

We analyse the 10,000 resulting trajectories as follows. First, we estimate the distribution of τ , which is the random variable corresponding to the minimal lockdown duration (in days) such that the incidence is always null afterward for various scenarios. To mimic what happened during the first lockdown we set the contact rate to η_{FR} for the first 55 days. We then set the contact rate to a fixed value (greater or equal than η_{FR}) until extinction is reached. As long as the effective reproductive number is lower than 1, the time to extinction is finite. Mathematically,

$$\tau = \inf_{s \in \mathbb{N}} \{Y_k = 0; \forall k \geq s\} \quad (5.5)$$

Second, we study the effect of finite lockdown extensions on the probability of extinction to understand the risk of epidemic rebound upon lockdown lifting. For simplicity, we assume that control measures are completely lifted once the lockdown is over. The probability of having no new cases at time t ($p_0(t)$) is estimated using the following formula

$$p_0(t) = \frac{1}{N} \sum_{k=1}^N \mathbb{1}_{\{Y_t^k=0\}} \quad (5.6)$$

where N is the number of simulations performed and Y_t^k the number of newly infected individuals in the k^{th} simulation at time t .

Third, we study the effect of initiating the first lockdown one month or two weeks earlier (in France, on February 17 or March 03 respectively) on the distribution of the time to extinction (τ). For comparison purposes, we assume that in any case the first 55 days of lockdown have the same contact rate ($\eta_{t \leq 55} = \eta_{\text{FR}}$) and then extend the lockdown indefinitely with variable intensities to estimate the time to extinction (τ) as described previously (equation 5.5).

Alternative models

To further study the effects of stochasticity, non-Markovian dynamics, and superspreading, we implemented two deterministic models. The first is Markovian, i.e. memoryless, and is based on a simpler model derived from a classical SEIR model. The second has a discrete-time structure, which allows capturing non-Markovian dynamics [Sofonea et al., 2020].

The SEAIRHD model

In this classical compartment model, hosts can belong to seven states : susceptible to infection (S), exposed (i.e. infected but not infectious, E), asymptomatic and infectious (A), infectious and symptomatic (I), removed (i.e. recovered or isolated, R), hospitalised who will die (H), or dead (D) (Fig. A.19). The model is described by a set of ODE detailed in the Appendix (equation system A.20). Since the model is deterministic, we can seed the simulations with a single exposed individual on day t_0 .

This model is solved numerically using the `Numpy` package in Python 3.8.3 to obtain a deterministic trajectory. Parameters were chosen with maximal likelihood given the observed daily mortality data, assuming that the daily mortality incidence is Poisson distributed, and independence between daily incidences (For more details, see the supplementary material). We also simulate a stochastic version of this model 1,000 times using a Gillespie algorithm with the package `TiPS` [Danesh et al., 2023b] in R v.3.6.3 [R Core Team, 2020a].

COVIDSIM : A non-Markovian deterministic model

Finally, we use an existing discrete-time model that has a similar structure to the continuous model mentioned above with an additional age-structure [Sofonea et al., 2020]. For comparison purposes, the generation time is set to be the same as in our DS model [Nishiura et al., 2020b],

and so the (non-exponential delay) from infection to death. However, two major differences are that this third model is not stochastic and does not allow for superspreading events. We restricted the parameter inference to the daily hospital mortality data described previously, with the main free parameter being the date of origin. We invite the reader to refer to [Sofonea et al., 2020] for the scripts and further details on this approach.

Model calibration

To allow for model comparison and improve estimates, we fix some key parameters based on existing values, focusing on the French COVID-19 epidemic. Table A.3 lists all the parameters used along with key references.

We compute the likelihood of the deterministic SEAIRHD model assuming a Poisson distribution of the daily mortality incidence data. Parameter inference with maximum likelihood is performed using the Nelder-Mead algorithm implemented by `Scipy.minimize` function in Python.

The parameters used for the non-Markovian deterministic model correspond to the maximum likelihood set of parameters used in [Sofonea et al., 2020].

Code and simulation results availability

The different scripts and simulation results are available on Gitlab : <https://gitlab.in2p3.fr/ete/origin-end-covid-19-epidemics>

Results

Origin of the epidemic wave

When neglecting host heterogeneity, using our DS framework, the median delay between the importation of the first case of the epidemic wave and the time mortality incidence reaches 100 deaths per day (March 23) is 67 days (equivalent to a first case on January 16 in France), with a 95% confidence interval (95% CI) between 62 and 79 days, *i.e.* between January 4 and 21 in France (Fig. 5.1). With this model, only 7% of the outbreaks die out before reaching the threshold.

Superspreading events, *i.e.* when the individual infectiousness \mathcal{B} follows a Gamma distribution, seem to have limited effects on these results : the median delay drops slightly to 64 days (January 19 in France), although with a larger 95% CI, between 54 and 85 days. Moreover, as expected [Lloyd-Smith et al., 2005], we observe a soar in the frequency of epidemic outbreaks dying out before reaching the threshold, which represent 75% of our simulations.

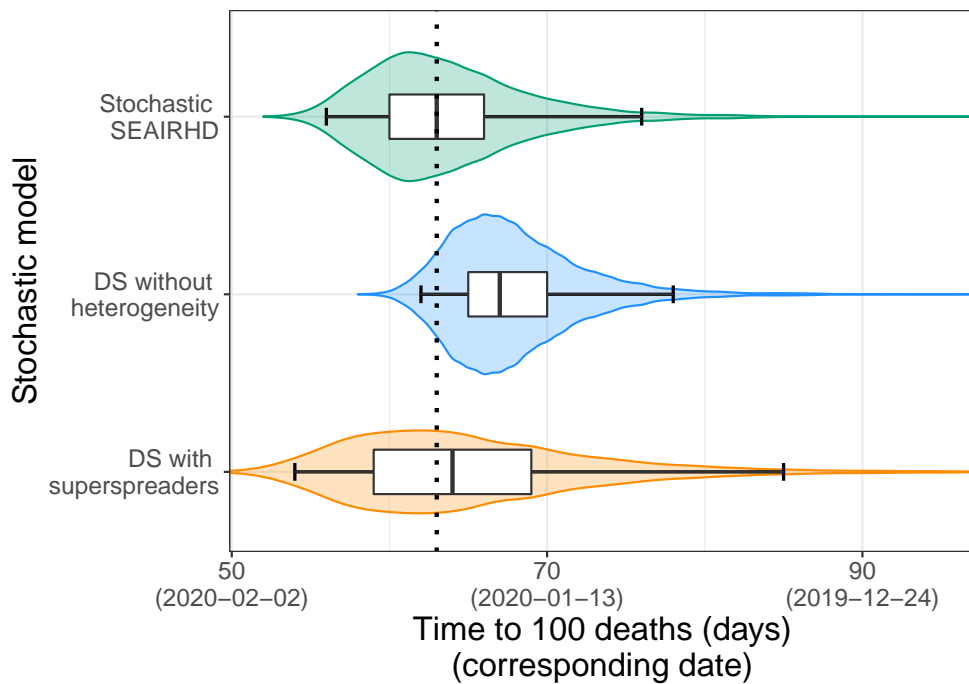


FIGURE 5.1 – **Estimated distribution of the number of days until daily mortality incidence reaches 100 cases.** The boxplots and the whiskers indicate the 2.5%, 25%, 50%, 75%, and 97.5% quantiles out of the 10,000 simulations. The red dashed line shows the estimates using the deterministic models.

When assuming deterministic and Markovian dynamics with our SEAIRHD model, the importation date of the first case of the epidemic wave that best fits the results is similar, with a delay of 63 days until daily mortality incidence reaches 100 cases. A stochastic implementation of the same model yields the same median delay of 63 days [95% CI : 56 - 76 days], which is comparable to the DS model. However, consistently with earlier studies [Sofonea et al., 2020; Grant, 2020], the ability of this memoryless model to capture the data is limited (Fig. A.20 in the Appendix). Finally, the maximum likelihood parameter estimates from a deterministic but non-Markovian model, COVIDSIM [Sofonea et al., 2020], restricted to the mortality data, indicates a similar delay of 63 days (January 20), [95% CI : 63 - 64 days].

We perform a sensitivity analysis of our results focusing on two parameters. First, we show that the median delay for daily mortality to reach 100 cases is increased by 5 days when the generation time standard deviation is decreased by one third (Fig. A.22). Therefore, the estimates remain within the confidence interval obtained for the starting date of the epidemic. Second, increasing the number of initially imported cases from 1 to 5 decreases the delay by 7 days, with a median of 60 days [95% CI : 57-64 days] without heterogeneity. However, when assuming a more realistic scenario where all those cases are not imported on the same day,

this impact of the delay was more limited (Fig. A.23). For example, if the 5 cases are imported during the first five days of the outbreak, the decrease is only of 5 days, with a median delay of 62 days [95% CI : 59 - 66 days].

Overall, non-Markovian dynamics or stochasticity do not tend to strongly impact the estimate of the delay for an epidemic to reach a daily mortality incidence of 100 cases. Introducing super-spreading events, however, slightly decreases the delay estimated and greatly increases its variance. As expected, the initial number of imported cases can have an impact on the estimates.

End of the epidemic wave with lockdown

Time to eradication

We estimate the distribution of the minimal lockdown duration to eradicate the epidemic (τ) by first neglecting superspreading events and starting from the end of the first-wave lockdown in France on May 11 (orange violins in Figure 5.2). When maintaining the constraints on social interactions to their full intensity ($\eta_{t>55} = 0.24$), a total of at least 8 months of lockdown, including the 55 days between Mar 17 and May 11, are required to reach a 97.5% extinction probability.

When accounting for individuals heterogeneity, we find that, everything else being equal, the quantiles of the time to eradication (τ) are always lower than the homogeneous cases. However, 7.23 months of lockdown at full intensity ($\eta_{t>55} = 0.24$) are still required to guarantee extinction in 97.5% of the cases (blue violins in Figure 5.2). Accounting for individual heterogeneity also reduces the variance of τ . This is expected because in this case, the majority of the infected people do not transmit, which increases the extinction probability [Lloyd-Smith et al., 2005].

The mean values of the time to eradication (τ) increases with the decrease in the intensity of the lockdown constraints post 55 first days of lockdown. As the contact rate of the population tends towards $1/R_0$ the mean values of τ diverge towards infinity. The dynamical process is said to be critical (resp. super-critical) if $\eta_t = 1/R_0$ (resp. $\eta_t \geq 1/R_0$). This result holds when assuming transmission heterogeneity.

We also compute the time to extinction with the deterministic SEAIRHD model after tuning the model using the parameters that best fitted the mortality incidence (Fig. 5.2). The time to extinction corresponds here to the minimum time where the incidence reaches zero.

Rebound risk

In our stochastic model, the incidence at time t (denoted $(Y_t)_{t \in \mathbb{N}}$) can alternate between zero and non-zero values. To evaluate the risk of epidemic rebound, we implement a finite lockdown extension after which

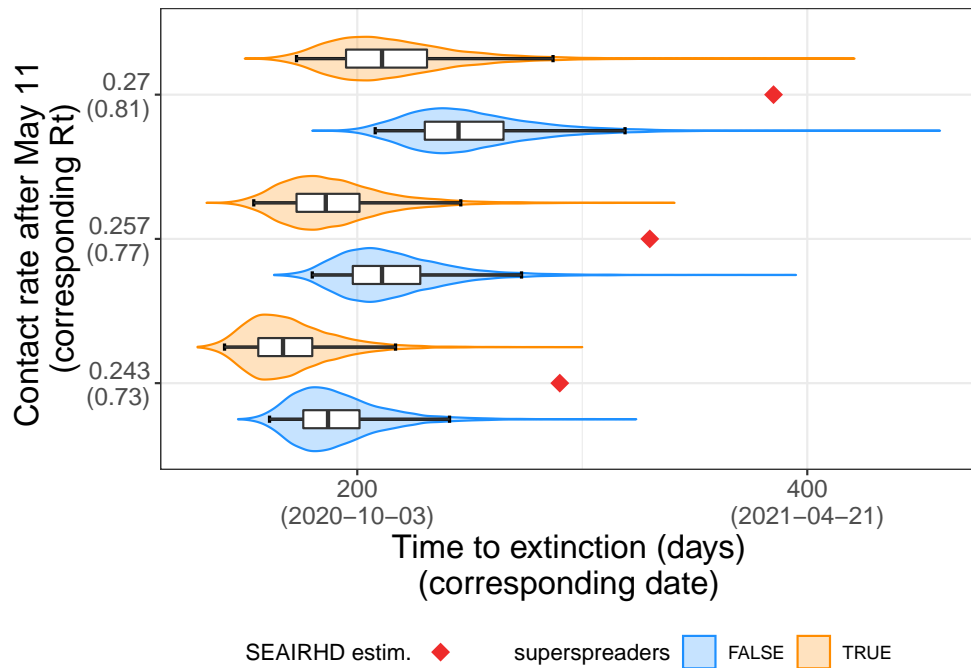


FIGURE 5.2 – **Effect of lockdown intensity, stochasticity, and superspreading events on the time to extinction (τ)**. The distributions of the time to extinction (in days since the start of the lockdown on March 17) for several lockdown intensities (η_t) after the first 55 days (i.e. after 11 May 2020) are plotted on the Y-axis using violin plots and boxplots. Results without transmission heterogeneity ($\mathcal{B} = \delta(R_0)$) are in orange. In blue, we assume a Gamma distribution for \mathcal{B} . Red diamonds show results from the deterministic Markovian model. The box extends from the lower to upper quartiles of the data. The whiskers expand from the 2.5% to the 97.5% quantiles.

all constraints are released ($\eta_t = 1$). This allows us to calculate $p_0(t)$, the probability to have 0 new cases after time t . In Figure A.24, we see a sharp decrease in $p_0(t)$ a few days after lockdown release.

The rebound risk is directly linked to the transmission heterogeneity. Assuming a higher individual transmission heterogeneity (i.e. lower k) drastically reduces the risk of rebound, as it also implies that most infectees do not transmit the disease.

Eradication and lockdown initiation date

We now turn to the consequence of implementing a lockdown a month or two weeks earlier. In France, this corresponds to Feb 17 and Mar 03 (at that time, a total of respectively 1 and 3 deaths were reported).

The results are shown in Figure A.25 for the case without host heterogeneity and Fig. 5.3 with superspreading events. Initiating the lockdown one month earlier, i.e. for France approximately 33 days after the onset of

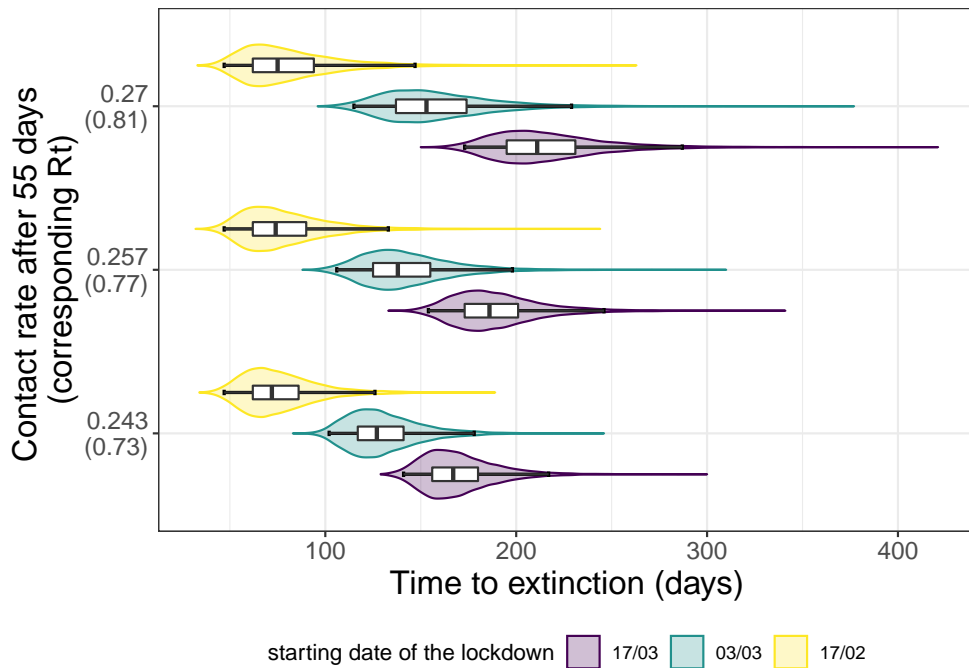


FIGURE 5.3 – **Effect of the lockdown intensity, stochasticity, and initiation date on the time to extinction (τ) under individual spreading heterogeneity assumption.** The distributions of the time to extinction (in days since the start of the lockdown) for several contact rate restrictions post 55 first days are plotted on the Y-axis using violin plots and boxplots. In this graph, we assume individual spreading heterogeneity. The colors indicate the different initiation date of the lockdown : in purple it starts on Feb 17, green Mar 03, and yellow on Mar 17 (official start). The box extends from the lower to upper quartiles of the data. The whiskers expand from the 2.5% to the 97.5% quantiles.

the epidemic wave, decreases the 97.5% quantile of the time to extinction by 91 days with transmission heterogeneity (97 days without heterogeneity) in the most restrictive scenario. If the onset of the lockdown is brought forward by two weeks (March 3rd), *i.e.* in France approximately 48 days after the onset of the epidemic, 97.5% of the extinction events occur before the 178th day of lockdown with transmission heterogeneity (199th day without heterogeneity). Hence a reduction of 39 (resp. 42) days of lockdown could be expected compared to the actual start (Mar 17).

These numbers increase with the easing of the constraints following the first 55 days of strict lockdown ($\eta_t = 0.24$). When assuming a lighter control in the following days (e.g. $\eta_{t>55} = 0.29$), one can notice that the increase in the quantiles of τ when starting the lockdown on Feb 17 is much lower than the two other cases.

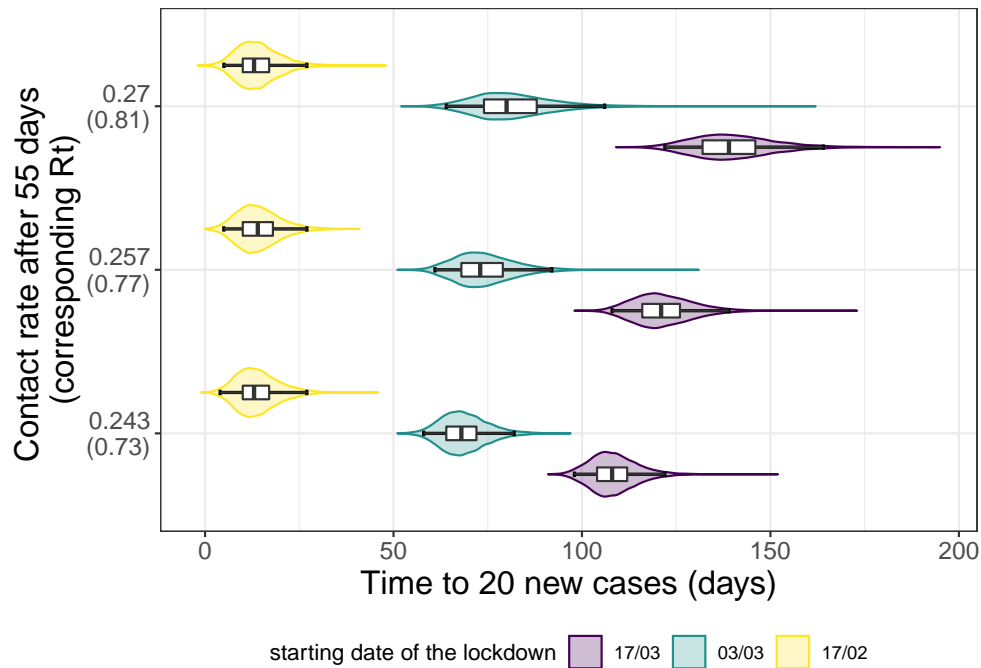


FIGURE 5.4 – **Effect of the lockdown intensity, stochasticity, and initiation date on the time to 20 new cases under individual spreading heterogeneity assumption.** The distributions of the time to 20 new cases (in days since the start of the lockdown) for several contact rate restrictions post 55 first days are plotted on the Y-axis using violin plots and boxplots. In this graph we assume individual spreading heterogeneity. The colors indicate the different initiation date of the lockdown : in purple it starts on Feb 17, green Mar 03 and yellow on Mar 17 (official start). The box extends from the lower to upper quartiles of the data. The whiskers expand from the 2.5% to the 97.5% quantiles.

Time to a threshold of 20 new cases per day

Finally, we study the distribution of the delay to reach 20 new cases per day, below which it is expected that a general lockdown is not required to control the epidemic. We evaluate the effect of lockdown intensity, initiation date and individual spreading heterogeneity on this delay.

The estimated distributions of the time to 20 new daily cases when accounting for superspreading events is displayed in Figure 5.4 (see Figure A.26 for the estimations without superspreaders). Our model suggests that initiating control measures one month earlier (mid-February) would have reduced the 97.5% quantile of the time to 20 new cases by 95 days under the strictest restrictions. In the mid-February scenario, we notice the time to 20 new cases occurs during the 55 first days of lockdown. Starting the lockdown early March does reduce the 97.5% quantile of the time to the threshold by 40 days. However, the first 55 days of lockdown are not sufficient to reach the limit of 20 new cases per day.

Discussion

In the early and final stages of an epidemic, stochastic forces may strongly affect transmission dynamics because infection prevalence is low. Using stochastic mathematical modelling, and assuming $R_0 = 3$, we estimate the time for a COVID-19 epidemic to reach an incidence of 100 deaths per day to be approximately 67 days, with a 95% probability between 62 and 79 days. In the case of France, where such incidence values were reached on Mar 23, this translates into an origin of the epidemic around January 16, with 95% probability between January 4 and 21. This is consistent with estimates obtained using virus genome data, although these should be interpreted with caution due to the uncertainties regarding the molecular clock estimates for the virus and the incomplete sampling in France [Danesh et al., 2020].

Accounting for superspreading events yields a later median date of origin (January 19 for France). This is expected because, in outbreaks that do not die out, superspreading events accelerate the initial dynamics [Lloyd-Smith et al., 2005]. However, this difference is not significant.

The 95% CI for the epidemic starting date generated by our different models overlap. This could originate from our use of mortality data. Since death occurs after a mean delay of 23 days after infection [Sofonea et al., 2020], by the time mortality incidence is detectable, transmission dynamics are largely deterministic. This also explains why introducing superspreading events mostly affects the variance of the estimate. Unfortunately, hospital admission data is not available for France until 18 March 2020, and screening data was initially performed with a very low sampling rate in the country (only severe cases were tested).

Care must be taken when comparing the estimates from our discrete stochastic model to that of earlier models. For instance, the non-Markovian deterministic COVIDSIM model [Sofonea et al., 2020], which estimates the date of onset to be slightly later (January 20), includes host age structure. Regarding the more classical deterministic and Markovian SEAIRHD model, its ability to fit the data is limited (Fig. A.20), except when only considering the exponential phase before the lockdown. This poor inference of underlying epidemiological dynamics is likely due to the absence of memory in the underlying processes, as stressed by earlier studies [Sofonea et al., 2020; Grant, 2020]. When incorporating memory on the hospitalization-to-death delay, we obtain a much better fit, and the time to the daily mortality of 100 cases is then comparable to that of the model without superspreading events.

We also estimate the median number of days of full intensity lockdown required to achieve extinction with a 95% confidence. In the French setting (i.e. introduction of the lockdown after 67 days of the epidemic), we find with our stochastic model that 187 (95% CI : [161, 241]) days of lockdown would be required to reach extinction in a homogeneous transmission scenario in 50% of the cases. Accounting for superspreading

events decreases the median estimate value by 20 days. Initiating the lockdown one month earlier strongly affects these estimates : a 30 days anticipated start reduces the mean number of days spent in full intensity lockdown by 95 days, i.e. a 51% reduction.

50% of the simulations reach the threshold of 20 new cases after 108 (95% CI : [98, 122]) days of lockdown at full intensity initiated mid-March. When initiating the constraint in mid-February, this threshold is reached in 13 (95% CI : [4, 27]) days. Since, in the latter scenario, the epidemic spread is more limited, the first 55 days of lockdown are decisive in the slowing down of the epidemic. This confirms that early interventions have a disproportionate impact on the epidemic dynamic.

Finally, we investigated the risk of an epidemic rebound upon lockdown lifting. In this scenario, super-spreading has a striking impact in limiting this risk, which is consistent with earlier work on outbreak emergence [Lloyd-Smith et al., 2005].

There are several limitations to this work. First, the generation time ω and the time from infection to death θ , remain largely unknown in France, as well as in many countries. Most serial interval estimates rely on contact tracing data from Asia [Liu et al., 2020a; Nishiura et al., 2020a], which could differ from the distribution in France, due to differences in contact structure, or non-pharmaceutical interventions. Although the generation time distribution is expected to affect epidemic dynamics, we show in Figure A.21 that the variance of this interval has little impact on our results.

Another important limitation about the estimation of the date of origin of the epidemic comes from the assumption that only a single infected person caused the epidemic. Most epidemics outside China were seeded by multiple importation events. The problem is that there is an identifiability issue because it is impossible to estimate both the number of initial infected cases and the time to a threshold of 100 deaths with incidence data only. However, some estimates made in the UK from phylogenetic data as well as the combination of prevalence and travel data show that the estimated number of importation events is less than 5 per day before the end of February, when the virus was beginning to circulate at higher levels throughout Europe [du Plessis et al., 2021]. Assuming a similar importation pattern in France, we show that the dynamic is only sensitive to the importation events within the first days after the beginning of the epidemic wave. While these events may have helped the epidemic to escape the stochastic phase faster, they are unlikely to strongly affect the estimated date of the beginning of the wave (Figure A.23). In a quite extreme scenario of 5 importations per day during 30 days, we estimate the median day of the epidemic beginning to be 16 days later (*i.e.* Feb 2 for France).

Another limitation comes from the lack of data regarding individual heterogeneity in COVID-19 epidemics. Such heterogeneity was supported by early limited data [Endo et al., 2020; Liu et al., 2020a] but recent ad-

ditional evidence from Chinese transmission chains further supports this result [Sun et al., 2020], although with a higher k parameter value than the one used here (0.30 versus 0.16 here), meaning a less heterogeneous transmission. Therefore, our assessment of superspreading events impact seems conservative.

These results have several implications. First, they can help reconcile the fact that cases may be detected long before the emergence of the transmission chains related to an epidemic wave. This is particularly important for an audience not familiar with stochasticity. Second, the estimate of the time required to ensure that the epidemic is gone can help inform public health decisions. In the case of the French epidemic, for instance, enforcing a strict lockdown from March 17 until epidemic extinction was practically unfeasible. However, this may not be the case if measures are taken early enough in the epidemic. Furthermore, our work also illustrates the risk of epidemic rebound as a function of the duration of the lockdown. Overall, this work calls for further studies, especially to assess the importance of super-spreading events in the global spread of SARS-CoV-2.

Acknowledgements

Thomas Beneteau is supported by a doctoral grant from the Ligue Contre le Cancer. We thank the CNRS, the IRD, and the IRD itrop HPC (South Green Platform) at IRD montpellier for providing HPC resources that have contributed to the research results reported within this paper. We also thank the ETE modelling team for discussion. The authors of this preprint declare that they have no financial conflict of interest with the content of this article. This article was submitted on medRxiv [Bénéteau et al., 2021]. Version 3 of this preprint has been peer-reviewed and recommended by *Peer Community In Mathematical and Computational Biology* (<https://doi.org/10.24072/pci.mcb.100004>).

CHAPITRE

6

Conclusion et perspectives



Dans ce manuscrit, j'ai détaillé trois approches de modélisation des infections par les papillomavirus. En particulier, j'ai utilisé trois formalismes différents pour décrire la dynamique des infections transitoires, évaluer la plausibilité de scénarios biologiques relatifs à la latence et proposer de nouvelles pistes d'explications sur les mécanismes responsables de l'extinction de ces infections.

À partir de l'étude clinique PAPCLEAR, nous avons montré que les épisodes de détection HPV sont très courants chez les jeunes adultes et sont généralement courts. Nos données, comme la littérature [Moscicki et al., 2013; Trottier et al., 2010], suggèrent que ces épisodes de détection traduisent préférentiellement de nouvelles acquisitions, sans que l'hypothèse de reprise d'infection latente ne puisse être totalement écartée.

Combinant les données de PAPCLEAR avec celles issues de suivis vaccinaux menés par *GlaxoSmithKline*, j'ai comparé des modèles de la dynamique infectieuse incorporant explicitement une phase latente à l'aide de chaînes de Markov cachées. Sans pour autant statuer sur l'existence de ces phases latentes, notamment chez les jeunes adultes, notre approche permet de quantifier leur implication sur la persistance de l'infection dans plusieurs scénarios de dynamique naturelle des infections HPV.

La persistance des infections HPV repose sur le maintien de copies virales dans les couches basales de l'épithélium. À l'aide d'un modèle stochastique, nous avons mis en lumière l'impact des événements aléatoires, aux échelles intra-cellulaires et inter-cellulaire, sur la clairance de ces infections. Cette approche fournit une piste d'explication de la clairance des infections dans les scénarios où la piste immunitaire n'est pas satisfaisante.

Les papillomavirus sont une préoccupation majeure en santé publique à cause de leur implication dans divers cancers, notamment celui du col de l'utérus. De fait, cela a orienté massivement les travaux sur le sujet [Goldie et al., 2006] et la modélisation des infections HPV reste encore majoritairement focalisée sur des approches cliniques d'évaluation de mesures sanitaires et d'identification de groupes à risques [Ryser et al., 2017].

Malgré l'existence d'un vaccin efficace et sûr, la couverture vaccinale demeure faible à l'échelle mondiale, avec de grandes disparités entre pays riches et pays plus défavorisés [Bruni et al., 2021]. Parmi les pays à hauts revenus, la France présente un taux de couverture vaccinale HPV particulièrement faible avec moins de 30 % des femmes ayant reçu la première dose et moins de 25 % la deuxième [Bruni et al., 2021].

En plus de la vaccination, le dépistage du cancer du col de l'utérus est un levier préventif efficace car il permet un traitement [Wilailak et al., 2021]. La survenue du cancer étant progressive, on peut surveiller ses stades d'avancement par l'étude des lésions précancéreuses lors de tests cytologiques. La majorité de ces lésions précancéreuses se résolvent spontanément [Doorbar et al., 2012], mais lorsqu'elles persistent, il est nécessaire de détruire ces cellules ou de les enlever pour éviter la survenue

de complications.

Depuis 2006, la vaccination est la pierre angulaire des politiques de prévention sur le long terme des cancers du col de l'utérus [Wilailak et al., 2021], car elle seule permet une diminution durable. Toutefois, les politiques de vaccination s'inscrivent sur la durée, les premiers résultats de ces politiques publiques ne sont visibles qu'après plusieurs décennies, ce qui n'est pas sans poser des difficultés pour convaincre l'opinion publique [Jit et al., 2014]. Il est ainsi nécessaire de maintenir la confiance et favoriser l'adhérence des personnes concernées à ces politiques publiques. Cependant, on dispose déjà d'éléments tangibles de l'efficacité des programmes de vaccination. Par exemple, en Australie, premier pays à mettre en application ces programmes à l'échelle nationale en 2007 [Patel et al., 2018], des travaux de modélisations estiment que le cancer du col de l'utérus devrait être éliminé d'ici 2035 [Hall et al., 2019].

Plusieurs autres obstacles se dressent devant la mise en place de politiques publiques ambitieuses contre les cancers induits par les HPV. D'abord, ces campagnes de prévention sont coûteuses, tant dans l'approvisionnement en vaccins [Hull et al., 2020] que dans la logistique pour les amener aux personnes visées [Wilailak et al., 2021]. Pour favoriser l'accès du vaccin aux enfants/adolescent·e·s concerné·e·s, les campagnes de vaccination s'effectuent généralement de concert avec les établissements scolaires. Cela nécessite la mise en place de mesures coûteuses, difficilement soutenables par les pays à revenus limités [Bruni et al., 2022].

De plus, même dans des situations où des campagnes de prévention existent, il faut encore s'assurer que les personnes ciblées par ces mesures acceptent d'y prendre part. Plusieurs facteurs peuvent influencer l'adhérence aux politiques publiques.

En premier lieu, il est important d'apporter à toutes et tous une information claire et précise. Le manque de connaissances sur les ratios bénéfice-risque de la vaccination HPV, ainsi que sur le cancer du col de l'utérus en général, est une cause majeure de la faible adhésion aux campagnes préventives. Les campagnes de vaccination ciblant principalement les enfants et adolescent·e·s, de préférence avant le début des relations sexuelles, l'aval des parents est généralement nécessaire. Les parents peuvent cependant avoir un a priori négatif, craignant pour la santé de leurs enfants et s'inquiétant de potentiels effets secondaires de la vaccination [Bonanni et al., 2018]. Certains perçoivent également le vaccin comme une incitation à la sexualité [Mattheij et al., 2012], notamment non-protégée [Bastani et al., 2011; Berenson and Rahman, 2012] alors que plusieurs études montrent que la vaccination n'induit pas de changements dans les pratiques sexuelles [Donken et al., 2018; Brouwer et al., 2019].

De plus, les infections par HPV comptent parmi les IST les plus courantes, mais la connaissance du public sur ce virus est limitée. En premier lieu, on note un manque de connaissance sur le lien entre le virus et le cancer du col de l'utérus [Gerend and Magloire, 2008; Marlow et al., 2013;

Cocchio et al., 2020] et sur l'association entre HPV et verrue génitale [Cocchio et al., 2020]. Contrairement à d'autres IST, la transmission du virus n'est que partiellement bloquée par le port de préservatifs [Winer et al., 2006]. Or, ce fait est généralement mal connu et dans une cohorte italienne, près de 75 % des personnes interrogées estiment le préservatif aussi efficace que la vaccination contre la transmission des HPV Cocchio et al. [2020]. Plus généralement, on note un manque d'information sur l'existence des mesures existantes ou sur les conditions d'accès aux politiques de prévention [Idehen et al., 2017; Marlow et al., 2015; Olsson et al., 2014].

Les études montrent qu'un niveau d'éducation plus faible est généralement associé à une moindre connaissance et adhésion aux politiques de prévention [Idehen et al., 2017; Marlow et al., 2015]. Cependant, ce critère ne suffit pas à lui seul à expliquer les disparités observées [Roland et al., 2016].

Un autre obstacle majeur, notamment dans les populations défavorisées, est la barrière de la langue, par exemple dans les populations immigrées [Idehen et al., 2017; Marlow et al., 2015; Abdullahi et al., 2009; Chen, 2013; Jacobs et al., 2005; Olsson et al., 2014].

Certaines croyances ou habitudes culturelles peuvent également freiner l'accès aux mesures préventives [Idehen et al., 2017]. En particulier, les frottis cervicaux peuvent être perçus comme gênants, voire stressants [Marlow et al., 2015]. Le dépistage urinaire peut ainsi constituer une alternative moins chère, moins intrusive et nécessitant moins de personnel médical aux frottis cervicaux [Pathak et al., 2014].

Favoriser l'accès à l'information est donc crucial pour augmenter l'efficacité des politiques vaccinales, qu'elles soient contre les HPV ou le SARS-CoV-2 [Haug et al., 2020]. C'est un processus long, qui implique de multiples acteurs [Cocchio et al., 2020]. Les supports papiers peuvent aider à la diffusion de l'information et au rappel des mesures existantes, mais plusieurs études pointent vers l'importance du dialogue avec un soignant [Marlow et al., 2015; Cocchio et al., 2020], notamment pour répondre aux interrogations ou apporter un soutien supplémentaire [Roland et al., 2016]. À l'heure du numérique, les applications téléphoniques, apparaissent aujourd'hui également comme des alternatives prometteuses pour la promotion de la vaccination chez un public adolescent [Woodall et al., 2021; Teitelman et al., 2020]. Pour les parents, les doutes et questionnements peuvent être levés en informant le-la patient·e et son entourage lors d'entretien avec des professionnels. Par exemple, dans une cohorte indienne, seuls 40 % des parents acceptaient spontanément de vacciner leurs enfants, mais ce chiffre grimpeait à 80 % après une courte session d'information [Chatterjee et al., 2016].

Même chez les personnes sensibilisées aux problématiques liées aux HPV, un test de dépistage positif, par rapport à un test cytologique, peut engendrer plus d'anxiété ou autres conséquences psychologiques néfastes. Le caractère sexuellement transmissible de l'infection peut aussi

être perçu comme stigmatisant [McCaffery et al., 2006].

Les différents travaux épidémiologiques des dernières décennies ont permis de mettre en évidence la prévalence élevée des infections HPV, tant chez la femme que chez l’homme [Bruni et al., 2023]. De plus, si divers travaux ont mis en lumière la courte persistance des infections HPV chez l’humain [Insinga et al., 2007; Schiffman et al., 2011a; Ramanakumar et al., 2016], les travaux de modélisation mathématique mécaniste sur la dynamique naturelle du papillomavirus humain demeurent limités [Ryser et al., 2017]. Ceci est particulièrement vrai pour les infections transitoires, généralement asymptomatiques, qui font l’objet de moins d’attention comparativement aux infections chroniques. Pourtant, bien que courtes, elles représentent la grande majorité des infections chez l’humain et leur compréhension pourrait aider à mieux cerner la biologie de l’infection et donc améliorer la prévention ou le soin [Alizon et al., 2017].

Récemment, la vision classique de la persistance des infections HPV a été questionnée par la possibilité d’entrée en phase latente. Sur la base d’études animales [Amella et al., 1994; Zhang et al., 1999; Maglennon et al., 2011, 2014], un modèle de latence du papillomavirus a été proposé dans lequel les papillomavirus peuvent être maintenus sous contrôle dans les cellules de la couche basale en faible nombre par le système immunitaire [Gravitt, 2012]. Ces infections peuvent redevenir productives, en particulier dans le cas de stimulations UV ou suite à des irritations mécaniques [Gravitt, 2011]. Les mécanismes moléculaires et immunitaires à l’origine du passage en latence sont encore mal connus chez l’humain [Doorbar, 2023], où la fréquence de la latence demeure une question ouverte [Malagón et al., 2023].

Dans la *Logique du Vivant* [Jacob, 1987], François Jacob écrit « pour qu’un objet soit accessible à l’analyse, il ne suffit pas de l’apercevoir. Il faut encore qu’une théorie soit prête à l’accueillir. [...] C’est elle qui détermine la forme de la question, donc les limites de la réponse ».

Dans ce contexte, en particulier lorsque la détection *in vivo* du virus est difficile et incertaine, la modélisation fournit un cadre solide pour la proposition et l’évaluation de scénarios explicatifs [Keeling and Rohani, 2008; Djidjou-Demasse et al., 2020].

La mise en évidence du rôle de certains papillomavirus humains dans la survenue de divers cancers, en particulier le cancer du col de l’utérus, remonte maintenant à plusieurs dizaines d’années [Boshart et al., 1984; zur Hausen, 1996; Mork et al., 2001]. Des progrès ont été accomplis dans la compréhension des phénomènes cancérigènes à l’échelle moléculaire. En particulier, l’augmentation de l’activité des gènes E6 et E7 favorise la croissance cellulaire, inhibe la différenciation et induit une instabilité chromosomale [Graham, 2017]. Dans la grande majorité des cancers (80 à 85 % exactement), ce changement d’expression des oncoprotéines résulte de l’intégration du génome viral dans celui de l’hôte [Pett and Coleman, 2007], mais les mécanismes conduisant à cette intégration sont

encore mal connus [Groves and Coleman, 2015]. Le schéma classique de progression cancéreuse fait intervenir une infection chronique HPV persistante sur plusieurs années dans les couches basales [Graham, 2017]. Néanmoins, des incertitudes demeurent quand à la transition vers les phases cancéreuses. Notamment, on compte plusieurs dizaines d’années entre les premières acquisitions HPV et la survenue de cancers [Schiffman and Solomon, 2013], durée bien supérieure aux estimations actuelles pour les durées d’infections HPV. Nos résultats suggèrent que même avec un compartiment latent, il est peu probable que les infections persistent sur plusieurs dizaines d’années. Néanmoins, nous nous sommes limités à une population assez jeune. Il serait très intéressant de réaliser des analyses statistiques similaires sur les phases latentes pour des populations plus âgées afin de corroborer ou ajuster nos estimations. Il est également possible que les cancers HPV puissent survenir suite à une exposition prolongée, mais interrompue, au virus. Dans tous les cas, la modélisation mathématique s’avère décisive pour proposer de nouveaux scénarios explicatifs robustes [Servedio et al., 2014] mais aussi pour comparer quantitativement et qualitativement les modèles mathématiques entre eux.

Grâce une bourse pour la mobilité financée par la Société Européenne de Biologie Évolutive, j’ai pu construire un projet avec Katrina Lythgoe sur les infections chroniques par le virus de l’hépatite B (HBV). Notre but est de modéliser la clairance des infections HBV sous traitement par analogue de nucléot(s)ides. Le virus de l’hépatite B a un cycle de vie propre aux *Hepadnaviridae*, impliquant plusieurs formes virales. En particulier, le virus persiste sous forme d’ADN circulaire fermé de manière covalente (ADNccc) dans le noyau [Liaw and Zoulim, 2015]. Le traitement par analogue de nucléot(s)ide ne permet pas d’éliminer le virus du noyau, mais il inhibe l’étape de transcription inverse, qui est indispensable à la formation de nouvelles particules virales. Avec le traitement, la charge virale des patients redescend rapidement à des niveaux nuls et limite la progression de la maladie vers des formes graves. Mais le traitement doit être pris à vie car la charge virale remonte rapidement à toute interruption [Trépo et al., 2014]. En effet, chez les patients traités, on n’observe que rarement (autour de 1 % par an) une clairance de l’infection [Downs et al., 2019]. À l’aide de modèles stochastiques inspirés des approches développées dans le chapitre 4, nous cherchons à évaluer des scénarios susceptibles d’expliquer cette persistance des infections HBV sous traitement de longue durée. En particulier, nous explorons la possibilité de persistance du virus dans une sous-population d’hépatocytes se divisant plus lentement que la majorité des autres cellules du foie [Lythgoe et al., 2021]. Ce projet aurait dû avoir lieu en hiver 2020, mais la survenue de la pandémie a retardé le début de la collaboration.

Comme beaucoup de personnes, la survenue du COVID-19 a bousculé le cours de ma thèse et m’a invité à dévier momentanément de mes travaux sur les papillomavirus. Durant le premier confinement, je me suis écarté de mes virus de prédilection, les oncovirus, pour travailler

avec les autres membres de l'équipe sur la modélisation de l'épidémie de SARS-CoV-2 en France. Notre modèle s'intéressait tout particulièrement à la description de l'épidémie au début et à la fin des vagues épidémiques, phases où la dynamique infectieuse est le fruit d'un nombre restreint d'individus. Les approches classiques en épidémiologie, comme les variantes de modèles SIR, ne permettent pas de prendre en compte le comportement non monotone de l'épidémie lorsque le nombre d'infectés est faible. Nous avons montré que même dans les scénarios les plus stricts, le temps nécessaire sous confinement avant l'extinction ou la mise sous contrôle de l'épidémie était assez important (> 100 jours). Malheureusement, cette question de persistance de l'épidémie à faible niveau n'a pas été anticipée par les pouvoirs publics lors de la première vague. Nos travaux renforcent également la nécessité d'intervenir tôt en contexte épidémique pour réduire au maximum la propagation du pathogène. Nous montrons que la mise en place de confinement précoce permet, en plus de la réduction du nombre de décès [Huang et al.], d'alléger le fardeau des confinements en diminuant le nombre de jours nécessaires pour atteindre des niveaux épidémiques contrôlables.

D'un point de vue global, ce travail se situe dans un ensemble de travaux menés collectivement au sein de l'équipe pour décrire et prédire la dynamique de la pandémie de COVID-19 avec des approches et des focales différentes. Sans entrer dans les détails, une des approches développées a mis en avant l'importance de la prise en compte des effets de mémoire (c'est-à-dire le caractère non-Markovien) pour saisir la dynamique épidémique sur le court terme [Sofonea et al., 2021]. Un autre travail a étudié plus en détails l'importance de la structure en âge dans un formalisme d'équations aux dérivées partielles [Reyné et al., 2022]. À l'aide de 196 séquences génétiques collectées, il a été possible de retracer la dynamique épidémique des premières semaines de l'épidémie en France à partir de modèles de phylodynamique [Danesh et al., 2020], mettant en exergue l'importance des données de séquençages génétiques dans la modélisation épidémique. La pandémie a également été le lieu de nombreux tests de dépistage par RT-qPCR. L'exploitation individuelle de ces données est limitée mais en grande quantité, elles peuvent fournir des informations pertinentes sur la dynamique épidémique [Alizon et al., 2022; Hay et al., 2021] ou sur la trajectoire évolutive des variants [Elie et al., 2022].

De manière plus générale encore, le contexte de la pandémie a mis en exergue l'importance des travaux de modélisation, mécaniste ou statistique, dans l'information et l'évaluation des politiques de santé publique [McBryde et al., 2020]. La diversité des approches et la reproductibilité des résultats sont encore plus cruciales dans l'information de politiques publiques dans un contexte de pandémie où l'instauration de mesures sanitaires peut avoir des conséquences secondaires sévères tant du point de vue humain qu'économique ou social [Brooks-Pollock et al., 2021]. Mais le rôle de la modélisation en épidémiologie ne se limite pas à un

travail descriptif et informatif. Elle peut être force de proposition et permettre d'étendre le champ des possibles en termes de décisions publiques [McBryde et al., 2020]. Néanmoins, l'interaction privilégiée entre modélisateurs et institutions publiques modifie les pratiques scientifiques. Ceci est particulièrement le cas en période épidémique majeure, où l'orientation de politiques publiques nécessite des réponses rapides. Cette urgence est souvent éloignée des temporalités classiques imposées par le système de revue par les pairs. Les manuscrits ayant suivi le chemin traditionnel de revue par les pairs paraissent obsolètes au moment de la publication des articles [Brooks-Pollock et al., 2021]. Ce contexte a favorisé l'essor du recours aux publications sous forme de *preprint*, jusqu'alors peu utilisés en sciences naturelles [Fraser et al., 2021]. Bien que cela permette un moyen rapide de partager des résultats dans un contexte de haute concurrence (autour de 30 000 *preprints* publiés sur le COVID-19 en 10 mois après le début de la pandémie [Fraser et al., 2021]) et d'incitations à la publication, il est légitime de s'interroger sur la qualité et la fiabilité de travaux avant revue par les pairs [Fraser et al., 2021]. Les rapports montrent que ceux-ci sont d'une meilleure qualité comparés à leurs équivalentes pré-publiées. Toutefois, cette différence reste limitée et les *preprints* restent des contributions scientifiques valides [Goodman et al., 1994; Pierie et al., 1996; Carneiro et al., 2020]. Ces résultats tendent à suggérer que la revue par les pairs permet plus de filtrer les papiers avant publication que d'améliorer la qualité du papier [Carneiro et al., 2020].

L'interaction privilégiée entre organes de santé publique et chercheurs en modélisation est bénéfique pour les deux protagonistes [Metcalf et al., 2015]. Toutefois, cela n'est pas sans difficulté pour les chercheurs en modélisation. L'interaction et le dialogue avec les instances publiques peut être chronophage et ce travail n'est pas forcément reconnu par la communauté scientifique, ces travaux de veille et de conseil ne constituant pas forcément un matériel publiable [Brooks-Pollock et al., 2021] et sont peu reconnus par les organismes de financement de la recherche comme l'ANR.

La modélisation mathématique offre des outils précieux dans la description de phénomènes biologiques, la proposition de nouveaux scénarios explicatifs et l'évaluation de politiques de santé publique [Djidjou-Demasse et al., 2020]. Néanmoins, peut-être plus que dans tout autre domaine, il est décisif que ces modèles parviennent aux personnes concernées. Bien que nécessaire à la description d'un modèle, une exposition trop directe et intense d'équations mathématiques diminue significativement la portée d'un travail de modélisation [Fawcett and Higginson, 2012]. Une présentation claire des hypothèses simplificatrices et des limites du modèle est essentiel à la transmission vers les pouvoirs publics et le corps médical [Basu and Andrews, 2013; Metcalf et al., 2015]. Au frein de la barrière des mathématiques vient s'ajouter le support de diffusion habituel de la science : des articles austères, le plus souvent rédigés en anglais [Drubin and Kellogg, 2012]. Dans les contextes épidémiques, il

est important que la science reste proche des citoyens [Ammon, 2011] et les chercheurs doivent jouer un rôle dans la diffusion et la vulgarisation de leurs travaux dans un format qui puissent être compris du plus grand nombre que ce soit sous forme de rapport en français [Bénéteau et al., 2020] ou d'interventions publiques dans les médias.

Bibliographie

- Taxonomy release history. <https://ictv.global/taxonomy/history>. Accessed : 2023-08-10. → pp. 3 et 5
- Décret n° 2020-548 du 11 mai 2020 prescrivant les mesures générales nécessaires pour faire face à l'épidémie de covid-19 dans le cadre de l'état d'urgence sanitaire. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000041865329/>, 2020. Accessed : 2023-09-07. → p. 102
- Aalen, O. Nonparametric Inference for a Family of Counting Processes. *The Annals of Statistics*, 6(4) :701–726, 1978. doi : 10.1214/aos/1176344247. → pp. 37, 54, 77 et 179
- Abdullahi, A., Copping, J., Kessel, A., Luck, M., and Bonell, C. Cervical screening : Perceptions and barriers to uptake among Somali women in Camden. *Public Health*, 123(10) :680–685, October 2009. ISSN 1476-5616. doi : 10.1016/j.puhe.2009.09.011. → p. 124
- Agresti, A. and Coull, B. A. Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52, 5 1998. ISSN 00031305. doi : 10.2307/2685469. → p. 90
- Alizon, S. *Évolution, écologie et pandémies : faire dialoguer Pasteur et Darwin*. Number S226 in Points. Éditions Points, Paris, 2020. ISBN 978-2-7578-8723-3. → p. 3
- Alizon, S. and Michalakis, Y. Adaptive virulence evolution : the good old fitness-based approach. *Trends in Ecology & Evolution*, 30(5) :248–254, May 2015. ISSN 0169-5347. doi : 10.1016/j.tree.2015.02.009. URL [https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347\(15\)00049-X](https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347(15)00049-X). Publisher : Elsevier. → p. 12
- Alizon, S. and Sofonea, M. T. SARS-CoV-2 virulence evolution : Avirulence theory, immunity and trade-offs. *Journal of Evolutionary Biology*, 34(12) :

1867–1877, December 2021. ISSN 1010-061X. doi : 10.1111/jeb.13896.
→ pp. 11 et 19

Alizon, S., Hurford, A., Mideo, N., and van Baalen, M. Virulence evolution and the trade-off hypothesis : history, current state of affairs and the future. *J. Evol. Biol.*, 22 :245–259, 2009. doi : 10.1111/j.1420-9101.2008.01658.x. tex.ids= AlizonEtal2009 tex.date-added : 2015-03-17 16 :08 :52 +0000 tex.date-modified : 2015-04-08 14 :47 :07 +0000 tex.local-url : /Users/alizon/Documents/Articles/Proches/AlizonEtal2009.pdf url : http://onlinelibrary.wiley.com/doi/10.1111/j.1420-9101.2008.01658.x/epdf.
→ pp. 11 et 12

Alizon, S., Murall, C. L., and Bravo, I. G. Why Human Papillomavirus Acute Infections Matter. *Viruses*, 9(10) :293, 2017. doi : 10.3390/v9100293.
→ p. 125

Alizon, S., Selinger, C., Sofonea, M. T., Haim-Boukobza, S., Giannoli, J.-M., Ninove, L., Pillet, S., Thibault, V., Rougemont, A. d., Tumiotta, C., Solis, M., Stephan, R., Bressollette-Bodin, C., Salmona, M., L’Honneur, A.-S., Behillil, S., Lefeuvre, C., Dina, J., Hantz, S., Hartard, C., Veyer, D., Delagrèverie, H. M., Fourati, S., Visseaux, B., Henquell, C., Lina, B., Foulongne, V., Burrel, S., and Group, o. b. o. t. S. C.-. s. Epidemiological and clinical insights from SARS-CoV-2 RT-PCR crossing threshold values, France, January to November 2020. *Eurosurveillance*, 27(6) :2100406, February 2022. ISSN 1560-7917. doi : 10.2807/1560-7917.ES.2022.27.6.2100406. URL <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2022.27.6.2100406>. Publisher : European Centre for Disease Prevention and Control. → p. 127

Allen, L. J. S. *An Introduction to Stochastic Processes with Applications to Biology*. Chapman and Hall/CRC, December 2010. ISBN 978-0-429-18460-4. doi : 10.1201/b12537. → p. 28

Allen, L. J. S. *Stochastic Population and Epidemic Models : Persistence and Extinction*. Springer International Publishing, Cham, 2015. ISBN 978-3-319-21553-2 978-3-319-21554-9. doi : 10.1007/978-3-319-21554-9. URL <https://link.springer.com/10.1007/978-3-319-21554-9>. → p. 28

Althouse, B. M., Wenger, E. A., Miller, J. C., Scarpino, S. V., Allard, A., Hébert-Dufresne, L., and Hu, H. Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2. *arXiv*, May 2020a. arXiv : 2005.13689. → p. 103

Althouse, B. M., Wenger, E. A., Miller, J. C., Scarpino, S. V., Allard, A., Hébert-Dufresne, L., and Hu, H. Superspreading events in the transmission dynamics of SARS-CoV-2 : Opportunities for interventions and control. *PLOS Biology*, 18(11) :e3000897, 2020b. doi : 10.1371/journal.pbio.3000897. → pp. 107 et 109

Amella, C. A., Lofgren, L. A., Ronn, A. M., Nouri, M., Shikowitz, M. J., and Steinberg, B. M. Latent infection induced with cottontail rabbit papillomavirus. A model for human papillomavirus latency. *The American Journal*

- of *Pathology*, 144(6) :1167, June 1994. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1887472/>. Publisher : American Society for Investigative Pathology. → pp. 64 et 125
- Ammon, U. *The Dominance of English as a Language of Science : Effects on Other Languages and Language Communities*. Walter de Gruyter, July 2011. ISBN 978-3-11-086948-4. → p. 129
- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. The proximal origin of SARS-CoV-2. *Nat Med*, pages 1–3, March 2020. ISSN 1546-170X. doi : 10.1038/s41591-020-0820-9. URL <https://www.nature.com/articles/s41591-020-0820-9>. Publisher : Nature Publishing Group. → p. 9
- Anderson, R. M. and May, R. M. *Infectious Diseases of Humans. Dynamics and Control*. Oxford University Press, Oxford, 1991. → p. 107
- Armitage, P. Multistage models of carcinogenesis. *Environmental Health Perspectives*, 63 :195–201, November 1985. ISSN 0091-6765. doi : 10.1289/ehp.8563195. URL <https://ehp.niehs.nih.gov/doi/10.1289/ehp.8563195>. → p. 42
- Ashcroft, P., Lehtinen, S., and Bonhoeffer, S. Test-trace-isolate-quarantine (TTIQ) intervention strategies after symptomatic COVID-19 case identification. *PloS One*, 17(2) :e0263597, 2022. ISSN 1932-6203. doi : 10.1371/journal.pone.0263597. → p. 102
- Athreya, K. B. and Ney, P. E. *Branching Processes*. Springer Berlin Heidelberg, 1972. ISBN 978-3-642-65373-5. doi : 10.1007/978-3-642-65371-1. → pp. 28, 88, 93 et 189
- Aubin, F., Prétet, J.-L., Jacquard, A.-C., Saunier, M., Carcopino, X., Jaroud, F., Pradat, P., Soubeyrand, B., Leocmach, Y., Mougin, C., Riethmuller, D., and EDiTH Study Group. Human papillomavirus genotype distribution in external acuminata condylomata : a Large French National Study (EDiTH IV). *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, 47(5) :610–615, September 2008. ISSN 1537-6591. doi : 10.1086/590560. → p. 17
- Averette, H. E., Weinstein, G. D., and Frost, P. Autoradiographic analysis of cell proliferation kinetics in human genital tissues. *American Journal of Obstetrics and Gynecology*, 108, 9 1970. ISSN 00029378. doi : 10.1016/0002-9378(70)90195-X. → pp. 83, 88, 90, 91, 93 et 99
- Bacaër, N. *A Short History of Mathematical Population Dynamics*. Springer London, London, 2011. ISBN 978-0-85729-114-1 978-0-85729-115-8. doi : 10.1007/978-0-85729-115-8. URL <https://link.springer.com/10.1007/978-0-85729-115-8>. → p. 20
- Bajaj, V., Gadi, N., Spihlman, A. P., Wu, S. C., Choi, C. H., and Moulton, V. R. Aging, Immunity, and COVID-19 : How Age Influences the Host Immune Response to Coronavirus Infections? *Frontiers in Physiology*, 11,

2021. ISSN 1664-042X. URL <https://www.frontiersin.org/articles/10.3389/fphys.2020.571416>. → p. 10
- Baltimore, D. Expression of animal virus genomes. *Bacteriological reviews*, 35(3), September 1971. ISSN 0005-3678. doi : 10.1128/br.35.3.235-241.1971. URL <https://pubmed.ncbi.nlm.nih.gov/4329869/>. Publisher : Bacteriol Rev. → p. 4
- Barberis, I., Bragazzi, N., Galluzzo, L., and Martini, M. The history of tuberculosis : from the first historical records to the isolation of Koch’s bacillus. *Journal of Preventive Medicine and Hygiene*, 58(1) :E9–E12, March 2017. ISSN 1121-2233. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5432783/>. → p. 12
- Bartoń, K. *MuMIn : Multi-Model Inference*, 2022. URL <https://CRAN.R-project.org/package=MuMIn>. R package version 1.46.0. → p. 182
- Bastani, R., Glenn, B. A., Tsui, J., Chang, L. C., Marchand, E. J., Taylor, V. M., and Singhal, R. Understanding Suboptimal Human Papillomavirus Vaccine Uptake Among Ethnic Minority Girls. *Cancer Epidemiology, Biomarkers & Prevention*, 20(7) :1463–1472, July 2011. ISSN 1055-9965. doi : 10.1158/1055-9965.EPI-11-0267. URL <https://doi.org/10.1158/1055-9965.EPI-11-0267>. → p. 123
- Basu, S. and Andrews, J. Complexity in Mathematical Models of Public Health Policies : A Guide for Consumers of Models. *PLOS Medicine*, 10(10) : e1001540, October 2013. ISSN 1549-1676. doi : 10.1371/journal.pmed.1001540. URL <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001540>. Publisher : Public Library of Science. → p. 128
- Bedford, T., Greninger, A. L., Roychoudhury, P., Starita, L. M., Famulare, M., Huang, M.-L., Nalla, A., Pepper, G., Reinhardt, A., Xie, H., Shrestha, L., Nguyen, T. N., Adler, A., Brandstetter, E., Cho, S., Giroux, D., Han, P. D., Fay, K., Frazar, C. D., Ilcisin, M., Lacombe, K., Lee, J., Kiavand, A., Richardson, M., Sibley, T. R., Truong, M., Wolf, C. R., Nickerson, D. A., Rieder, M. J., Englund, J. A., Investigators†, T. S. F. S., Hadfield, J., Huddleston, E. B., Huddleston, J., Moncla, L. H., Müller, N. F., Neher, R. A., Deng, X., Gu, W., Federman, S., Chiu, C., Duchin, J. S., Gautom, R., Melly, G., Hiatt, B., Dykema, P., Lindquist, S., Queen, K., Tao, Y., Uehara, A., Tong, S., MacCannell, D., Armstrong, G. L., Baird, G. S., Chu, H. Y., Shendure, J., and Jerome, K. R. Cryptic transmission of SARS-CoV-2 in Washington state. *Science*, 2020. doi : 10.1126/science.abc0523. → p. 106
- Bénéteau, T., Elie, B., Sofonea, M. T., and Alizon, S. Estimating dates of origin and end of COVID-19 epidemics. *medRxiv*, page 2021.01.19.21250080, January 2021. doi : 10.1101/2021.01.19.21250080. → p. 120
- Berenson, A. B. and Rahman, M. Gender Differences Among Low Income Women in Their Intent to Vaccinate Their Sons and Daughters Against Human Papillomavirus Infection. *Journal of Pediatric and Adolescent Gynecology*, 25(3) :218–220, June 2012. ISSN 1083-3188, 1873-4332. doi :

- 10.1016/j.jpag.2012.01.003. URL [https://www.jpagonline.org/article/S1083-3188\(12\)00007-1/fulltext](https://www.jpagonline.org/article/S1083-3188(12)00007-1/fulltext). Publisher : Elsevier. → p. 123
- Berkson, J. and Gage, R. P. Calculation of survival rates for cancer. *Proceedings of the Staff Meetings. Mayo Clinic*, 25(11) :270–286, May 1950. ISSN 0092-699X. → p. 34
- Bernoulli, D. and Chapelle, D. Essai d’une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l’inoculation pour la prévenir. Mémoires de Mathématique et de Physique, tirés des registres de l’Académie Royale des Sciences de l’année 1760, pages 1-45, May 2023. URL <https://inria.hal.science/hal-04100467>. → p. 20
- Bidot, C., Lamboni, M., and Monod, H. *multisensi : Multivariate Sensitivity Analysis*, 2018. URL <https://CRAN.R-project.org/package=multisensi>. R package version 2.1-1. → p. 92
- Bienaymé, I.-J. *De la loi de multiplication et de la durée des familles, probabilités...* Cosson, 1845. → p. 28
- Bogaerts, K., Komárek, A., and Lesaffre, E. *Survival Analysis with Interval-Censored Data*. Chapman and Hall/CRC, November 2017. ISBN 978-1-315-11694-5. doi : 10.1201/9781315116945. → p. 178
- Bonanni, P., Zanella, B., Santomauro, F., Lorini, C., Bechini, A., and Boccalini, S. Safety and perception : What are the greatest enemies of HPV vaccination programmes? *Vaccine*, 36(36) :5424–5429, August 2018. ISSN 0264-410X. doi : 10.1016/j.vaccine.2017.05.071. URL <https://www.sciencedirect.com/science/article/pii/S0264410X17307302>. → p. 123
- Bosch, F. X., Lorincz, A., Muñoz, N., Meijer, C. J. L. M., and Shah, K. V. The causal relation between human papillomavirus and cervical cancer. *Journal of clinical pathology*, 55 :244–265, 4 2002. ISSN 0021-9746. doi : 10.1136/jcp.55.4.244. URL <https://pubmed.ncbi.nlm.nih.gov/11919208https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1769629/>. → p. 15
- Bosetti, P., Kiem, C. T., Yazdanpanah, Y., Fontanet, A., Lina, B., Colizza, V., and Cauchemez, S. Impact of mass testing during an epidemic rebound of SARS-CoV-2 : a modelling study using the example of France. *Eurosurveillance*, 26(1) :2001978, January 2021. ISSN 1560-7917. doi : 10.2807/1560-7917.ES.2020.26.1.2001978. URL <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.26.1.2001978>. Publisher : European Centre for Disease Prevention and Control. → pp. 102 et 103
- Boshart, M., Gissmann, L., Ikenberg, H., Kleinheinz, A., Scheurlen, W., and zur Hausen, H. A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer. *The EMBO Journal*, 3(5) :1151–1157, May 1984. ISSN 0261-4189. doi : 10.1002/j.1460-2075.1984.tb01944.x. URL <https://www.embopress.org/doi/abs/10.1002/j.1460-2075.1984.tb01944.x>. Publisher : John Wiley & Sons, Ltd. → pp. 15, 64 et 125

- Brauer, F. Compartmental Models in Epidemiology. *Mathematical Epidemiology*, 1945 :19–79, 2008. doi : 10.1007/978-3-540-78911-6_2. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7122373/>. → p. 20
- Bravo, I. G. and Félez-Sánchez, M. Papillomaviruses : Viral evolution, cancer and evolutionary medicine. *Evolution, Medicine, and Public Health*, 2015 (1) :32–51, January 2015. ISSN 2050-6201. doi : 10.1093/emph/eov003. URL <https://doi.org/10.1093/emph/eov003>. → p. 12
- Bravo, I. G., de Sanjosé, S., and Gottschling, M. The clinical importance of understanding the evolution of papillomaviruses. *Trends in Microbiology*, 18, 10 2010. ISSN 0966842X. doi : 10.1016/j.tim.2010.07.008. → p. 98
- Britton, T. and Scalia Tomba, G. Estimation in emerging epidemics : biases and remedies. *Journal of The Royal Society Interface*, 16(150) :20180670, January 2019. doi : 10.1098/rsif.2018.0670. URL <https://royalsocietypublishing.org/doi/full/10.1098/rsif.2018.0670>. → p. 106
- Brooks-Pollock, E., Danon, L., Jombart, T., and Pellis, L. Modelling that shaped the early COVID-19 pandemic response in the UK. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 376(1829) : 20210001, May 2021. doi : 10.1098/rstb.2021.0001. URL <https://royalsocietypublishing.org/doi/10.1098/rstb.2021.0001>. Publisher : Royal Society. → pp. 2, 127 et 128
- Brotherton, J. M. L. Impact of HPV vaccination : Achievements and future challenges. *Papillomavirus Research (Amsterdam, Netherlands)*, 7 :138–140, June 2019. ISSN 2405-8521. doi : 10.1016/j.pvr.2019.04.004. → pp. 61 et 80
- Brotman, R. M., Shardell, M. D., Gajer, P., Tracy, J. K., Zenilman, J. M., Ravel, J., and Gravitt, P. E. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. *J Infect Dis*, 210(11) :1723–33, 2014. doi : 10.1093/infdis/jiu330. → p. 49
- Brouwer, A. F., Delinger, R. L., Eisenberg, M. C., Campredon, L. P., Walline, H. M., Carey, T. E., and Meza, R. HPV vaccination has not increased sexual activity or accelerated sexual debut in a college-aged cohort of men and women. *BMC Public Health*, 19 :821, June 2019. ISSN 1471-2458. doi : 10.1186/s12889-019-7134-1. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6593582/>. → p. 123
- Brown, F. The history of research in foot-and-mouth disease. *Virus Research*, 91(1) :3–7, January 2003. ISSN 0168-1702. doi : 10.1016/s0168-1702(02)0268-x. → p. 3
- Bruni, L., Diaz, M., Castellsagué, M., Ferrer, E., Bosch, F. X., and de Sanjosé, S. Cervical Human Papillomavirus Prevalence in 5 Continents : Meta-Analysis of 1 Million Women with Normal Cytological Findings. *The Journal of Infectious Diseases*, 202(12) :1789–1799, December 2010. ISSN 0022-1899. doi : 10.1086/657321. URL <https://doi.org/10.1086/657321>. → pp. 52, 79 et 80

- Bruni, L., Saura-Lázaro, A., Montoliu, A., Brotons, M., Alemany, L., Diallo, M. S., Afsar, O. Z., LaMontagne, D. S., Mosina, L., Contreras, M., Velandia-González, M., Pastore, R., Gacic-Dobo, M., and Bloem, P. HPV vaccination introduction worldwide and WHO and UNICEF estimates of national HPV immunization coverage 2010–2019. *Preventive Medicine*, 144 :106399, 2021. doi : <https://doi.org/10.1016/j.ypmed.2020.106399>. → p. 122
- Bruni, L., Serrano, B., Roura, E., Alemany, L., Cowan, M., Herrero, R., Poljak, M., Murillo, R., Broutet, N., Riley, L. M., and de Sanjose, S. Cervical cancer screening programmes and age-specific coverage estimates for 202 countries and territories worldwide : a review and synthetic analysis. *The Lancet Global Health*, 10(8) :e1115–e1127, August 2022. ISSN 2214-109X. doi : 10.1016/S2214-109X(22)00241-8. URL <https://www.sciencedirect.com/science/article/pii/S2214109X22002418>. → pp. 16 et 123
- Bruni, L., Albero, G., Rowley, J., Alemany, L., Arbyn, M., Giuliano, A. R., Markowitz, L. E., Broutet, N., and Taylor, M. Global and regional estimates of genital human papillomavirus prevalence among men : a systematic review and meta-analysis. *The Lancet Global Health*, 11(9) :e1345–e1362, September 2023. ISSN 2214-109X. doi : 10.1016/S2214-109X(23)00305-4. Publisher : Elsevier. → p. 125
- Bureau, A., Shiboski, S., and Hughes, J. P. Applications of continuous time hidden Markov models to the study of misclassified disease outcomes. *Statistics in Medicine*, 22(3) :441–462, 2003. ISSN 1097-0258. doi : 10.1002/sim.1270. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.1270>. _eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.1270>. → p. 65
- Burnham, K. P. and Anderson, D. R. Multimodel Inference : Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2) : 261–304, 2004. doi : 10.1177/0049124104268644. → p. 179
- Bénéteau, T., Elie, B., Alizon, S., and Sofonea, M. T. Stochasticité et épidémie de COVID-19. Research Report 8, Centre National de la Recherche Scientifique (CNRS); Institut de Recherche pour le Développement (IRD); Université de Montpellier (UM), FRA., May 2020. URL <https://hal.science/hal-02950660>. → pp. 105 et 129
- Böhmer, P. Theorie der unabhängigen Wahrscheinlichkeiten. *Rapports Mémoires et Procès verbaux de Septieme Congres International dActuaires Amsterdam*, 2 :327–343, 1912. → p. 34
- Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D., and Lawley, T. D. Massive expansion of human gut bacteriophage diversity. *Cell*, 184(4) :1098–1109.e9, February 2021. ISSN 0092-8674. doi : 10.1016/j.cell.2021.01.029. URL <https://www.sciencedirect.com/science/article/pii/S0092867421000726>. → p. 3
- Cao, Y., Gillespie, D. T., and Petzold, L. R. Efficient step size selection for the tau-leaping simulation method. *The Journal of Chemical Physics*, 124(4) : 044109, January 2006. ISSN 0021-9606. doi : 10.1063/1.2159468. → p. 33

- Carabelli, A. M., Peacock, T. P., Thorne, L. G., Harvey, W. T., Hughes, J., de Silva, T. I., Peacock, S. J., Barclay, W. S., de Silva, T. I., Towers, G. J., and Robertson, D. L. SARS-CoV-2 variant biology : immune escape, transmission and fitness. *Nature Reviews Microbiology*, 21(3) :162–177, March 2023. ISSN 1740-1534. doi : 10.1038/s41579-022-00841-7. URL <https://www.nature.com/articles/s41579-022-00841-7>. Number : 3 Publisher : Nature Publishing Group. → p. 19
- Carneiro, C. F. D., Queiroz, V. G. S., Moulin, T. C., Carvalho, C. A. M., Haas, C. B., Rayêe, D., Henshall, D. E., De-Souza, E. A., Amorim, F. E., Boos, F. Z., Guercio, G. D., Costa, I. R., Hajdu, K. L., van Egmond, L., Modrák, M., Tan, P. B., Abdill, R. J., Burgess, S. J., Guerra, S. F. S., Bortoluzzi, V. T., and Amaral, O. B. Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical literature. *Research Integrity and Peer Review*, 5(1) :16, December 2020. ISSN 2058-8615. doi : 10.1186/s41073-020-00101-3. URL <https://doi.org/10.1186/s41073-020-00101-3>. → p. 128
- Carnell, R. *lhs : Latin Hypercube Samples*, 2020. URL <https://CRAN.R-project.org/package=lhs>. R package version 1.1.1. → pp. 92 et 195
- Carter, J., Koutsky, L., Hughes, J., Lee, S., Kuypers, J., Kiviat, N., and Galloway, D. Comparison of human papillomavirus types 16, 18, and 6 capsid antibody responses following incident infection. *The Journal of infectious diseases*, 181(6), June 2000. ISSN 0022-1899. doi : 10.1086/315498. URL <https://pubmed.ncbi.nlm.nih.gov/10837170/>. Publisher : J Infect Dis. → p. 84
- Carter, J. J., Wipf, G. C., Hagensee, M. E., McKnight, B., Habel, L. A., Lee, S.-K., Kuypers, J., Kiviat, N., Dating, J. R., Koutsky, L. A., Watts, D. H., Holmes, K. K., and Galloway, D. A. Use of human papillomavirus type 6 capsids to detect antibodies in people with genital warts. *Journal of Infectious Diseases*, 172, 7 1995. ISSN 0022-1899. doi : 10.1093/infdis/172.1.11. → p. 86
- Center for Disease Control and Prevention. Website : CDC Respiratory Virus Updates, 2023. URL <https://www.cdc.gov/respiratory-viruses/whats-new/>. → pp. 2 et 19
- Chan, P. K. S., Luk, A. C. S., Luk, T. N. M., Lee, K.-F., Cheung, J. L. K., Ho, K.-M., and Lo, K.-K. Distribution of human papillomavirus types in anogenital warts of men. *Journal of Clinical Virology : The Official Publication of the Pan American Society for Clinical Virology*, 44(2) :111–114, February 2009. ISSN 1873-5967. doi : 10.1016/j.jcv.2008.11.001. → p. 17
- Chatterjee, S., Chattopadhyay, A., Samanta, L., and Panigrahi, P. HPV and Cervical Cancer Epidemiology - Current Status of HPV Vaccination in India. *Asian Pacific journal of cancer prevention : APJCP*, 17(8) :3663–3673, 2016. ISSN 2476-762X. → p. 124
- Chaturvedi, A. K., Katki, H. A., Hildesheim, A., Rodríguez, A. C., Quint, W., Schiffman, M., Doorn, L.-J. V., Carolina Porras, . S. W., Gonzalez,

- P., Sherman, M. E., Herrero, R., and Group, t. C. Human Papillomavirus Infection with Multiple Types : Pattern of Coinfection and Risk of Cervical Disease. *J Infect Dis*, 203 :910–920, 2011. doi : 10.1093/infdis/jiq139. → p. 72
- Chen, W.-T. Chinese female immigrants english-speaking ability and breast and cervical cancer early detection practices in the New York metropolitan area. *Asian Pacific journal of cancer prevention : APJCP*, 14(2) :733–738, 2013. ISSN 2476-762X. doi : 10.7314/apjcp.2013.14.2.733. → p. 124
- Chesson, H. W., Dunne, E. F., Hariri, S., and Markowitz, L. E. The estimated lifetime probability of acquiring human papillomavirus in the united states. *Sexually Transmitted Diseases*, 41, 2014a. ISSN 0148-5717. URL https://journals.lww.com/stdjournal/Fulltext/2014/11000/The_Estimate_d_Lifetime_Probability_of_Acquiring.4.aspx. → pp. 15 et 86
- Chesson, H. W., Dunne, E. F., Hariri, S., and Markowitz, L. E. The estimated lifetime probability of acquiring human papillomavirus in the United States. *Sexually Transmitted Diseases*, 41(11) :660–664, 2014b. ISSN 1537-4521. doi : 10.1097/OLQ.000000000000193. → p. 52
- Chiu, Y.-F. and Sugden, B. Plasmid partitioning by human tumor viruses. *Journal of Virology*, 92, 2 2018. ISSN 0022-538X. doi : 10.1128/JVI.02170-17. URL <https://jvi.asm.org/content/92/9/e02170-17>. → p. 87
- Clark, R. and Kupper, T. Old meets new : the interaction between innate and adaptive immunity. *The Journal of Investigative Dermatology*, 125(4) : 629–637, October 2005. ISSN 0022-202X. doi : 10.1111/j.0022-202X.2005.23856.x. → p. 10
- Clayton, E., Doupé, D. P., Klein, A. M., Winton, D. J., Simons, B. D., and Jones, P. H. A single type of progenitor cell maintains normal epidermis. *Nature*, 446 :185–189, 3 2007. ISSN 0028-0836. doi : 10.1038/nature05574. URL <http://www.nature.com/articles/nature05574>. → pp. 83, 87, 90 et 91
- Cocchio, S., Bertoncetto, C., Baldovin, T., Fonzo, M., Bennici, S. E., Buja, A., Majori, S., and Baldo, V. Awareness of HPV and drivers of HPV vaccine uptake among university students : A quantitative, cross-sectional study. *Health & Social Care in the Community*, 28(5) :1514–1524, 2020. ISSN 1365-2524. doi : 10.1111/hsc.12974. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/hsc.12974>. _eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/hsc.12974>. → p. 124
- Cohn, L., Homer, R. J., Marinov, A., Rankin, J., and Bottomly, K. Induction of Airway Mucus Production By T Helper 2 (Th2) Cells : A Critical Role For Interleukin 4 In Cell Recruitment But Not Mucus Production. *The Journal of Experimental Medicine*, 186(10) :1737–1747, November 1997. ISSN 0022-1007. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2199146/>. → p. 8

- Contreras, S., Dehning, J., Loidolt, M., Zierenberg, J., Spitzner, F. P., Urrea-Quintero, J. H., Mohr, S. B., Wilczek, M., Wibrál, M., and Priesemann, V. The challenges of containing SARS-CoV-2 via test-trace-and-isolate. *Nature Communications*, 12(1) :378, January 2021. ISSN 2041-1723. doi : 10.1038/s41467-020-20699-8. URL <https://www.nature.com/articles/s41467-020-20699-8>. Number : 1 Publisher : Nature Publishing Group. → p. 102
- Coutlée, F., Gravitt, P., Kornegay, J., Hankins, C., Richardson, H., Lapointe, N., Voyer, H., Group, T. C. W. H. S., and Franco, E. Use of PGMV Primers in L1 Consensus PCR Improves Detection of Human Papillomavirus DNA in Genital Samples. *Journal of Clinical Microbiology*, 40(3) :902–907, March 2002. doi : 10.1128/JCM.40.3.902-907.2002. → pp. 48, 53 et 68
- Cox, D. R. Regression Models and Life-Tables. In Kotz, S. and Johnson, N. L., editors, *Breakthroughs in Statistics : Methodology and Distribution*, Springer Series in Statistics, pages 527–541. Springer, New York, NY, 1992. doi : 10.1007/978-1-4612-4380-9_37. → pp. 39 et 55
- Cox, F. E. History of the discovery of the malaria parasites and their vectors. *Parasites & Vectors*, 3(1) :5, February 2010. ISSN 1756-3305. doi : 10.1186/1756-3305-3-5. URL <https://doi.org/10.1186/1756-3305-3-5>. → p. 12
- Crow, J. M. Hpv : The global burden. *Nature*, 488 :S2–S3, 8 2012. ISSN 0028-0836. doi : 10.1038/488S2a. URL <http://www.nature.com/articles/488S2a>. → p. 98
- Dalton, J. P., editor. *Fasciolosis*. CABI Publ, Wallingford, 1999. ISBN 978-0-85199-260-0. → p. 12
- Danesh, G., Elie, B., Michalakis, Y., Sofonea, M. T., Bal, A., Behillil, S., Destras, G., Boutolleau, D., Burrel, S., Marcelin, A.-G., Plantier, J.-C., Thibault, V., Simon-Lorière, E., der Werf, S. v., Lina, B., Josset, L., Enouf, V., and Alizon, S. Early phylodynamics analysis of the COVID-19 epidemic in France. *medRxiv*, 2020.06.03.20119925, ver. 3 peer-reviewed and recommended by *PCI in Evolutionary Biology*, 2020. doi : 10.24072/pci.evolbiol.100107. → pp. 103, 106, 118 et 127
- Danesh, G., Saulnier, E., Gascuel, O., Choisy, M., and Alizon, S. Tips : Rapidly simulating trajectories and phylogenies from compartmental models. *Methods in Ecology and Evolution*, 14(2) :487–495, 2023a. doi : <https://doi.org/10.1111/2041-210X.14038>. URL <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14038>. → p. 33
- Danesh, G., Saulnier, E., Gascuel, O., Choisy, M., and Alizon, S. Tips : Rapidly simulating trajectories and phylogenies from compartmental models. *Methods in Ecology and Evolution*, 14(2) :487–495, 2023b. doi : <https://doi.org/10.1111/2041-210X.14038>. URL <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14038>. → pp. 111 et 203

- Dawkins, R., Krebs, J. R., Maynard Smith, J., and Holliday, R. Arms races between and within species. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 205(1161) :489–511, January 1997. doi : 10.1098/rspb.1979.0081. URL <https://royalsocietypublishing.org/doi/10.1098/rspb.1979.0081>. Publisher : Royal Society. → p. 12
- de Martel, C., Plummer, M., Vignat, J., and Franceschi, S. Worldwide burden of cancer attributable to hpv by site, country and hpv type. *International Journal of Cancer*, 141 :664–670, 8 2017. ISSN 00207136. doi : 10.1002/ijc.30716. URL <http://doi.wiley.com/10.1002/ijc.30716>. → pp. 15, 16, 86 et 98
- de Sanjosé, S., Brotons, M., and Pavón, M. A. The natural history of human papillomavirus infection. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 47 :2–13, February 2018. ISSN 1521-6934. doi : 10.1016/j.bpobgyn.2017.08.015. URL <https://www.sciencedirect.com/science/article/pii/S1521693417301335>. → pp. 13, 15 et 212
- de Wit, E., van Doremalen, N., Falzarano, D., and Munster, V. J. SARS and MERS : recent insights into emerging coronaviruses. *Nature Reviews. Microbiology*, 14(8) :523–534, August 2016. ISSN 1740-1534. doi : 10.1038/nrmicro.2016.81. → p. 18
- Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society : Series B (Methodological)*, 39(1) :1–22, 1977. ISSN 2517-6161. doi : 10.1111/j.2517-6161.1977.tb01600.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1977.tb01600.x>. _eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1977.tb01600.x>. → p. 45
- Denton, K. J. Liquid based cytology in cervical cancer screening. *BMJ (Clinical research ed.)*, 335(7609) :1–2, July 2007. ISSN 1756-1833. doi : 10.1136/bmj.39262.506528.47. → p. 16
- Diekmann, O., Heesterbeek, J. A., and Metz, J. A. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.*, 28(4) :365–82, 1990. doi : 10.1007/BF00178324. → p. 26
- Dilley, S. E., Peral, S., Straughn, J. M., and Scarinci, I. C. The challenge of HPV vaccination uptake and opportunities for solutions : Lessons learned from Alabama. *Preventive Medicine*, 113 :124–131, 2018. doi : 10.1016/j.jpmed.2018.05.021. → p. 182
- Dillner, J., Arbyn, M., Unger, E., and Dillner, L. Monitoring of human papillomavirus vaccination. *Clin Exp Immunol*, 163(1) :17–25, 2011. doi : 10.1111/j.1365-2249.2010.04268.x. → pp. 54, 69 et 178
- Dimmock, N. J., Easton, A. J., and Leppard, K. N. *Introduction to modern virology*. John Wiley & Sons, Nashville, TN, 7 edition, March 2016. → pp. 3, 4 et 11

- Ding, D.-C., Chang, Y.-C., Liu, H.-W., and Chu, T.-Y. Long-term persistence of human papillomavirus in environments. *Gynecologic Oncology*, 121(1) : 148–151, April 2011. ISSN 1095-6859. doi : 10.1016/j.ygyno.2010.11.040. → p. 16
- Djaafara, B. A., Imai, N., Hamblion, E., Impouma, B., Donnelly, C. A., and Cori, A. A quantitative framework to define the end of an outbreak : Application to ebola virus disease. *Am J Epidemiol*, page kwaa212, 2020. ISSN 0002-9262, 1476-6256. doi : 10.1093/aje/kwaa212. → p. 106
- Djidjou-Demasse, R., Selinger, C., and Sofonea, M. T. Épidémiologie mathématique et modélisation de la pandémie de Covid-19 : enjeux et diversité. *Revue Francophone des Laboratoires*, 2020(526) :63–69, November 2020. ISSN 1773035X. doi : 10.1016/S1773-035X(20)30315-4. URL <https://linkinghub.elsevier.com/retrieve/pii/S1773035X20303154>. → pp. 2, 20, 22, 83, 125 et 128
- Donken, R., Ogilvie, G. S., Bettinger, J. A., Sadarangani, M., and Goldman, R. D. Effect of human papillomavirus vaccination on sexual behaviour among young females. *Canadian Family Physician*, 64(7) :509–513, July 2018. ISSN 0008-350X. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6042675/>. → p. 123
- Doorbar, J. The papillomavirus life cycle. *Journal of Clinical Virology*, 32 : 7–15, March 2005. ISSN 1386-6532. doi : 10.1016/j.jcv.2004.12.006. URL <https://www.sciencedirect.com/science/article/pii/S1386653204003671>. → p. 15
- Doorbar, J. Molecular biology of human papillomavirus infection and cervical cancer. *Clinical Science*, 110 :525–541, 5 2006. ISSN 0143-5221. doi : 10.1042/CS20050369. URL <https://portlandpress.com/clinsci/article/110/5/525/68249/Molecular-biology-of-human-papillomavirus>. → pp. 7, 14, 84, 87, 89, 91 et 99
- Doorbar, J. Latent papillomavirus infections and their regulation. *Curr Opin Virol*, 3(4) :416–21, 2013. doi : 10.1016/j.coviro.2013.06.003. → p. 99
- Doorbar, J. The human Papillomavirus twilight zone – Latency, immune control and subclinical infection. *Tumour Virus Research*, 16 :200268, December 2023. ISSN 2666-6790. doi : 10.1016/j.tvr.2023.200268. URL <https://www.sciencedirect.com/science/article/pii/S2666679023000150>. → pp. 15, 48, 64, 70, 79, 80 et 125
- Doorbar, J., Quint, W., Banks, L., Bravo, I. G., Stoler, M., Broker, T. R., and Stanley, M. A. The biology and life-cycle of human papillomaviruses. *Vaccine*, 30 :F55–F70, 11 2012. ISSN 0264410X. doi : 10.1016/j.vaccine.2012.06.083. URL <https://linkinghub.elsevier.com/retrieve/pii/S0264410X12009735>. → pp. 13, 14, 83, 84, 87, 97, 98 et 122
- Doorbar, J., Egawa, N., Griffin, H., Kranjec, C., and Murakami, I. Human papillomavirus molecular biology and disease association. *Reviews in Medical*

- Virology*, 25, 3 2015. ISSN 1052-9276. doi : 10.1002/rmv.1822. → pp. 13, 14 et 98
- Doupe, D. P., Alcolea, M. P., Roshan, A., Zhang, G., Klein, A. M., Simons, B. D., and Jones, P. H. A single progenitor population switches behavior to maintain and repair esophageal epithelium. *Science*, 337 :1091–1093, 8 2012. ISSN 0036-8075. doi : 10.1126/science.1218835. URL <https://www.sciencemag.org/lookup/doi/10.1126/science.1218835>. → pp. 87, 95 et 97
- Downs, L. O., Smith, D. A., Lumley, S. F., Patel, M., McNaughton, A. L., Mokaya, J., Ansari, M. A., Salih, H., Várnai, K. A., Freeman, O., Cripps, S., Phillips, J., Collier, J., Woods, K., Channon, K., Davies, J., Barnes, E., Jeffery, K., and Matthews, P. C. Electronic Health Informatics Data To Describe Clearance Dynamics of Hepatitis B Surface Antigen (HBsAg) and e Antigen (HBeAg) in Chronic Hepatitis B Virus Infection. *mBio*, 10(3) : e00699–19, June 2019. ISSN 2150-7511. doi : 10.1128/mBio.00699-19. → p. 126
- Drubin, D. G. and Kellogg, D. R. English as the universal language of science : opportunities and challenges. *Molecular Biology of the Cell*, 23(8) :1399–1399, April 2012. ISSN 1059-1524. doi : 10.1091/mbc.e12-02-0108. URL <https://www.molbiolcell.org/doi/full/10.1091/mbc.e12-02-0108>. Publisher : American Society for Cell Biology (mboc). → p. 128
- Drugs, I. o. M. U. C. o. t. E. o. A., Arrow, K. J., Panosian, C., and Gelband, H. A Brief History of Malaria. In *Saving Lives, Buying Time : Economics of Malaria Drugs in an Age of Resistance*. National Academies Press (US), 2004. URL <https://www.ncbi.nlm.nih.gov/books/NBK215638/>. → p. 12
- du Plessis, L., McCrone, J. T., Zarebski, A. E., Hill, V., Ruis, C., Gutierrez, B., Raghwan, J., Ashworth, J., Colquhoun, R., Connor, T. R., Faria, N. R., Jackson, B., Loman, N. J., O’Toole, A., Nicholls, S. M., Parag, K. V., Scher, E., Vasylyeva, T. I., Volz, E. M., Watts, A., Bogoch, I. I., Khan, K., Consortium†, C.-. G. U. C.-U., Aanensen, D. M., Kraemer, M. U. G., Rambaut, A., and Pybus, O. G. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science*, January 2021. ISSN 0036-8075, 1095-9203. doi : 10.1126/science.abf2946. URL <https://science.sciencemag.org/content/early/2021/01/07/science.abf2946>. Publisher : American Association for the Advancement of Science Section : Research Article. → p. 119
- Ebert, D. and Fields, P. D. Host–parasite co-evolution and its genomic signature. *Nature Reviews Genetics*, 21(12) :754–768, December 2020. ISSN 1471-0064. doi : 10.1038/s41576-020-0269-1. URL <https://www.nature.com/articles/s41576-020-0269-1>. Number : 12 Publisher : Nature Publishing Group. → p. 3
- Efron, B. The Efficiency of Cox’s Likelihood Function for Censored Data. *Journal of the American Statistical Association*, 72(359) :557–565, September 1977. ISSN 0162-1459. doi : 10.1080/01621459.1977.1048

0613. URL <https://www.tandfonline.com/doi/abs/10.1080/01621459.1977.10480613>. Publisher : Taylor & Francis _eprint : <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1977.10480613>. → p. 39
- Egawa, N., Nakahara, T., Ohno, S., Narisawa-Saito, M., Yugawa, T., Fujita, M., Yamato, K., Natori, Y., and Kiyono, T. The e1 protein of human papillomavirus type 16 is dispensable for maintenance replication of the viral genome. *Journal of Virology*, 86, 3 2012. ISSN 0022-538X. doi : 10.1128/JVI.06450-11. → pp. 84, 98 et 99
- Egawa, N. and Doorbar, J. The low-risk papillomaviruses. *Virus Research*, 231, 3 2017. ISSN 01681702. doi : 10.1016/j.virusres.2016.12.017. → p. 98
- Egawa, N., Egawa, K., Griffin, H., and Doorbar, J. Human papillomaviruses ; epithelial tropisms, and the development of neoplasia. *Viruses*, 7, 7 2015. ISSN 1999-4915. doi : 10.3390/v7072802. → p. 98
- Eichner, M. and Dietz, K. Eradication of poliomyelitis : when can one be sure that polio virus transmission has been terminated? *American journal of epidemiology*, 143(8) :816–822, 1996. → p. 106
- Elie, B., Roquebert, B., Sofonea, M. T., Trombert-Paolantoni, S., Foulongne, V., Guedj, J., Haim-Boukoba, S., and Alizon, S. Variant-specific SARS-CoV-2 within-host kinetics. *Journal of Medical Virology*, 94(8) :3625–3633, 2022. ISSN 1096-9071. doi : 10.1002/jmv.27757. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.27757>. _eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmv.27757>. → p. 127
- Endo, A., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott, S., Kucharski, A. J., and Funk, S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res*, 5 :67, April 2020. ISSN 2398-502X. doi : 10.12688/wellcomeopenres.15842.1. URL <https://wellcomeopenresearch.org/articles/5-67/v1>. → pp. 103, 107, 109 et 119
- Etti, M., Calvert, A., Galiza, E., Lim, S., Khalil, A., Le Doare, K., and Heath, P. T. Maternal vaccination : a review of current evidence and recommendations. *American Journal of Obstetrics and Gynecology*, 226(4) :459–474, April 2022. ISSN 0002-9378. doi : 10.1016/j.ajog.2021.10.041. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8582099/>. → p. 9
- European Study Group on Heterosexual Transmission of HIV. Comparison of female to male and male to female transmission of HIV in 563 stable couples. European Study Group on Heterosexual Transmission of HIV. *BMJ : British Medical Journal*, 304(6830) :809–813, March 1992. ISSN 0959-8138. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1881672/>. → p. 8
- Farhat, S., Nakagawa, M., and Moscicki, A.-B. Cell-mediated immune responses to human papillomavirus 16 e6 and e7 antigens as measured by interferon gamma enzyme-linked immunospot in women with cleared or persistent human papillomavirus infection. *International Journal of Gynecological Cancer*, 19 :508–512, 5 2009. ISSN 1048-891X. doi : 10.1111/IGC.0b

013e3181a388c4. URL <https://clsjournal.ascls.org/lookup/doi/10.1111/IGC.0b013e3181a388c4>. → p. 86

Fawcett, T. W. and Higginson, A. D. Heavy use of equations impedes communication among biologists. *Proceedings of the National Academy of Sciences*, 109(29) :11735–11739, July 2012. doi : 10.1073/pnas.1205259109. URL <https://www.pnas.org/doi/abs/10.1073/pnas.1205259109>. Publisher : Proceedings of the National Academy of Sciences. → p. 128

Ferdinands, J. M., Rao, S., Dixon, B. E., Mitchell, P. K., DeSilva, M. B., Irving, S. A., Lewis, N., Natarajan, K., Stenhjem, E., Grannis, S. J., Han, J., McEvoy, C., Ong, T. C., Naleway, A. L., Reese, S. E., Embi, P. J., Dascomb, K., Klein, N. P., Griggs, E. P., Liao, I.-C., Yang, D.-H., Fadel, W. F., Grisel, N., Goddard, K., Patel, P., Murthy, K., Birch, R., Valvi, N. R., Arndorfer, J., Zerbo, O., Dickerson, M., Raiyani, C., Williams, J., Bozio, C. H., Blanton, L., Link-Gelles, R., Barron, M. A., Gaglani, M., Thompson, M. G., and Fireman, B. Waning of vaccine effectiveness against moderate and severe covid-19 among adults in the US from the VISION network : test negative, case-control study. 379 :e072141. ISSN 1756-1833. doi : 10.1136/bmj-2022-072141. URL <https://www.bmj.com/content/379/bmj-2022-072141>. Publisher : British Medical Journal Publishing Group Section : Research. → p. 40

Ferrière, R. and Tran, V. C. Stochastic and deterministic models for age-structured populations with genetically variable traits. *ESAIM : Proceedings*, 27 :289–310, 2009. ISSN 1270-900X. doi : 10.1051/proc/2009033. URL <https://www.esaim-proc.org/articles/proc/abs/2009/02/proc092717/proc092717.html>. Publisher : EDP Sciences. → p. 34

Forien, R., Pang, G., and Pardoux, Estimating the state of the COVID-19 epidemic in France using a model with memory. *Royal Society Open Science*, 8(3) :202327, March 2021. doi : 10.1098/rsos.202327. URL <https://royalsocietypublishing.org/doi/10.1098/rsos.202327>. Publisher : Royal Society. → pp. 102 et 103

Forni, D., Cagliani, R., Clerici, M., and Sironi, M. Molecular Evolution of Human Coronavirus Genomes. *Trends in Microbiology*, 25(1) :35–48, January 2017. ISSN 0966-842X. doi : 10.1016/j.tim.2016.09.001. URL <https://www.sciencedirect.com/science/article/pii/S0966842X16301330>. → p. 18

Forterre, P. Defining Life : The Virus Viewpoint. *Origins of Life and Evolution of Biospheres*, 40(2) :151–160, April 2010. ISSN 1573-0875. doi : 10.1007/s11084-010-9194-1. URL <https://doi.org/10.1007/s11084-010-9194-1>. → p. 3

Forterre, P. To be or not to be alive : How recent discoveries challenge the traditional definitions of viruses and life. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 59 :100–108, October 2016. ISSN 1369-8486. doi : 10.1016/j.shpsc.2016.02.013. URL <https://www.sciencedirect.com/science/article/pii/S1369848616300085>. → p. 3

- Forterre, P. and Gaïa, M. Les virus et l'émergence des cellules eucaryotes modernes. *Médecine/Sciences*, 38(12) :990, 2022. doi : 10.1051/medsci/2022164. URL <https://hal.science/hal-03897504>. → pp. 3 et 4
- Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálffy, M., Nanni, F., and Coates, J. A. The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLOS Biology*, 19(4) :e3000959, April 2021. ISSN 1545-7885. doi : 10.1371/journal.pbio.3000959. URL <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000959>. Publisher : Public Library of Science. → p. 128
- French, R. K. and Holmes, E. C. An Ecosystems Perspective on Virus Evolution and Emergence. *Trends in Microbiology*, 28(3) :165–175, March 2020. ISSN 0966-842X. doi : 10.1016/j.tim.2019.10.010. URL <https://www.sciencedirect.com/science/article/pii/S0966842X19302653>. → p. 3
- Fuchs, E. and Nowak, J. Building epithelial tissues from skin stem cells. *Cold Spring Harbor Symposia on Quantitative Biology*, 73, 1 2008. ISSN 0091-7451. doi : 10.1101/sqb.2008.73.032. → p. 99
- Gaymard, A., Bosetti, P., Feri, A., Destras, G., Enouf, V., Andronico, A., Burrel, S., Behillil, S., Sauvage, C., Bal, A., Morfin, F., Van Der Werf, S., Josset, L., ANRS MIE AC43 COVID-19, French viro COVID group, Blanquart, F., Coignard, B., Cauchemez, S., and Lina, B. Early assessment of diffusion and possible expansion of SARS-CoV-2 Lineage 20I/501Y.V1 (B.1.1.7, variant of concern 202012/01) in France, January to March 2021. *Euro Surveillance : Bulletin Européen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, 26(9) :2100133, March 2021. ISSN 1560-7917. doi : 10.2807/1560-7917.ES.2021.26.9.2100133. → p. 104
- Geraets, D. T., Struijk, L., Kleter, B., Molijn, A., van Doorn, L.-J., Quint, W. G. V., and Colau, B. The original SPF10 LiPA25 algorithm is more sensitive and suitable for epidemiologic HPV research than the SPF10 INNO-LiPA Extra. *J Virol Meth*, 215-216 :22–29, 2015a. ISSN 0166-0934. doi : 10.1016/j.jviromet.2015.01.001. → pp. 52 et 53
- Geraets, D. T., Struijk, L., Kleter, B., Molijn, A., van Doorn, L.-J., Quint, W. G. V., and Colau, B. The original SPF10 LiPA25 algorithm is more sensitive and suitable for epidemiologic HPV research than the SPF10 INNO-LiPA Extra. *Journal of Virological Methods*, 215-216 :22–29, April 2015b. ISSN 1879-0984. doi : 10.1016/j.jviromet.2015.01.001. → pp. 49, 68, 78 et 79
- Gerend, M. A. and Magloire, Z. F. Awareness, Knowledge, and Beliefs about Human Papillomavirus in a Racially Diverse Sample of Young Adults. *Journal of Adolescent Health*, 42(3) :237–242, March 2008. ISSN 1054-139X, 1879-1972. doi : 10.1016/j.jadohealth.2007.08.022. URL [https://www.jahonline.org/article/S1054-139X\(07\)00404-1/fulltext](https://www.jahonline.org/article/S1054-139X(07)00404-1/fulltext). Publisher : Elsevier. → p. 123

- Ghebreyesus, T. A. and Jakab, Z. *Global strategy to accelerate the elimination of cervical cancer as a public health problem*. 2020. ISBN 978-92-4-001410-7. OCLC : 1390785466. → p. 16
- Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25) :2340–2361, December 1977. ISSN 0022-3654. doi : 10.1021/j100540a008. URL <https://doi.org/10.1021/j100540a008>. Publisher : American Chemical Society. → p. 32
- Gillison, M. L. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *Journal of the National Cancer Institute*, 92, 5 2000. ISSN 14602105. doi : 10.1093/jnci/92.9.709. → p. 98
- Goldie, S. J., Goldhaber-Fiebert, J. D., and Garnett, G. P. Chapter 18 : Public health policy for cervical cancer prevention : The role of decision science, economic evaluation, and mathematical modeling. *Vaccine*, 24 :S155–S163, August 2006. ISSN 0264-410X. doi : 10.1016/j.vaccine.2006.05.112. URL <https://www.sciencedirect.com/science/article/pii/S0264410X0606591>. → p. 122
- Goodman, M. T., Shvetsov, Y. B., McDuffie, K., Wilkens, L. R., Zhu, X., Thompson, P. J., Ning, L., Killeen, J., Kamemoto, L., and Hernandez, B. Y. Prevalence, Acquisition, and Clearance of Cervical Human Papillomavirus Infection among Women with Normal Cytology : Hawaii Human Papillomavirus Cohort Study. *Cancer Research*, 68(21) :8813–8824, 2008. doi : 10.1158/0008-5472.CAN-08-1380. → pp. 50 et 61
- Goodman, S. N., Berlin, J., Fletcher, S. W., and Fletcher, R. H. Manuscript Quality before and after Peer Review and Editing at *Annals of Internal Medicine*. *Annals of Internal Medicine*, 121(1) :11–21, July 1994. ISSN 0003-4819. doi : 10.7326/0003-4819-121-1-199407010-00003. URL <https://www.acpjournals.org/doi/10.7326/0003-4819-121-1-199407010-00003>. Publisher : American College of Physicians. → p. 128
- Graham, A. L., Allen, J. E., and Read, A. F. Evolutionary Causes and Consequences of Immunopathology. *Annual Review of Ecology, Evolution, and Systematics*, 36(1) :373–397, December 2005. ISSN 1543-592X, 1545-2069. doi : 10.1146/annurev.ecolsys.36.102003.152622. URL <https://www.annualreviews.org/doi/10.1146/annurev.ecolsys.36.102003.152622>. → p. 11
- Graham, S. The human papillomavirus replication cycle, and its links to cancer progression : A comprehensive review. *Clinical Science*, 131(17), 2017. ISSN 01435221. doi : 10.1042/CS20160786. Publisher : Portland Press, Ltd. → pp. 14, 125 et 126
- Grant, A. Dynamics of COVID-19 epidemics : SEIR models underestimate peak infection rates and overestimate epidemic duration. *medRxiv*, page 2020.04.02.20050674, April 2020. ISSN 2005-0674. doi : 10.1101/2020.04.02.20050674. URL <https://www.medrxiv.org/content/10.1101/2020.04.02.20050674v2>. Publisher : Cold Spring Harbor Laboratory Press. → pp. 107, 113 et 118

- Grassly, N. C., Fraser, C., and Garnett, G. P. Host immunity and synchronized epidemics of syphilis across the United States. *Nature*, 433(7024) :417–421, January 2005. ISSN 1476-4687. doi : 10.1038/nature03072. → p. 23
- Gravitt, P. and Winer, R. Natural History of HPV Infection across the Lifespan : Role of Viral Latency. *Viruses*, 9(10) :267, September 2017. ISSN 1999-4915. doi : 10.3390/v9100267. URL <http://www.mdpi.com/1999-4915/9/10/267>. → pp. 15, 48, 52, 64, 70, 78 et 79
- Gravitt, P. E. The known unknowns of HPV natural history. *J Clin Invest*, 121(12) :4593–4599, 2011. doi : 10.1172/JCI57149. → pp. 41, 125 et 212
- Gravitt, P. E. Evidence and Impact of Human Papillomavirus Latency. *The Open Virology Journal*, 6(1), December 2012. ISSN 18743579. doi : 10.2174/1874357901206010198. → pp. 15, 48, 64, 84 et 125
- Groves, I. J. and Coleman, N. Pathogenesis of human papillomavirus-associated mucosal disease. *The Journal of Pathology*, 235(4) :527–538, March 2015. ISSN 1096-9896. doi : 10.1002/path.4496. → p. 126
- Haccou, P., Jagers, P., and Vatutin, V. A. *Branching Processes : Variation, Growth, and Extinction of Populations*. Cambridge University Press, Cambridge, 2005. ISBN 978-0-511-62913-6. OCLC : 776965631. → p. 28
- Hall, M. T., Simms, K. T., Lew, J.-B., Smith, M. A., Brotherton, J. M., Saville, M., Frazer, I. H., and Canfell, K. The projected timeframe until cervical cancer elimination in Australia : a modelling study. *The Lancet Public Health*, 4(1) :e19–e27, January 2019. ISSN 2468-2667. doi : 10.1016/S2468-2667(18)30183-X. URL [https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667\(18\)30183-X/fulltext](https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(18)30183-X/fulltext). Publisher : Elsevier. → p. 123
- Harper, D. M., Franco, E. L., Wheeler, C., Ferris, D. G., Jenkins, D., Schuind, A., Zahaf, T., Innis, B., Naud, P., De Carvalho, N. S., Roteli-Martins, C. M., Teixeira, J., Blatter, M. M., Korn, A. P., Quint, W., and Dubin, G. Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women : a randomised controlled trial. *The Lancet*, 364(9447) :1757–1765, November 2004. ISSN 0140-6736. doi : 10.1016/S0140-6736(04)17398-4. URL <https://www.sciencedirect.com/science/article/pii/S0140673604173984>. → p. 67
- Harper, D. M., Franco, E. L., Wheeler, C. M., Moscicki, A.-B., Romanowski, B., Roteli-Martins, C. M., Jenkins, D., Schuind, A., Costa Clemens, S. A., Dubin, G., and HPV Vaccine Study group. Sustained efficacy up to 4.5 years of a bivalent L1 virus-like particle vaccine against human papillomavirus types 16 and 18 : follow-up from a randomised control trial. *Lancet (London, England)*, 367(9518) :1247–1255, April 2006. ISSN 1474-547X. doi : 10.1016/S0140-6736(06)68439-0. → p. 67
- Harris, T. E. Branching processes. *The Annals of Mathematical Statistics*, 19, 12 1948. ISSN 0003-4851. doi : 10.1214/aoms/1177730146. → p. 88

- Hartfield, M. and Alizon, S. Introducing the outbreak threshold in epidemiology. *PLoS Pathog.*, 6 :e1003277, 2013. doi : 10.1371/journal.ppat.1003277. → pp. 106 et 109
- Haug, N., Geyrhofer, L., Londei, A., Dervic, E., Desvars-Larrive, A., Loreto, V., Piniór, B., Thurner, S., and Klimek, P. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behaviour*, 4(12) :1303–1312, December 2020. ISSN 2397-3374. doi : 10.1038/s41562-020-01009-0. URL <https://www.nature.com/articles/s41562-020-01009-0>. Number : 12 Publisher : Nature Publishing Group. → p. 124
- Haute Autorité de Santé. Avis de la commission de la transparence sur le gardasil 9. https://has-sante.fr/jcms/p_3160954/fr/gardasil-9-vaccin-papillomavirus-humain-9-valent-recombinant-adsorbe, 2020. Accessed : 2023-08-10. → p. 17
- Hay, J. A., Kennedy-Shaffer, L., Kanjilal, S., Lennon, N. J., Gabriel, S. B., Lipsitch, M., and Mina, M. J. Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science*, 373(6552) :eabh0635, June 2021. doi : 10.1126/science.abh0635. URL <https://www.science.org/doi/10.1126/science.abh0635>. Publisher : American Association for the Advancement of Science. → p. 127
- He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., Zhang, F., Cowling, B. J., Li, F., and Leung, G. M. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med*, pages 1–4, April 2020. ISSN 1546-170X. doi : 10.1038/s41591-020-0869-5. URL <https://www.nature.com/articles/s41591-020-0869-5>. Publisher : Nature Publishing Group. → p. 107
- Hellewell, J., Abbott, S., Gimma, A., Bosse, N. I., Jarvis, C. I., Russell, T. W., Munday, J. D., Kucharski, A. J., Edmunds, W. J., Sun, F., Flasche, S., Quilty, B. J., Davies, N., Liu, Y., Clifford, S., Klepac, P., Jit, M., Diamond, C., Gibbs, H., Zandvoort, K. v., Funk, S., and Eggo, R. M. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 0(0), February 2020. ISSN 2214-109X. doi : 10.1016/S2214-109X(20)30074-7. URL [https://www.thelancet.com/journals/langlo/article/PIIS2214-109X\(20\)30074-7/abstract](https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(20)30074-7/abstract). → p. 107
- Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E., and Hatfull, G. F. Evolutionary relationships among diverse bacteriophages and prophages : All the world’s a phage. *Proceedings of the National Academy of Sciences*, 96 (5) :2192–2197, March 1999. doi : 10.1073/pnas.96.5.2192. URL <https://www.pnas.org/doi/10.1073/pnas.96.5.2192>. Publisher : Proceedings of the National Academy of Sciences. → p. 3
- Herbenick, D., Reece, M., Schick, V., Sanders, S. A., Dodge, B., and Fortenberry, J. D. Sexual behavior in the United States : results from a national probability sample of men and women ages 14-94. *The Journal of Sexual Medicine*, 7 Suppl 5 :255–265, October 2010. ISSN 1743-6109. doi : 10.1111/j.1743-6109.2010.02012.x. → p. 60

- Herfs, M. and Crum, C. P. Squamocolumnar junction ablation—tying up loose ends? *Nature Reviews Clinical Oncology*, 12(7) :378–380, July 2015. ISSN 1759-4782. doi : 10.1038/nrclinonc.2015.104. URL <https://www.nature.com/articles/nrclinonc.2015.104>. → p. 15
- Herfs, M., Yamamoto, Y., Laury, A., Wang, X., Nucci, M. R., McLaughlin-Drubin, M. E., Münger, K., Feldman, S., McKeon, F. D., Xian, W., and Crum, C. P. A discrete population of squamocolumnar junction cells implicated in the pathogenesis of cervical cancer. *Proc Natl Acad Sci USA*, 109(26) :10516–21, 2012. doi : 10.1073/pnas.1202684109. → p. 15
- Herse, F. and Reissell, E. The annual costs associated with human papillomavirus types 6, 11, 16, and 18 infections in Finland. *Scandinavian Journal of Infectious Diseases*, 43(3) :209–215, March 2011. ISSN 0036-5548. doi : 10.3109/00365548.2010.541492. URL <https://doi.org/10.3109/00365548.2010.541492>. Publisher : Taylor & Francis _eprint : <https://doi.org/10.3109/00365548.2010.541492>. → p. 13
- Hethcote, H. W. The mathematics of infectious diseases. *SIAM review*, 42(4) : 599–653, 2000. doi : 10.1137/S0036144500371907. → p. 25
- Hirsch, M. W. and Smale, S. *Differential equations, dynamical systems, and linear algebra*. Number v. 60 in Pure and applied mathematics ; a series of monographs and textbooks. Academic Press, New York, 1974. ISBN 978-0-12-349550-1. → p. 26
- Huang, X., Shao, X., Xing, L., Hu, Y., Sin, D. D., and Zhang, X. The impact of lockdown timing on COVID-19 transmission across US counties. 38. ISSN 2589-5370. doi : 10.1016/j.eclinm.2021.101035. URL [https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370\(21\)00315-1/fulltext](https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(21)00315-1/fulltext). Publisher : Elsevier. → p. 127
- Huisman, J. S., Scire, J., Caduff, L., Fernandez-Cassi, X., Ganesanandamoorthy, P., Kull, A., Scheidegger, A., Stachler, E., Boehm, A. B., Hughes, B., Knudson, A., Topol, A., Wigginton, K. R., Wolfe, M. K., Kohn, T., Ort, C., Stadler, T., and Julian, T. R. Wastewater-Based Estimation of the Effective Reproductive Number of SARS-CoV-2. *Environmental Health Perspectives*, 130(5) :57011, May 2022. ISSN 1552-9924. doi : 10.1289/EHP10050. → p. 20
- Hull, R., Mbele, M., Makhafola, T., Hicks, C., Wang, S.-M., Reis, R. M., Mehrotra, R., Mkhize-Kwitshana, Z., Kibiki, G., Bates, D. O., and Dlamini, Z. Cervical cancer in low and middle-income countries. *Oncology Letters*, 20(3) : 2058–2074, September 2020. ISSN 1792-1074. doi : 10.3892/ol.2020.11754. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7400218/>. → p. 123
- Hurford, A., Cownden, D., and Day, T. Next-generation tools for evolutionary invasion analyses. *Journal of The Royal Society Interface*, 7(45) :561–571, December 2009. doi : 10.1098/rsif.2009.0448. URL <https://royalsocietypublishing.org/doi/full/10.1098/rsif.2009.0448>. Publisher : Royal Society. → p. 26

- Hurvich, C. M. and Tsai, C.-L. A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection. *Journal of Time Series Analysis*, 14 (3) :271–279, 1993. doi : 10.1111/j.1467-9892.1993.tb00144.x. → pp. 54 et 179
- Hutzler, F., Richlan, F., Leitner, M. C., Schuster, S., Braun, M., and Hawelka, S. Anticipating trajectories of exponential growth. *Royal Society Open Science*, 8(4) :201574, April 2021. doi : 10.1098/rsos.201574. URL <https://royalsocietypublishing.org/doi/10.1098/rsos.201574>. Publisher : Royal Society. → p. 2
- IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. *Biological Agents*, volume 100B of *IARC monographs on the evaluation of carcinogenic risks to humans*. International Agency for Research on Cancer, Lyon, France, 2012. ISBN 978-92-832-1319-2 978-92-832-0134-2. → pp. 13 et 53
- Idehen, E. E., Korhonen, T., Castaneda, A., Juntunen, T., Kangasniemi, M., Pietilä, A.-M., and Koponen, P. Factors associated with cervical cancer screening participation among immigrants of Russian, Somali and Kurdish origin : a population-based study in Finland. *BMC Women's Health*, 17(1) : 19, March 2017. ISSN 1472-6874. doi : 10.1186/s12905-017-0375-1. URL <https://doi.org/10.1186/s12905-017-0375-1>. → p. 124
- Ikesu, R., Taguchi, A., Hara, K., Kawana, K., Tsuruga, T., Tomio, J., and Osuga, Y. Prognosis of high-risk human papillomavirus-related cervical lesions : A hidden Markov model analysis of a single-center cohort in Japan. *Cancer Medicine*, 11(3) :664–675, 2022. ISSN 2045-7634. doi : 10.1002/cam4.4470. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/cam4.4470>. _eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/cam4.4470>. → p. 65
- Insinga, R. P., Dasbach, E. J., Elbasha, E. H., Liaw, K.-L., and Barr, E. Incidence and duration of cervical human papillomavirus 6, 11, 16, and 18 infections in young women : An evaluation from multiple analytic perspectives. *Cancer Epidemiology Biomarkers & Prevention*, 16 :709–715, 4 2007. ISSN 1055-9965. doi : 10.1158/1055-9965.EPI-06-0846. URL <http://cebp.aacrjournals.org/cgi/doi/10.1158/1055-9965.EPI-06-0846>. → pp. 15, 48, 49, 64, 80, 86 et 125
- International HPV reference center. HPV reference clones, karolinska institutet. https://www.hpvcenter.se/human_reference_clones/. Accessed : 2023-08-07. → p. 12
- Jackson, C. Multi-State Models for Panel Data : The msm Package for R. *Journal of Statistical Software*, 38 :1–28, January 2011a. ISSN 1548-7660. doi : 10.18637/jss.v038.i08. URL <https://doi.org/10.18637/jss.v038.i08>. → p. 44
- Jackson, C. H. Multi-state models for panel data : The msm package for R. *Journal of Statistical Software*, 38(8) :1–29, 2011b. doi : 10.18637/jss.v038.i08. → pp. 69 et 70

- Jackson, C. B., Farzan, M., Chen, B., and Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology*, 23(1) :3–20, January 2022. ISSN 1471-0080. doi : 10.1038/s41580-021-00418-x. URL <https://www.nature.com/articles/s41580-021-00418-x>. Number : 1 Publisher : Nature Publishing Group. → p. 19
- Jacob, F. *La logique du vivant : une histoire de l'hérédité*. Number 2 in Tel. Gallimard, Paris, repr edition, 1987. ISBN 978-2-07-029334-6. → p. 125
- Jacobs, E. A., Karavolos, K., Rathouz, P. J., Ferris, T. G., and Powell, L. H. Limited English proficiency and breast and cervical cancer screening in a multiethnic population. *American Journal of Public Health*, 95(8) :1410–1416, August 2005. ISSN 0090-0036. doi : 10.2105/AJPH.2004.041418. → p. 124
- Jansen, E. E. L., Zielonke, N., Gini, A., Anttila, A., Segnan, N., Vokó, Z., Ivanuš, U., McKee, M., de Koning, H. J., de Kok, I. M. C. M., Veerus, P., Anttila, A., Heinävaara, S., Sarkeala, T., Csanádi, M., Pitter, J., Széles, G., Vokó, Z., Minozzi, S., Segnan, N., Senore, C., van Ballegooijen, M., Driesprong de Kok, I., Gini, A., Heijnsdijk, E., Jansen, E., de Koning, H., Lansdorp – Vogelaar, I., van Ravesteyn, N., Zielonke, N., Ivanus, U., Jarm, K., Mlakar, D. N., Primic-Žakelj, M., McKee, M., and Priaulx, J. Effect of organised cervical cancer screening on cervical cancer mortality in Europe : a systematic review. *European Journal of Cancer*, 127 :207–223, March 2020. ISSN 0959-8049. doi : 10.1016/j.ejca.2019.12.013. URL <https://www.sciencedirect.com/science/article/pii/S0959804919308688>. → p. 62
- Jit, M., Brisson, M., Portnoy, A., and Hutubessy, R. Cost-effectiveness of female human papillomavirus vaccination in 179 countries : a PRIME modelling study. *The Lancet. Global Health*, 2(7) :e406–414, July 2014. ISSN 2214-109X. doi : 10.1016/S2214-109X(14)70237-2. → p. 123
- Joura, E. A., Giuliano, A. R., Iversen, O.-E., Bouchard, C., Mao, C., Mehlsen, J., Moreira, Jr, E. D., Ngan, Y., Petersen, L. K., Lazcano-Ponce, E., Piti-suttithum, P., Restrepo, J. A., Stuart, G., Woelber, L., Yang, Y. C., Cuzick, J., Garland, S. M., Huh, W., Kjaer, S. K., Bautista, O. M., Chan, I. S. F., Chen, J., Gesser, R., Moeller, E., Ritter, M., Vuocolo, S., Luxembourg, A., and Broad Spectrum HPV Vaccine Study. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. *N Engl J Med*, 372(8) : 711–23, 2015. doi : 10.1056/NEJMoa1405044. → p. 17
- Kamal, M. Pap Smear Collection and Preparation : Key Points. *CytoJournal*, 19 :24, March 2022. ISSN 0974-5963. doi : 10.25259/CMAS_03_05_2021. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9063692/>. → p. 64
- Kaplan, E. L. and Meier, P. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282) : 457–481, 1958. ISSN 0162-1459. doi : 10.2307/2281868. URL <https://www.jstor.org/stable/2281868>. Publisher : [American Statistical Association, Taylor & Francis, Ltd.]. → p. 37

- Keeling, M. J. and Rohani, P. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2008. → pp. 2, 24, 83 et 125
- Kermack, W. O., McKendrick, A. G., and Walker, G. T. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772) :700–721, January 1997a. doi : 10.1098/rspa.1927.0118. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1927.0118>. Publisher : Royal Society. → pp. 20, 23 et 26
- Kermack, W. O., McKendrick, A. G., and Walker, G. T. Contributions to the mathematical theory of epidemics. II. —The problem of endemicity. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 138(834) :55–83, January 1997b. doi : 10.1098/rspa.1932.0171. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1932.0171>. Publisher : Royal Society. → pp. 20, 23 et 26
- Kermack, W. O., McKendrick, A. G., and Walker, G. T. Contributions to the mathematical theory of epidemics. III.—Further studies of the problem of endemicity. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 141(843) :94–122, January 1997c. doi : 10.1098/rspa.1933.0106. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1933.0106>. Publisher : Royal Society. → pp. 20, 23 et 26
- Kines, R. C., Thompson, C. D., Lowy, D. R., Schiller, J. T., and Day, P. M. The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding. *Proc Natl Acad Sci USA*, 106(48) : 20458–63, 2009. doi : 10.1073/pnas.0908502106. → p. 14
- Klein, A. M., Doupé, D. P., Jones, P. H., and Simons, B. D. Kinetics of cell division in epidermal maintenance. *Physical Review E*, 76 :021910, 8 2007. ISSN 1539-3755. doi : 10.1103/PhysRevE.76.021910. URL <https://link.aps.org/doi/10.1103/PhysRevE.76.021910>. → p. 100
- Kleinbaum, D. G. and Klein, M. *Survival Analysis*. Springer New York, New York, NY, 2012. ISBN 978-1-4419-6645-2. doi : 10.1007/978-1-4419-6646-9. → pp. 34, 35, 38 et 40
- Kleter, B., van Doorn, L. J., Schrauwen, L., Molijn, A., Sastrowijoto, S., ter Schegget, J., Lindeman, J., ter Harmsel, B., Burger, M., and Quint, W. Development and clinical evaluation of a highly sensitive PCR-reverse hybridization line probe assay for detection and identification of anogenital human papillomavirus. *Journal of Clinical Microbiology*, 37(8) :2508–2517, August 1999. ISSN 0095-1137. doi : 10.1128/JCM.37.8.2508-2517.1999. → pp. 48 et 68
- Kleter, B., Doorn, L.-J. v., Schegget, J. t., Schrauwen, L., Krimpen, K. v., Burger, M., Harmsel, B. t., and Quint, W. Novel Short-Fragment PCR Assay for Highly Sensitive Broad-Spectrum Detection of Anogenital Human

- Papillomaviruses. *The American Journal of Pathology*, 153(6) :1731–1739, December 1998a. ISSN 0002-9440, 1525-2191. doi : 10.1016/S0002-9440(10)65688-X. URL [https://ajp.amjpathol.org/article/S0002-9440\(10\)65688-X/fulltext](https://ajp.amjpathol.org/article/S0002-9440(10)65688-X/fulltext). Publisher : Elsevier. → pp. 48 et 68
- Kleter, B., van Doorn, L.-J., ter Schegget, J., Schrauwen, L., van Krimpen, K., Burger, M., ter Harnsel, B., and Quint, W. Novel Short-Fragment PCR Assay for Highly Sensitive Broad-Spectrum Detection of Anogenital Human Papillomaviruses. *Am J Pathol*, 153(6) :1731–1739, 1998b. ISSN 0002-9440. doi : 10.1016/S0002-9440(10)65688-X. → p. 53
- Konno, R., Tamura, S., Dobbelaere, K., and Yoshikawa, H. Efficacy of human papillomavirus type 16/18 AS04-adjuvanted vaccine in Japanese women aged 20 to 25 years : final analysis of a phase 2 double-blind, randomized controlled trial. *International Journal of Gynecological Cancer : Official Journal of the International Gynecological Cancer Society*, 20(5) :847–855, July 2010. ISSN 1525-1438. doi : 10.1111/IGC.0b013e3181da2128. → p. 68
- Konno, R., Yoshikawa, H., Okutani, M., Quint, W., V Suryakiran, P., Lin, L., and Struyf, F. Efficacy of the human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical intraepithelial neoplasia and cervical infection in young Japanese women. *Human Vaccines & Immunotherapeutics*, 10(7) :1781–1794, 2014. ISSN 2164-554X. doi : 10.4161/hv.28712. → pp. 68, 185 et 216
- Koonin, E. V. and Starokadomskyy, P. Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. *Studies in History and Philosophy of Science Part C : Studies in History and Philosophy of Biological and Biomedical Sciences*, 59 :125–134, October 2016. ISSN 1369-8486. doi : 10.1016/j.shpsc.2016.02.016. URL <https://www.sciencedirect.com/science/article/pii/S1369848616300103>. → p. 3
- Koonin, E. V., Dolja, V. V., Krupovic, M., and Kuhn, J. H. Viruses Defined by the Position of the Virosphere within the Replicator Space. *Microbiology and Molecular Biology Reviews*, 85(4) :e00193–20, September 2021. doi : 10.1128/MMBR.00193-20. URL <https://journals.asm.org/doi/full/10.1128/membr.00193-20>. Publisher : American Society for Microbiology. → p. 3
- Kornberg, A. and Baker, T. A. *DNA replication*. University Science Books, Sausalito, Calif, 2. ed., paperback ed edition, 2005. ISBN 978-1-891389-44-3. → p. 7
- Krupovic, M., Dolja, V. V., and Koonin, E. V. Origin of viruses : primordial replicators recruiting capsids from hosts. *Nature Reviews. Microbiology*, 17 (7) :449–458, July 2019. ISSN 1740-1534. doi : 10.1038/s41579-019-0205-6. → p. 4
- Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F.,

- Ling, A.-E., Humphrey, C. D., Shieh, W.-J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., DeRisi, J., Yang, J.-Y., Cox, N., Hughes, J. M., LeDuc, J. W., Bellini, W. J., Anderson, L. J., and SARS Working Group. A novel coronavirus associated with severe acute respiratory syndrome. *The New England Journal of Medicine*, 348(20) :1953–1966, May 2003. ISSN 1533-4406. doi : 10.1056/NEJMoa030781. → p. 18
- Kurtz, T. G. Limit Theorems for Sequences of Jump Markov Processes Approximating Ordinary Differential Processes. *Journal of Applied Probability*, 8(2) :344–356, 1971. ISSN 0021-9002. doi : 10.2307/3211904. URL <https://www.jstor.org/stable/3211904>. Publisher : Applied Probability Trust. → pp. 28 et 32
- Kurtz, T. G. Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes. *Journal of Applied Probability*, 7(1) :49–58, 1970. ISSN 0021-9002. doi : 10.2307/3212147. URL <https://www.jstor.org/stable/3212147>. Publisher : Applied Probability Trust. → pp. 28 et 32
- Kurtz, T. G. The Relationship between Stochastic and Deterministic Models for Chemical Reactions. *The Journal of Chemical Physics*, 57(7) :2976–2978, September 2003. ISSN 0021-9606. doi : 10.1063/1.1678692. URL <https://doi.org/10.1063/1.1678692>. → pp. 28 et 32
- Ky, D. X. and Tuyen, L. T. A Higher order Markov model for time series forecasting. *International Journal of Applied Mathematics and Statistics™*, 57(3) :1–18, April 2018. ISSN 0973-7545. URL <http://www.ceser.in/ceser/index.php/ijamas/article/view/5473>. Number : 3. → p. 33
- Kyriakidis, N. C., López-Cortés, A., González, E. V., Grimaldos, A. B., and Prado, E. O. SARS-CoV-2 vaccines strategies : a comprehensive review of phase 3 candidates. *NPJ Vaccines*, 6 :28, February 2021. ISSN 2059-0105. doi : 10.1038/s41541-021-00292-w. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7900244/>. → p. 17
- Law, C. G. and Brookmeyer, R. Effects of mid-point imputation on the analysis of doubly censored data. *Statistics in Medicine*, 11(12), 1992. ISSN 02776715. doi : 10.1002/sim.4780111204. → pp. 49 et 178
- Ledwaba, T., Dlamini, Z., Naicker, S., and Bhoola, K. Molecular genetics of human cervical cancer : role of papillomavirus and the apoptotic cascade. *Biological Chemistry*, 385, 1 2004. ISSN 1431-6730. doi : 10.1515/BC.2004.083. URL <https://www.degruyter.com/view/j/bchm.2004.385.issue-8/bc.2004.083/bc.2004.083.xml>. → p. 99
- Levins, R. The Strategy of Model Building in Population Biology. *American Scientist*, 54(4) :421–431, 1966. ISSN 0003-0996. URL <https://www.jstor.org/stable/27836590>. Publisher : Sigma Xi, The Scientific Research Society. → p. 21
- Li, J. Persistence and extinction in continuous age-structured population models. *Computers & Mathematics with Applications*, 15(6) :511–523, January 1988. ISSN 0898-1221. doi : 10.1016/0898-1221(88)90277-5. URL <https://>

[//www.sciencedirect.com/science/article/pii/S0898122188902775](https://www.sciencedirect.com/science/article/pii/S0898122188902775).
→ p. 34

- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T. T. Y., Wu, J. T., Gao, G. F., Cowling, B. J., Yang, B., Leung, G. M., and Feng, Z. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *New England Journal of Medicine*, January 2020. doi : 10.1056/NEJMoa2001316. URL <https://www.nejm.org/doi/10.1056/NEJMoa2001316>. → p. 203
- Liaw, Y.-F. and Zoulim, F., editors. *Hepatitis B Virus in Human Diseases*. Molecular and Translational Medicine. Springer International Publishing, Cham, Switzerland, 1 edition, November 2015. → p. 126
- Lin, D.-Y., Gu, Y., Wheeler, B., Young, H., Holloway, S., Sunny, S.-K., Moore, Z., and Zeng, D. Effectiveness of Covid-19 Vaccines over a 9-Month Period in North Carolina. *New England Journal of Medicine*, 386(10) :933–941, March 2022. ISSN 0028-4793. doi : 10.1056/NEJMoa2117128. URL <https://doi.org/10.1056/NEJMoa2117128>. Publisher : Massachusetts Medical Society _eprint : <https://doi.org/10.1056/NEJMoa2117128>. → p. 19
- Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S.-m., Yuan, B., Kinoshita, R., and Nishiura, H. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation : A Statistical Analysis of Publicly Available Case Data. *Journal of Clinical Medicine*, 9(2) :538, February 2020. doi : 10.3390/jcm9020538. URL <https://www.mdpi.com/2077-0383/9/2/538>. → pp. 110 et 200
- Liu, Y., Eggo, R. M., and Kucharski, A. J. Secondary attack rate and superspreading events for SARS-CoV-2. *The Lancet*, 395(10227) :e47, March 2020a. ISSN 0140-6736, 1474-547X. doi : 10.1016/S0140-6736(20)30462-1. URL [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(20\)30462-1/abstract](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30462-1/abstract). → pp. 49, 109 et 119
- Liu, Y., Ning, Z., Chen, Y., Guo, M., Liu, Y., Gali, N. K., Sun, L., Duan, Y., Cai, J., Westerdahl, D., Liu, X., Xu, K., Ho, K.-f., Kan, H., Fu, Q., and Lan, K. Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. *Nature*, pages 1–6, April 2020b. ISSN 1476-4687. doi : 10.1038/s41586-020-2271-3. URL <https://www.nature.com/articles/s41586-020-2271-3>. Publisher : Nature Publishing Group. → p. 103
- Liu, Z., Rashid, T., and Nyitray, A. G. Penises not required : a systematic review of the potential for human papillomavirus horizontal transmission that is non-sexual or does not include penile penetration. *Sexual Health*, 13 (1) :10–21, February 2016. ISSN 1448-5028. doi : 10.1071/SH15089. → p. 16

- Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., and Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438 (7066) :355–9, 2005. doi : 10.1038/nature04153. → pp. 102, 103, 107, 109, 112, 114, 118, 119 et 200
- Lucas, W. Viral Capsids and Envelopes : Structure and Function. In *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd, 2010. ISBN 978-0-470-01590-2. doi : 10.1002/9780470015902.a0001091.pub2. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470015902.a0001091.pub2>. _eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470015902.a0001091.pub2>. → pp. 5 et 6
- Lythgoe, K. A., Lumley, S. F., Pellis, L., McKeating, J. A., and Matthews, P. C. Estimating hepatitis B virus cccDNA persistence in chronic infection†. *Virus Evolution*, 7(1), January 2021. ISSN 2057-1577. doi : 10.1093/ve/veaa063. → p. 126
- Ma, S. T. C., Yeung, A. C., Chan, P. K. S., and Graham, C. A. Transvaginal ultrasound probe contamination by the human papillomavirus in the emergency department. *Emergency medicine journal : EMJ*, 30(6) :472–475, June 2013. ISSN 1472-0213. doi : 10.1136/emered-2012-201407. → p. 17
- MacDonald, I. L. and Zucchini, W. *Hidden Markov and other models for discrete-valued time series*. Number 70 in Monographs on statistics and applied probability. Chapman & Hall, London ; New York, 1st ed edition, 1997. ISBN 978-0-412-55850-4. → p. 45
- MacDonald, Iain L., W. Z. *Hidden Markov Models for Time Series : An Introduction Using R*. Chapman and Hall/CRC, New York, April 2009. ISBN 978-0-429-13953-6. doi : 10.1201/9781420010893. → pp. 42 et 44
- Magal, P. and Ruan, S. Susceptible-infectious-recovered models revisited : from the individual level to the population level. *Mathematical Biosciences*, 250 :26–40, April 2014. ISSN 1879-3134. doi : 10.1016/j.mbs.2014.02.001. → p. 20
- Maglennon, G. A., McIntosh, P. B., and Doorbar, J. Immunosuppression Facilitates the Reactivation of Latent Papillomavirus Infections. *Journal of Virology*, 88(1) :710, January 2014. doi : 10.1128/JVI.02589-13. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3911712/>. Publisher : American Society for Microbiology (ASM). → pp. 64 et 125
- Maglennon, G. A., McIntosh, P., and Doorbar, J. Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression. *Virology*, 414(2) :153–163, June 2011. ISSN 00426822. doi : 10.1016/j.virol.2011.03.019. URL <https://linkinghub.elsevier.com/retrieve/pii/S0042682211001449>. → pp. 64 et 125
- Mahase, E. Coronavirus : covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. *BMJ*, 368 :m641, February 2020. ISSN 1756-1833. doi : 10.1136/bmj.m641. URL <https://www.bmj.co>

- [m/content/368/bmj.m641](#). Publisher : British Medical Journal Publishing Group Section : News. → p. 18
- Malagón, T., Trottier, H., El-Zein, M., Villa, L., and Franco, E. Human papillomavirus intermittence and risk factors associated with first detections and redetections in the Ludwig-McGill cohort study of adult women. *The Journal of infectious diseases*, February 2023. ISSN 1537-6613. doi : 10.1093/infdis/jiad043. URL <https://pubmed.ncbi.nlm.nih.gov/36790831/>. Publisher : J Infect Dis. → pp. 49, 54, 56, 60, 61, 69, 79, 125 et 178
- Mantel, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 50(3) :163–170, March 1966. ISSN 0069-0112. → pp. 38 et 55
- Maqsood, R., Rodgers, R., Rodriguez, C., Handley, S. A., Ndao, I. M., Tarr, P. I., Warner, B. B., Lim, E. S., and Holtz, L. R. Discordant transmission of bacteria and viruses from mothers to babies at birth. *Microbiome*, 7(1) : 156, December 2019. ISSN 2049-2618. doi : 10.1186/s40168-019-0766-7. → p. 8
- Markov, A. Extension of the law of large numbers to quantities, depending on each other (1906). reprint. *Journal Électronique d’Histoire des Probabilités et de la Statistique [electronic only]*, 2(1b) :Article 10, 12 p., electronic only–Article 10, 12 p., electronic only, 2006. URL <http://eudml.org/doc/128778>. → p. 28
- Markov, A. A. The theory of algorithms. *Trudy Matematicheskogo Instituta Imeni VA Steklova*, 42 :3–375, 1954. → p. 28
- Markowitz, L. E., Gee, J., Chesson, H., and Stokley, S. Ten Years of Human Papillomavirus Vaccination in the United States. *Academic Pediatrics*, 18 (2, Supplement) :S3–S10, March 2018. ISSN 1876-2859. doi : 10.1016/j.acap.2017.09.014. URL <https://www.sciencedirect.com/science/article/pii/S1876285917304941>. → p. 16
- Marlow, L. A. V., Wardle, J., and Waller, J. Understanding cervical screening non-attendance among ethnic minority women in England. *British Journal of Cancer*, 113(5) :833–839, September 2015. ISSN 0007-0920. doi : 10.1038/bjc.2015.248. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4559824/>. → p. 124
- Marlow, L. A. V., Zimet, G. D., McCaffery, K. J., Ostini, R., and Waller, J. Knowledge of human papillomavirus (HPV) and HPV vaccination : An international comparison. *Vaccine*, 31(5) :763–769, January 2013. ISSN 0264-410X. doi : 10.1016/j.vaccine.2012.11.083. URL <https://www.sciencedirect.com/science/article/pii/S0264410X12017343>. → p. 123
- Marshall, G. and Jones, R. H. Multi-state models and diabetic retinopathy. *Statistics in Medicine*, 14(18) :1975–1983, 1995. ISSN 1097-0258. doi : 10.1002/sim.4780141804. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780141804>. _eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.4780141804>. → pp. 44 et 70

- Martin-Hirsch, P. P., Paraskevaïdis, E., Bryant, A., Dickinson, H. O., and Keep, S. L. Surgery for cervical intraepithelial neoplasia. *The Cochrane database of systematic reviews*, (6) :CD001318, June 2010. ISSN 1469-493X. doi : 10.1002/14651858.CD001318.pub2. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4170911/>. → p. 16
- Mattheij, I., Pollock, A. M., and Brhlikova, P. Do cervical cancer data justify HPV vaccination in India? Epidemiological data sources and comprehensiveness. *Journal of the Royal Society of Medicine*, 105(6) :250–262, June 2012. ISSN 1758-1095. doi : 10.1258/jrsm.2012.110343. → p. 123
- Matthewson, J. Trade-offs in model-building : A more target-oriented approach. *Studies in History and Philosophy of Science Part A*, 42(2) :324–333, June 2011. ISSN 0039-3681. doi : 10.1016/j.shpsa.2010.11.040. URL <https://www.sciencedirect.com/science/article/pii/S0039368110001196>. → p. 21
- Max Roser, Hannah Ritchie, E. O.-O. and Hasell, J. Coronavirus pandemic (covid-19). *Our World in Data*, 2020. <https://ourworldindata.org/coronavirus>. → p. 106
- McBride, E., Tatar, O., Rosberger, Z., Rockliffe, L., Marlow, L. A. V., Moss-Morris, R., Kaur, N., Wade, K., and Waller, J. Emotional response to testing positive for human papillomavirus at cervical cancer screening : a mixed method systematic review with meta-analysis. *Health Psychology Review*, 15 (3) :395–429, July 2021. ISSN 1743-7199. doi : 10.1080/17437199.2020.1762106. URL <https://doi.org/10.1080/17437199.2020.1762106>. Publisher : Routledge _eprint : <https://doi.org/10.1080/17437199.2020.1762106>. → p. 62
- McBryde, E. S., Meehan, M. T., Adegboye, O. A., Adekunle, A. I., Caldwell, J. M., Pak, A., Rojas, D. P., Williams, B. M., and Trauer, J. M. Role of modelling in COVID-19 policy development. *Paediatric Respiratory Reviews*, 35 :57–60, September 2020. ISSN 1526-0542. doi : 10.1016/j.prrv.2020.06.013. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7301791/>. → pp. 2, 127 et 128
- McCaffery, K., Waller, J., Nazroo, J., and Wardle, J. Social and psychological impact of HPV testing in cervical screening : a qualitative study. *Sexually Transmitted Infections*, 82(2) :169–174, April 2006. ISSN 1368-4973. doi : 10.1136/sti.2005.016436. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2564695/>. → p. 125
- McPhillips, M. G., Ozato, K., and McBride, A. A. Interaction of bovine papillomavirus e2 protein with brd4 stabilizes its association with chromatin. *Journal of Virology*, 79 :8920–8932, 7 2005. ISSN 0022-538X. doi : 10.1128/JVI.79.14.8920-8932.2005. URL <https://JVI.asm.org/content/79/14/8920>. → pp. 84 et 91
- Meleard, S. and Bansaye, V. *Stochastic Models for Structured Populations : Scaling Limits and Long Time Behavior*. Springer International Publishing,

- Cham, 2015. ISBN 978-3-319-21710-9 978-3-319-21711-6. doi : 10.1007/978-3-319-21711-6. URL <https://link.springer.com/10.1007/978-3-319-21711-6>. → p. 28
- Metcalf, C. J. E., Edmunds, W. J., and Lessler, J. Six challenges in modelling for public health policy. *Epidemics*, 10 :93–96, March 2015. ISSN 1755-4365. doi : 10.1016/j.epidem.2014.08.008. URL <https://www.sciencedirect.com/science/article/pii/S1755436514000620>. → p. 128
- Metropolis, N. and Ulam, S. The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247) :335–341, 1949. ISSN 0162-1459. doi : 10.2307/2280232. URL <https://www.jstor.org/stable/2280232>. Publisher : [American Statistical Association, Taylor & Francis, Ltd.]. → p. 32
- Ministère de la santé et de la prévention. Website : Lancement de la campagne de vaccination contre le Covid-19, 2023. URL <https://sante.gouv.fr/actualites/actualites-du-ministere/article/lancement-de-la-campagne-de-vaccination-contre-le-covid-19>. → p. 19
- Mirabello, L., Yeager, M., Yu, K., Clifford, G. M., Xiao, Y., Zhu, B., Cullen, M., Boland, J. F., Wentzensen, N., Nelson, C. W., Raine-Bennett, T., Chen, Z., Bass, S., Song, L., Yang, Q., Steinberg, M., Burdett, L., Dean, M., Roberson, D., Mitchell, J., Lorey, T., Franceschi, S., Castle, P. E., Walker, J., Zuna, R., Kreimer, A. R., Beachler, D. C., Hildesheim, A., Gonzalez, P., Porras, C., Burk, R. D., and Schiffman, M. Hpv16 e7 genetic conservation is critical to carcinogenesis. *Cell*, 170 :1164–1174.e6, 9 2017. ISSN 00928674. doi : 10.1016/j.cell.2017.08.001. URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867417308899>. → pp. 84 et 98
- Mistry, N., Wibom, C., and Evander, M. Cutaneous and mucosal human papillomaviruses differ in net surface charge, potential impact on tropism. *Virology Journal*, 5, 2008. ISSN 1743-422X. doi : 10.1186/1743-422X-5-118. → p. 98
- Montazersaheb, S., Hosseiniyan Khatibi, S. M., Hejazi, M. S., Tarhriz, V., Farjami, A., Ghasemian Sorbeni, F., Farahzadi, R., and Ghasemnejad, T. COVID-19 infection : an overview on cytokine storm and related interventions. *Virology Journal*, 19(1) :92, May 2022. ISSN 1743-422X. doi : 10.1186/s12985-022-01814-1. URL <https://doi.org/10.1186/s12985-022-01814-1>. → p. 10
- Moody, C. Mechanisms by which hpv induces a replication competent environment in differentiating keratinocytes. *Viruses*, 9 :261, 9 2017. ISSN 1999-4915. doi : 10.3390/v9090261. URL <http://www.mdpi.com/1999-4915/9/9/261>. → pp. 14, 84, 87, 88, 89 et 91
- Mora, C., McKenzie, T., Gaw, I. M., Dean, J. M., von Hammerstein, H., Knudson, T. A., Setter, R. O., Smith, C. Z., Webster, K. M., Patz, J. A., and Franklin, E. C. Over half of known human pathogenic diseases can be aggravated by climate change. *Nature Climate Change*, 12(9) :869–875, September 2022. ISSN 1758-6798. doi : 10.1038/s41558-022-01426-1. URL

<https://www.nature.com/articles/s41558-022-01426-1>. Number : 9
Publisher : Nature Publishing Group. → p. 9

Mork, J., Lie, A. K., Glattre, E., Clark, S., Hallmans, G., Jellum, E., Koskela, P., Møller, B., Pukkala, E., Schiller, J. T., Wang, Z., Youngman, L., Lehtinen, M., and Dillner, J. Human Papillomavirus Infection as a Risk Factor for Squamous-Cell Carcinoma of the Head and Neck. *New England Journal of Medicine*, 344(15), April 2001. ISSN 0028-4793. doi : 10.1056/NEJM200104123441503. → p. 125

Moscicki, A.-B., Ma, Y., Farhat, S., Darragh, T. M., Pawlita, M., Galloway, D. A., and Shiboski, S. Redetection of cervical human papillomavirus type 16 (HPV16) in women with a history of HPV16. *The Journal of Infectious Diseases*, 208(3) :403–412, August 2013. ISSN 1537-6613. doi : 10.1093/infdis/jit175. → pp. 61, 64 et 122

Msemburi, W., Karlinsky, A., Knutson, V., Aleshin-Guendel, S., Chatterji, S., and Wakefield, J. The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature*, 613(7942) :130–137, January 2023. ISSN 1476-4687. doi : 10.1038/s41586-022-05522-2. URL <https://www.nature.com/articles/s41586-022-05522-2>. Number : 7942 Publisher : Nature Publishing Group. → p. 2

Murall, C. L., Jackson, R., Zehbe, I., Boulle, N., Segondy, M., and Alizon, S. Epithelial stratification shapes infection dynamics. *PLoS Comput Biol*, 15 (1) :e1006646, 2019a. doi : 10.1371/journal.pcbi.1006646. → p. 83

Murall, C. L., McCann, K. S., and Bauch, C. T. Revising ecological assumptions about human papillomavirus interactions and type replacement. *Journal of Theoretical Biology*, 350 :98–109, 6 2014. ISSN 0022-5193. doi : 10.1016/j.jtbi.2013.12.028. URL <http://www.sciencedirect.com/science/article/pii/S0022519314000022>. → p. 87

Murall, C. L., Bauch, C. T., and Day, T. Could the human papillomavirus vaccines drive virulence evolution? *Proceedings of the Royal Society B : Biological Sciences*, 282 :20141069, 1 2015. doi : 10.1098/rspb.2014.1069. URL <https://royalsocietypublishing.org/doi/full/10.1098/rspb.2014.1069>. → p. 87

Murall, C. L., Rahmoun, M., Selinger, C., Baldellou, M., Bernat, C., Bonneau, M., Boué, V., Buisson, M., Christophe, G., D’Auria, G., Taroni, F. D., Foulongne, V., Froissart, R., Graf, C., Grasset, S., Groc, S., Hirtz, C., Jaussent, A., Lajoie, J., Lorcy, F., Picot, E., Picot, M.-C., Ravel, J., Reynes, J., Rousset, T., Seddiki, A., Teirlinck, M., Tribout, V., Tuailon, , Waterboer, T., Jacobs, N., Bravo, I. G., Segondy, M., Boulle, N., and Alizon, S. Natural history, dynamics, and ecology of human papillomaviruses in genital infections of young women : protocol of the PAPCLEAR cohort study. *BMJ Open*, 9(6) :e025129, June 2019b. doi : 10.1136/bmjopen-2018-025129. URL <http://bmjopen.bmj.com/content/9/6/e025129.abstract>. → pp. 48, 53, 56, 61, 66, 67, 178 et 214

- Mushegian, A. R. Are There 1031 Virus Particles on Earth, or More, or Fewer? *Journal of Bacteriology*, 202(9) :e00052–20, April 2020. ISSN 0021-9193. doi : 10.1128/JB.00052-20. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7148134/>. → p. 3
- Méthot, P.-O. Why do parasites harm their host? On the origin and legacy of Theobald Smith's "law of declining virulence"—1900-1980. *History and Philosophy of the Life Sciences*, 34(4) :561–601, 2012. ISSN 0391-9714. → p. 12
- Naud, P. S., Roteli-Martins, C. M., De Carvalho, N. S., Teixeira, J. C., de Borba, P. C., Sanchez, N., Zahaf, T., Catteau, G., Geeraerts, B., and Descamps, D. Sustained efficacy, immunogenicity, and safety of the HPV-16/18 AS04-adjuvanted vaccine. *Human Vaccines & Immunotherapeutics*, 10(8) :2147–2162, June 2014. ISSN 2164-5515. doi : 10.4161/hv.29532. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4896780/>. → p. 67
- Nelson, C. W. and Mirabello, L. Human papillomavirus genomics : Understanding carcinogenicity. *Tumour Virus Research*, 15 :200258, June 2023. ISSN 2666-6790. doi : 10.1016/j.tvr.2023.200258. URL <https://www.sciencedirect.com/science/article/pii/S2666679023000058>. → p. 13
- Nelson, W. Hazard Plotting for Incomplete Failure Data. *Journal of Quality Technology*, 1(1) :27–52, 1969. doi : 10.1080/00224065.1969.11980344. → pp. 37, 77 et 179
- Nelson, W. Theory and Applications of Hazard Plotting for Censored Failure Data. *Technometrics*, 14(4) :945–966, 1972. doi : 10.1080/00401706.1972.10488991. → pp. 37, 54, 77 et 179
- Nightingale, F. -. A. d. t. *Notes on matters affecting the health, efficiency and hospital administration of the British army, funded chiefly on the experience of the late war / by Florence Nightingale... ; presented by request to Secretary of State for war*. 1858. URL <https://gallica.bnf.fr/ark:/12148/bpt6k8561918>. → p. 8
- Nightingale, F. -. A. d. t. *Des soins à donner aux malades : ce qu'il faut faire, ce qu'il faut éviter / par Miss Nightingale ; ouvrage traduit de l'anglais avec l'autorisation de l'auteur, précédé d'une lettre [16 juillet 1862] de M. Guizot et d'une introduction par M. Daremberg*. 1862. URL <https://gallica.bnf.fr/ark:/12148/bpt6k5426595k>. → p. 8
- Nishimura, H., Yeh, P. T., Oguntade, H., Kennedy, C. E., and Narasimhan, M. HPV self-sampling for cervical cancer screening : a systematic review of values and preferences. *BMJ Global Health*, 6(5) :e003743, May 2021. ISSN 2059-7908. doi : 10.1136/bmjgh-2020-003743. URL <https://gh.bmj.com/content/6/5/e003743>. Publisher : BMJ Specialist Journals Section : Original research. → p. 62
- Nishiura, H., Miyamatsu, Y., and Mizumoto, K. Objective Determination of End of MERS Outbreak, South Korea, 2015. *Emerging Infectious Diseases*, 22(1) :146–148, January 2016. ISSN 1080-6040. doi : 10.3201/eid2201.1513

83. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4696716/>.
→ p. 106
- Nishiura, H., Linton, N. M., and Akhmetzhanov, A. R. Serial interval of novel coronavirus (COVID-19) infections. *International Journal of Infectious Diseases*, 93 :284–286, April 2020a. doi : 10.1016/j.ijid.2020.02.060. URL <https://doi.org/10.1016/j.ijid.2020.02.060>. → pp. 107, 108, 119 et 200
- Nishiura, H., Linton, N. M., and Akhmetzhanov, A. R. Serial interval of novel coronavirus (COVID-19) infections. *International Journal of Infectious Diseases*, 0(0), March 2020b. ISSN 1201-9712. doi : 10.1016/j.ijid.2020.02.060. URL [https://www.ijidonline.com/article/S1201-9712\(20\)30119-3/abstract](https://www.ijidonline.com/article/S1201-9712(20)30119-3/abstract). → pp. 103, 111, 203, 204 et 205
- O’Driscoll, M., Ribeiro Dos Santos, G., Wang, L., Cummings, D. A. T., Azman, A. S., Paireau, J., Fontanet, A., Cauchemez, S., and Salje, H. Age-specific mortality and immunity patterns of sars-cov-2. *Nature*, 590(7844) :140–145, Feb 2021. ISSN 1476-4687. doi : 10.1038/s41586-020-2918-0. URL <https://doi.org/10.1038/s41586-020-2918-0>. → p. 18
- Olsson, E., Lau, M., Lifvergren, S., and Chakhunashvili, A. Community collaboration to increase foreign-born women’s participation in a cervical cancer screening program in Sweden : a quality improvement project. *International Journal for Equity in Health*, 13 :62, August 2014. ISSN 1475-9276. doi : 10.1186/s12939-014-0062-x. → p. 124
- Our World in Data. Cumulative confirmed covid-19 death. https://ourworldindata.org/covid-deaths?country=FRA~OWID_WRL, 2023. Accessed : 2023-08-16. → p. 18
- Ozili, P. K. and Arun, T. Spillover of COVID-19 : Impact on the Global Economy, March 2020. URL <https://papers.ssrn.com/abstract=3562570>. → p. 2
- Patel, C., Brotherton, J. M., Pillsbury, A., Jayasinghe, S., Donovan, B., McCartney, K., and Marshall, H. The impact of 10 years of human papillomavirus (HPV) vaccination in Australia : what additional disease burden will a nonavalent vaccine prevent ? *Eurosurveillance*, 23 :1700737, October 2018. ISSN 1025-496X. doi : 10.2807/1560-7917.ES.2018.23.41.1700737. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6194907/>. → p. 123
- Patel, P., Borkowf, C. B., Brooks, J. T., Lasry, A., Lansky, A., and Mermin, J. Estimating per-act HIV transmission risk : a systematic review. *AIDS (London, England)*, 28(10) :1509–1519, June 2014. ISSN 0269-9370. doi : 10.1097/QAD.0000000000000298. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6195215/>. → p. 8
- Pathak, N., Dodds, J., Zamora, J., and Khan, K. Accuracy of urinary human papillomavirus testing for presence of cervical HPV : systematic review and meta-analysis. *BMJ*, 349 :g5264, September 2014. ISSN 1756-1833. doi : 10.1136/bmj.g5264. URL <https://www.bmj.com/content/349/bmj.g5264>.

Publisher : British Medical Journal Publishing Group Section : Research.
→ p. 124

Paul, P., Hammer, A., Rositch, A. F., Burke, A. E., Viscidi, R. P., Silver, M. I., Campos, N., Youk, A. O., and Gravitt, P. E. Rates of New Human Papillomavirus Detection and Loss of Detection in Middle-aged Women by Recent and Past Sexual Behavior. *The Journal of Infectious Diseases*, 223 (8) :1423–1432, April 2021. ISSN 0022-1899. doi : 10.1093/infdis/jiaa557. URL <https://doi.org/10.1093/infdis/jiaa557>. → pp. 49, 60, 61, 64, 69, 79 et 178

Peck, K. M. and Lauring, A. S. Complexities of Viral Mutation Rates. *Journal of Virology*, 92(14) :10.1128/jvi.01031–17, June 2018. doi : 10.1128/jvi.01031-17. URL <https://journals.asm.org/doi/10.1128/jvi.01031-17>. Publisher : American Society for Microbiology. → p. 13

Peto, R. and Peto, J. Asymptotically Efficient Rank Invariant Test Procedures. *Journal of the Royal Statistical Society. Series A (General)*, 135(2) :185–207, 1972. doi : 10.2307/2344317. → pp. 38 et 55

Pett, M. and Coleman, N. Integration of high-risk human papillomavirus : a key event in cervical carcinogenesis? *The Journal of Pathology*, 212(4) : 356–367, August 2007. ISSN 0022-3417. doi : 10.1002/path.2192. → p. 125

Pierie, J.-P. E., Walvoort, H. C., and Overbeke, A. J. P. Readers' evaluation of effect of peer review and editing on quality of articles in the Netherlands Tijdschrift voor Geneeskunde. *The Lancet*, 348(9040) :1480–1483, November 1996. ISSN 0140-6736, 1474-547X. doi : 10.1016/S0140-6736(96)05016-7. URL [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(96\)05016-7/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(96)05016-7/fulltext). Publisher : Elsevier. → p. 128

Plummer, M., Vaccarella, S., and Franceschi, S. Multiple human papillomavirus infections : the exception or the rule? *J Infect Dis*, 203(7) :891–3, 2011. doi : 10.1093/infdis/jiq146. → p. 52

Plummer, M., de Martel, C., Vignat, J., Ferlay, J., Bray, F., and Franceschi, S. Global burden of cancers attributable to infections in 2012 : a synthetic analysis. *The Lancet Global Health*, 4 :e609–e616, 9 2016a. ISSN 2214109X. doi : 10.1016/S2214-109X(16)30143-7. URL <https://linkinghub.elsevier.com/retrieve/pii/S2214109X16301437>. → p. 86

Plummer, M., de Martel, C., Vignat, J., Ferlay, J., Bray, F., and Franceschi, S. Global burden of cancers attributable to infections in 2012 : a synthetic analysis. *Lancet Global Health*, 4(9) :e609–e616, 2016b. doi : 10.1016/S2214-109X(16)30143-7. → p. 15

Poisson, S.-D. -. A. d. t. *Recherches sur la probabilité des jugements en matière criminelle et en matière civile / par S.-D. Poisson,...* 1837. URL <https://gallica.bnf.fr/ark:/12148/bpt6k110193z>. → p. 102

- Pudney, J., Quayle, A. J., and Anderson, D. J. Immunological Microenvironments in the Human Vagina and Cervix : Mediators of Cellular Immunity Are Concentrated in the Cervical Transformation Zone1. *Biology of Reproduction*, 73(6) :1253–1263, December 2005. ISSN 0006-3363. doi : 10.1095/biolreprod.105.043133. URL <https://doi.org/10.1095/biolreprod.105.043133>. → p. 84
- Pullano, G., Di Domenico, L., Sabbatini, C. E., Valdano, E., Turbelin, C., Debin, M., Guerrisi, C., Kengne-Kuetche, C., Souty, C., Hanslik, T., Blanchon, T., Boëlle, P.-Y., Figoni, J., Vaux, S., Campèse, C., Bernard-Stoecklin, S., and Colizza, V. Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature*, 590(7844) :134–139, February 2021. ISSN 1476-4687. doi : 10.1038/s41586-020-03095-6. URL <https://www.nature.com/articles/s41586-020-03095-6>. Number : 7844 Publisher : Nature Publishing Group. → p. 102
- Pyeon, D., Pearce, S. M., Lank, S. M., Ahlquist, P., and Lambert, P. F. Establishment of human papillomavirus infection requires cell cycle progression. *PLoS pathogens*, 5(2) :e1000318, February 2009. ISSN 1553-7374. doi : 10.1371/journal.ppat.1000318. → p. 14
- Pytynia, K. B., Dahlstrom, K. R., and Sturgis, E. M. Epidemiology of hpv-associated oropharyngeal cancer. *Oral Oncology*, 50, 5 2014. ISSN 13688375. doi : 10.1016/j.oraloncology.2013.12.019. → p. 98
- Quint, W. G., Scholte, G., van Doorn, L. J., Kleter, B., Smits, P. H., and Lindeman, J. Comparative analysis of human papillomavirus infections in cervical scrapes and biopsy specimens by general SPF(10) PCR and HPV genotyping. *The Journal of Pathology*, 194(1) :51–58, May 2001. ISSN 0022-3417. doi : 10.1002/path.855. → pp. 48 et 68
- R Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020a. URL <https://www.R-project.org/>. → pp. 111 et 203
- R Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020b. URL <https://www.R-project.org/>. → p. 92
- Raftery, A. E. A Model for High-Order Markov Chains. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(3) :528–539, 1985. ISSN 0035-9246. URL <https://www.jstor.org/stable/2345788>. Publisher : [Royal Statistical Society, Wiley]. → p. 33
- Ramanakumar, A. V., Naud, P., Roteli-Martins, C. M., de Carvalho, N. S., de Borba, P. C., Teixeira, J. C., Blatter, M., Moscicki, A.-B., Harper, D. M., Romanowski, B., Tyring, S. K., Ramjattan, B., Schuind, A., Dubin, G., and Franco, E. L. Incidence and duration of type-specific human papillomavirus infection in high-risk HPV-naïve women : results from the control arm of a phase II HPV-16/18 vaccine trial. *BMJ Open*, 6(8) :e011371, 2016. doi : 10.1136/bmjopen-2016-011371. → pp. 15, 49, 50, 60, 61, 69, 80 et 125

- Rampersad, S. and Tennant, P. Replication and Expression Strategies of Viruses. *Viruses*, pages 55–82, 2018. doi : 10.1016/B978-0-12-811257-1.00003-6. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7158166/>. → p. 5
- Rautava, J. and Syrjänen, S. Biology of human papillomavirus infections in head and neck carcinogenesis. *Head and Neck Pathology*, 6, 7 2012. ISSN 1936-055X. doi : 10.1007/s12105-012-0367-2. → p. 88
- Rewar, S. and Mirdha, D. Transmission of Ebola Virus Disease : An Overview. *Annals of Global Health*, 80(6) :444–451, November 2014. ISSN 2214-9996. doi : 10.1016/j.aogh.2015.02.005. URL <https://www.sciencedirect.com/science/article/pii/S2214999615000107>. → p. 11
- Reyné, B., Richard, Q., Selinger, C., Sofonea, M. T., Djidjou-Demasse, R., and Alizon, S. Non-Markovian modelling highlights the importance of age structure on Covid-19 epidemiological dynamics. *Mathematical Modelling of Natural Phenomena*, 17 :7, 2022. ISSN 0973-5348, 1760-6101. doi : 10.1051/mmnp/2022008. URL <https://www.mmnp-journal.org/articles/mmnp/abs/2022/01/mmnp220048/mmnp220048.html>. Publisher : EDP Sciences. → pp. 33, 34 et 127
- Riedel, S. Edward Jenner and the history of smallpox and vaccination. *Proceedings (Baylor University Medical Center)*, 18(1) :21–25, January 2005. ISSN 0899-8280. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1200696/>. → p. 12
- Roden, R. B., Lowy, D. R., and Schiller, J. T. Papillomavirus is resistant to desiccation. *The Journal of Infectious Diseases*, 176(4) :1076–1079, October 1997. ISSN 0022-1899. doi : 10.1086/516515. → p. 16
- Roland, K. B., Benard, V. B., Greek, A., Hawkins, N. A., and Lin, L. Changes in Knowledge and Beliefs About Human Papillomavirus and Cervical Cancer Screening Intervals in Low-Income Women After an Educational Intervention. *Journal of Primary Care & Community Health*, 7(2) :88–95, April 2016. ISSN 2150-1319. doi : 10.1177/2150131915624869. URL <https://doi.org/10.1177/2150131915624869>. Publisher : SAGE Publications Inc. → p. 124
- Roques, L., Klein, E. K., Papaïx, J., Sar, A., and Soubeyrand, S. Impact of Lockdown on the Epidemic Dynamics of COVID-19 in France. *Frontiers in Medicine*, 7, 2020. ISSN 2296-858X. URL <https://www.frontiersin.org/articles/10.3389/fmed.2020.00274>. → pp. 102 et 103
- Roshan, A., Murai, K., Fowler, J., Simons, B. D., Nikolaidou-Neokosmidou, V., and Jones, P. H. Human keratinocytes have two interconvertible modes of proliferation. *Nature Cell Biology*, 18, 2 2016. ISSN 1465-7392. doi : 10.1038/ncb3282. Note available

. → pp. 87, 95, 97 et 99
- Rositch, A. F., Burke, A. E., Viscidi, R. P., Silver, M. I., Chang, K., and Gravitt, P. E. Contributions of Recent and Past Sexual Partnerships on Incident Human Papillomavirus Detection : Acquisition and Reactivation

- in Older Women. *Cancer Research*, 72(23) :6183–6190, November 2012. ISSN 0008-5472. doi : 10.1158/0008-5472.CAN-12-2635. URL <https://doi.org/10.1158/0008-5472.CAN-12-2635>. → p. 64
- Ross, R. An application of the theory of probabilities to the study of a priori pathometry.—Part I. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 92(638) : 204–230, January 1997. doi : 10.1098/rspa.1916.0007. URL <https://royalsocietypublishing.org/doi/10.1098/rspa.1916.0007>. Publisher : Royal Society. → p. 20
- Ross, R. and Hudson, H. P. An application of the theory of probabilities to the study of a priori pathometry.—Part II. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 93(650) :212–225, January 1997a. doi : 10.1098/rspa.1917.0014. URL <https://royalsocietypublishing.org/doi/10.1098/rspa.1917.0014>. Publisher : Royal Society. → p. 20
- Ross, R. and Hudson, H. P. An application of the theory of probabilities to the study of a priori pathometry.—Part III. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 93(650) :225–240, January 1997b. doi : 10.1098/rspa.1917.0015. URL <https://royalsocietypublishing.org/doi/10.1098/rspa.1917.0015>. Publisher : Royal Society. → p. 20
- Roteli-Martins, C. M., Naud, P., De Borba, P., Teixeira, J. C., De Carvalho, N. S., Zahaf, T., Sanchez, N., Geeraerts, B., and Descamps, D. Sustained immunogenicity and efficacy of the HPV-16/18 AS04-adjuvanted vaccine : up to 8.4 years of follow-up. *Human Vaccines & Immunotherapeutics*, 8(3) : 390–397, March 2012. ISSN 2164-554X. doi : 10.4161/hv.18865. → p. 67
- Rubinstein, R. Y. and Kroese, D. P. *Simulation and the Monte Carlo method*. Wiley series in probability and statistics. Wiley, Hoboken, New Jersey, third edition edition, 2017. ISBN 978-1-118-63216-1. → p. 33
- Ryser, M. D., Myers, E. R., and Durrett, R. Hpv clearance and the neglected role of stochasticity. *PLoS Comput. Biol.*, 11 :e1004113, 3 2015. ISSN 1553-7358. doi : 10.1371/journal.pcbi.1004113. → pp. 83, 84, 87, 90, 96, 98, 99, 100 et 192
- Ryser, M. D., Gravitt, P. E., and Myers, E. R. Mechanistic mathematical models : An underused platform for hpv research. *Papillomavirus Research*, 3 :46–49, 6 2017. ISSN 2405-8521. doi : 10.1016/j.pvr.2017.01.004. URL <http://www.sciencedirect.com/science/article/pii/S2405852116300623>. → pp. 83, 87, 122 et 125
- Salje, H., Tran Kiem, C., Lefrancq, N., Courtejoie, N., Bosetti, P., Paireau, J., Andronico, A., Hozé, N., Richet, J., Dubost, C.-L., Le Strat, Y., Lessler, J., Levy-Bruhl, D., Fontanet, A., Opatowski, L., Boelle, P.-Y., and Cauchemez, S. Estimating the burden of SARS-CoV-2 in France. *Science (New York, N.Y.)*, 369(6500) :208–211, July 2020. ISSN 1095-9203. doi : 10.1126/science.abc3517. → pp. 102 et 103

- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. *Global Sensitivity Analysis. The Primer*. John Wiley & Sons, Ltd, 12 2007. ISBN 9780470725184. doi : 10.1002/9780470725184. → pp. 21 et 92
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, 181, 2 2010. ISSN 00104655. doi : 10.1016/j.cpc.2009.09.018. → pp. 21 et 92
- Sanjose, S. d., Quint, W. G., Alemany, L., Geraets, D. T., Klaustermeier, J. E., Lloveras, B., Tous, S., Felix, A., Bravo, L. E., Shin, H.-R., Vallejos, C. S., Ruiz, P. A. d., Lima, M. A., Guimera, N., Clavero, O., Alejo, M., Llombart-Bosch, A., Cheng-Yang, C., Tatti, S. A., Kasamatsu, E., Iljazovic, E., Odida, M., Prado, R., Seoud, M., Grce, M., Usubutun, A., Jain, A., Suarez, G. A. H., Lombardi, L. E., Banjo, A., Menéndez, C., Domingo, E. J., Velasco, J., Nessa, A., Chichareon, S. C. B., Qiao, Y. L., Lerma, E., Garland, S. M., Sasagawa, T., Ferrera, A., Hammouda, D., Mariani, L., Pelayo, A., Steiner, I., Oliva, E., Meijer, C. J., Al-Jassar, W. F., Cruz, E., Wright, T. C., Puras, A., Llave, C. L., Tzardi, M., Agorastos, T., Garcia-Barriola, V., Clavel, C., Ordi, J., Andújar, M., Castellsagué, X., Sánchez, G. I., Nowakowski, A. M., Bornstein, J., Muñoz, N., and Bosch, F. X. Human papillomavirus genotype attribution in invasive cervical cancer : a retrospective cross-sectional worldwide study. *The Lancet Oncology*, 11(11) :1048–1056, November 2010. ISSN 1470-2045, 1474-5488. doi : 10.1016/S1470-2045(10)70230-8. URL [https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045\(10\)70230-8/fulltext](https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(10)70230-8/fulltext). Publisher : Elsevier. → p. 16
- Schiffman, M. and Solomon, D. Cervical-cancer screening with human papillomavirus and cytologic cotesting. *New England Journal of Medicine*, 369 : 2324–2331, 12 2013. ISSN 0028-4793. doi : 10.1056/NEJMcp1210379. URL <http://www.nejm.org/doi/10.1056/NEJMcp1210379>. → pp. 15, 41, 48, 86 et 126
- Schiffman, M., Herrero, R., DeSalle, R., Hildesheim, A., Wacholder, S., Rodriguez, A. C., Bratti, M. C., Sherman, M. E., Morales, J., Guillen, D., Alfaro, M., Hutchinson, M., Wright, T. C., Solomon, D., Chen, Z., Schussler, J., Castle, P. E., and Burk, R. D. The carcinogenicity of human papillomavirus types reflects viral evolution. *Virology*, 337 :76–84, 6 2005. ISSN 00426822. doi : 10.1016/j.virol.2005.04.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S0042682205002096>. → pp. 16 et 98
- Schiffman, M., Wentzensen, N., Wacholder, S., Kinney, W., Gage, J. C., and Castle, P. E. Human papillomavirus testing in the prevention of cervical cancer. *JNCI : Journal of the National Cancer Institute*, 103 :368–383, 3 2011a. ISSN 1460-2105. doi : 10.1093/jnci/djq562. URL <https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djq562>. → pp. 15, 86 et 125

- Schiffman, M., Wentzensen, N., Wacholder, S., Kinney, W., Gage, J. C., and Castle, P. E. Human Papillomavirus Testing in the Prevention of Cervical Cancer. *J Natl Cancer Inst*, 103(5) :368–383, March 2011b. ISSN 0027-8874. doi : 10.1093/jnci/djq562. URL <https://academic.oup.com/jnci/article/103/5/368/905734>. Publisher : Oxford Academic. → p. 52
- Schiller, J. T., Day, P. M., and Kines, R. C. Current understanding of the mechanism of hpv infection. *Gynecologic Oncology*, 118 :S12–S17, 6 2010. ISSN 00908258. doi : 10.1016/j.ygyno.2010.04.004. URL <https://linkinghub.elsevier.com/retrieve/pii/S0090825810002908>. → p. 99
- Schoenfeld, D. Partial Residuals for The Proportional Hazards Regression Model. *Biometrika*, 69(1) :239, 1982. doi : 10.2307/2335876. → pp. 40 et 55
- Scott, M., Nakagawa, M., and Moscicki, A.-B. Cell-mediated immune response to human papillomavirus infection. *Clinical Diagnostic Laboratory Immunology*, 8, 3 2001. ISSN 1071-412X. doi : 10.1128/CDLI.8.2.209-220.2001. → p. 86
- Serrano, B., Brotons, M., Bosch, F. X., and Bruni, L. Epidemiology and burden of HPV-related disease. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 47 :14–26, February 2018. ISSN 1521-6934. doi : 10.1016/j.bpobgyn.2017.08.006. URL <https://www.sciencedirect.com/science/article/pii/S1521693417301244>. → p. 15
- Servedio, M. R., Brandvain, Y., Dhole, S., Fitzpatrick, C. L., Goldberg, E. E., Stern, C. A., Cleve, J. V., and Yeh, D. J. Not Just a Theory—The Utility of Mathematical Models in Evolutionary Biology. *PLOS Biology*, 12(12) : e1002017, December 2014. ISSN 1545-7885. doi : 10.1371/journal.pbio.1002017. URL <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002017>. Publisher : Public Library of Science. → pp. 80, 84 et 126
- Shew, M. L., Ermel, A. C., Tong, Y., Tu, W., Qadadri, B., and Brown, D. R. Episodic detection of human papillomavirus within a longitudinal cohort of young women. *Journal of Medical Virology*, 87(12), December 2015. ISSN 01466615. doi : 10.1002/jmv.24284. → pp. 64 et 69
- Shorter, E. Ignaz Semmelweis : The etiology, concept, and prophylaxis of childbed fever. *Medical History*, 28(3) :334, July 1984. ISSN 0025-7273. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1139454/>. → p. 8
- Simon, A. K., Hollander, G. A., and McMichael, A. Evolution of the immune system in humans from infancy to old age. *Proceedings of the Royal Society B : Biological Sciences*, 282(1821) :20143085, December 2015. ISSN 0962-8452. doi : 10.1098/rspb.2014.3085. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4707740/>. → p. 10
- Slack, E., Hapfelmeier, S., Stecher, B., Velykoredko, Y., Stoel, M., Lawson, M. A. E., Geuking, M. B., Beutler, B., Tedder, T. F., Hardt, W.-D., Bercik, P., Verdu, E. F., McCoy, K. D., and Macpherson, A. J. Innate and adaptive

- immunity cooperate flexibly to maintain host-microbiota mutualism. *Science (New York, N.Y.)*, 325(5940) :617–620, July 2009. ISSN 1095-9203. doi : 10.1126/science.1172747. → p. 10
- Smeach, S. C. and Gold, H. J. Stochastic and deterministic models for the kinetic behavior of certain structured enzyme systems II : Consecutive two enzyme systems. *Journal of Theoretical Biology*, 51(1) :79–96, May 1975. ISSN 0022-5193. doi : 10.1016/0022-5193(75)90140-X. URL <https://www.sciencedirect.com/science/article/pii/002251937590140X>. → pp. 21 et 32
- Snow, J. On the Mode of Communication of Cholera. *Edinburgh Medical Journal*, 1(7) :668–670, January 1856. ISSN 0367-1038. → p. 8
- Snow, J. On Continuous Molecular Changes, More Particularly in Their Relation to Epidemic Diseases. *Reviews of Infectious Diseases*, 7(3) :441–447, 1985. ISSN 0162-0886. URL <https://www.jstor.org/stable/4453638>. Publisher : Oxford University Press. → p. 8
- Sobol, I. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, 55, 2 2001. ISSN 03784754. doi : 10.1016/S0378-4754(00)00270-6. → pp. 21 et 92
- Sofonea, M. T., Reyné, B., Elie, B., Djidjou-Demasse, R., Selinger, C., Michalakis, Y., and Alizon, S. Epidemiological monitoring and control perspectives : application of a parsimonious modelling framework to the COVID-19 dynamics in France. *medRxiv*, page 2020.05.22.20110593, 2020. doi : 10.1101/2020.05.22.20110593. → pp. 103, 107, 109, 110, 111, 112, 113, 118, 200 et 201
- Sofonea, M. T., Reyné, B., Elie, B., Djidjou-Demasse, R., Selinger, C., Michalakis, Y., and Alizon, S. Memory is key in capturing COVID-19 epidemiological dynamics. *Epidemics*, 35 :100459, June 2021. ISSN 1755-4365. doi : 10.1016/j.epidem.2021.100459. URL <https://www.sciencedirect.com/science/article/pii/S1755436521000189>. → pp. 34, 102, 103 et 127
- Song, P., Li, W., Xie, J., Hou, Y., and You, C. Cytokine storm induced by SARS-CoV-2. *Clinica Chimica Acta; International Journal of Clinical Chemistry*, 509 :280–287, October 2020. ISSN 0009-8981. doi : 10.1016/j.cca.2020.06.017. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7283076/>. → p. 10
- Soper, B. C., Nygård, M., Abdulla, G., Meng, R., and Nygård, J. F. A hidden Markov model for population-level cervical cancer screening data. *Statistics in Medicine*, 39(25) :3569–3590, 2020. ISSN 1097-0258. doi : 10.1002/sim.8681. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8681>. eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.8681>. → p. 65
- Stanley, M. A. Epithelial cell responses to infection with human papillomavirus. *Clinical Microbiology Reviews*, 25 :215–222, 4 2012. ISSN 0893-8512. doi :

- 10.1128/CMR.05028-11. URL <https://cmr.asm.org/content/25/2/215>.
→ pp. 88 et 89
- Stanley, M. Immune responses to human papillomavirus. *Vaccine*, 24(S1) : S16–22, 2006. doi : 10.1016/j.vaccine.2005.09.002. → pp. 84 et 88
- Stanley, M. Immunobiology of hpv and hpv vaccines. *Gynecologic Oncology*, 109 :S15–S21, 5 2008. ISSN 00908258. doi : 10.1016/j.ygyno.2008.02.003. URL <https://linkinghub.elsevier.com/retrieve/pii/S0090825808001054>. → pp. 86 et 98
- Stanley, M. Pathology and epidemiology of hpv infection in females. *Gynecologic Oncology*, 117 :S5–S10, 5 2010. ISSN 00908258. doi : 10.1016/j.ygyno.2010.01.024. URL <https://linkinghub.elsevier.com/retrieve/pii/S0090825810000879>. → pp. 8 et 88
- Stanley, M. Host defence and persistent human papillomavirus infection. *Current Opinion in Virology*, 51 :106–110, December 2021. ISSN 1879-6265. doi : 10.1016/j.coviro.2021.09.010. → p. 84
- Steele, M. K., Couture, A., Reed, C., Iuliano, D., Whitaker, M., Fast, H., Hall, A. J., MacNeil, A., Cadwell, B., Marks, K. J., and Silk, B. J. Estimated Number of COVID-19 Infections, Hospitalizations, and Deaths Prevented Among Vaccinated Persons in the US, December 2020 to September 2021. *JAMA Network Open*, 5(7) :e2220385, July 2022. ISSN 2574-3805. doi : 10.1001/jamanetworkopen.2022.20385. URL <https://doi.org/10.1001/jamanetworkopen.2022.20385>. → p. 19
- Strickler, H. D., Burk, R. D., Fazzari, M., Anastos, K., Minkoff, H., Massad, L. S., Hall, C., Bacon, M., Levine, A. M., Watts, D. H., Silverberg, M. J., Xue, X., Schlecht, N. F., Melnick, S., and Palefsky, J. M. Natural history and possible reactivation of human papillomavirus in human immunodeficiency virus-positive women. *Journal of the National Cancer Institute*, 97(8) :577–586, April 2005. ISSN 1460-2105. doi : 10.1093/jnci/dji073. → p. 64
- Stubenrauch, F. and Laimins, L. A. Human papillomavirus life cycle : active and latent phases. *Seminars in Cancer Biology*, 9(6) :379–386, December 1999. ISSN 1044-579X. doi : 10.1006/scbi.1999.0141. URL <https://www.sciencedirect.com/science/article/pii/S1044579X99901414>. → p. 14
- Sun, K., Wang, W., Gao, L., Wang, Y., Luo, K., Ren, L., Zhan, Z., Chen, X., Zhao, S., Huang, Y., Sun, Q., Liu, Z., Litvinova, M., Vespignani, A., Ajelli, M., Viboud, C., and Yu, H. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science*, November 2020. ISSN 0036-8075, 1095-9203. doi : 10.1126/science.abe2424. URL <https://science.sciencemag.org/content/early/2020/11/23/science.abe2424>. Publisher : American Association for the Advancement of Science Section : Research Article. → pp. 103, 109 et 120
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. Global Cancer Statistics 2020 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA : A Cancer Journal for Clinicians*, 71(3) :209–249, 2021. ISSN

- 1542-4863. doi : 10.3322/caac.21660. URL <https://onlinelibrary.wiley.com/doi/abs/10.3322/caac.21660>. _eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21660>. → p. 64
- Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology*, 5(10) :801–812, October 2007. ISSN 1740-1534. doi : 10.1038/nrmicro1750. URL <https://www.nature.com/articles/nrmicro1750>. Number : 10 Publisher : Nature Publishing Group. → p. 3
- Symonds, M. R. E. and Moussalli, A. A brief guide to model selection, multi-model inference and model averaging in behavioural ecology using Akaike’s information criterion. *Behavioral Ecology and Sociobiology*, 65(1) :13–21, 2011. doi : 10.1007/s00265-010-1037-6. → p. 179
- Teitelman, A. M., Gregory, E. F., Jayasinghe, J., Wermers, Z., Koo, J. H., Morone, J. F., Leri, D. C., Davis, A., and Feemster, K. A. Vaccipack, A Mobile App to Promote Human Papillomavirus Vaccine Uptake Among Adolescents Aged 11 to 14 Years : Development and Usability Study. *JMIR Nursing*, 3(1) :e19503, October 2020. ISSN 2562-7600. doi : 10.2196/19503. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8279454/>. → p. 124
- Tellier, R. COVID-19 : the case for aerosol transmission. *Interface Focus*, 12(2) :20210072, February 2022. doi : 10.1098/rsfs.2021.0072. URL <https://royalsocietypublishing.org/doi/full/10.1098/rsfs.2021.0072>. Publisher : Royal Society. → pp. 7 et 17
- Theiler, R. N., Farr, S. L., Karon, J. M., Paramsothy, P., Viscidi, R., Duerr, A., Cu-Uvin, S., Sobel, J., Shah, K., Klein, R. S., and Jamieson, D. J. High-risk human papillomavirus reactivation in human immunodeficiency virus-infected women : risk factors for cervical viral shedding. *Obstetrics and Gynecology*, 115(6) :1150–1158, June 2010. ISSN 1873-233X. doi : 10.1097/AOG.0b013e3181e00927. → p. 64
- Therneau, T. M. *coxme : Mixed Effects Cox Models*, 2022a. URL <https://CRAN.R-project.org/package=coxme>. R package version 2.2-17. → p. 182
- Therneau, T. M. *A Package for Survival Analysis in R*, 2022b. URL <https://CRAN.R-project.org/package=survival>. R package version 3.4-0. → p. 182
- Thompson, E. L., Vamos, C. A., Sappenfield, W. M., Straub, D. M., and Daley, E. M. Relationship status impacts primary reasons for interest in the HPV vaccine among young adult women. *Vaccine*, 34(27) :3119–3124, June 2016. ISSN 0264-410X. doi : 10.1016/j.vaccine.2016.04.063. URL <https://www.sciencedirect.com/science/article/pii/S0264410X16302286>. → p. 62
- Thompson, R. N., Morgan, O. W., and Jalava, K. Rigorous surveillance is necessary for high confidence in end-of-outbreak declarations for Ebola and other infectious diseases. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 374(1776) :20180431, 2019. doi : 10.1098/rstb.2018.0431. → p. 106

- Tomaić, V. Functional roles of e6 and e7 oncoproteins in hpv-induced malignancies at diverse anatomical sites. *Cancers*, 8, 2016. ISSN 20726694. doi : 10.3390/cancers8100095. → pp. 13, 84 et 98
- Trapman, P., Ball, F., Dhersin, J.-S., Tran, V. C., Wallinga, J., and Britton, T. Inferring R0 in emerging epidemics—the effect of common population structure is small. *J. R. Soc. Interface*, 13(121) :20160288, 2016. doi : 10.1098/rsif.2016.0288. → p. 108
- Trottier, H., Ferreira, S., Thomann, P., Costa, M. C., Sobrinho, J. S., Prado, J. C. M., Rohan, T. E., Villa, L. L., and Franco, E. L. Human papillomavirus infection and reinfection in adult women : the role of sexual activity and natural immunity. *Cancer Research*, 70(21) :8569–8577, November 2010. ISSN 1538-7445. doi : 10.1158/0008-5472.CAN-10-0621. → pp. 61, 64 et 122
- Trépo, C., Chan, H. L. Y., and Lok, A. Hepatitis B virus infection. *The Lancet*, 384(9959) :2053–2063, December 2014. ISSN 0140-6736. doi : 10.1016/S0140-6736(14)60220-8. URL <https://www.sciencedirect.com/science/article/pii/S0140673614602208>. → p. 126
- Twort, F. W. AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES. *The Lancet*, 186(4814) :1241–1243, December 1915. ISSN 0140-6736, 1474-547X. doi : 10.1016/S0140-6736(01)20383-3. URL [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(01\)20383-3/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(01)20383-3/fulltext). Publisher : Elsevier. → p. 3
- Valen, L. V. 19. A New Evolutionary Law (1973). In *19. A New Evolutionary Law (1973)*, pages 284–314. University of Chicago Press, August 2014. ISBN 978-0-226-11550-4. doi : 10.7208/9780226115504-022. URL <https://www.degruyter.com/document/doi/10.7208/9780226115504-022/html>. → p. 12
- Vallée, J., Cadot, E., Grillo, F., Parizot, I., and Chauvin, P. The combined effects of activity space and neighbourhood of residence on participation in preventive health-care activities : The case of cervical screening in the Paris metropolitan area (France). *Health & Place*, 16(5) :838–852, September 2010. ISSN 1353-8292. doi : 10.1016/j.healthplace.2010.04.009. URL <https://www.sciencedirect.com/science/article/pii/S1353829210000456>. → p. 61
- van Baalen, M. Coevolution of recovery ability and virulence. *Proceedings. Biological Sciences*, 265(1393) :317–325, February 1998. ISSN 0962-8452. doi : 10.1098/rspb.1998.0298. → p. 12
- van Doorn, L.-J., Quint, W., Kleter, B., Molijn, A., Colau, B., Martin, M.-T., Kravang-In, n., Torrez-Martinez, N., Peyton, C. L., and Wheeler, C. M. Genotyping of human papillomavirus in liquid cytology cervical specimens by the PGMY line blot assay and the SPF(10) line probe assay. *Journal of Clinical Microbiology*, 40(3) :979–983, March 2002. ISSN 0095-1137. doi : 10.1128/JCM.40.3.979-983.2002. → pp. 48 et 68

- van Doorn, L.-J., Molijn, A., Kleter, B., Quint, W., and Colau, B. Highly effective detection of human papillomavirus 16 and 18 DNA by a testing algorithm combining broad-spectrum and type-specific PCR. *Journal of Clinical Microbiology*, 44(9) :3292–3298, September 2006. ISSN 0095-1137. doi : 10.1128/JCM.00539-06. → p. 68
- Van Doorslaer, K. Evolution of the Papillomaviridae. *Virology*, 445(1) :11–20, October 2013. ISSN 0042-6822. doi : 10.1016/j.virol.2013.05.012. URL <https://www.sciencedirect.com/science/article/pii/S0042682213002924>. → p. 12
- Van Hede, D., Langers, I., Delvenne, P., and Jacobs, N. Origin and immunoescape of uterine cervical cancer. *La Presse Médicale*, 43 :e413–e421, 12 2014. ISSN 07554982. doi : 10.1016/j.lpm.2014.09.005. URL <https://linkinghub.elsevier.com/retrieve/pii/S0755498214004709>. → p. 86
- Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P. G. T., Fu, H., Dighe, A., Griffin, J. T., Baguelin, M., Bhatia, S., Boonyasiri, A., Cori, A., Cucunubá, Z., FitzJohn, R., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Laydon, D., Nedjati-Gilani, G., Riley, S., Elsland, S. v., Volz, E., Wang, H., Wang, Y., Xi, X., Donnelly, C. A., Ghani, A. C., and Ferguson, N. M. Estimates of the severity of coronavirus disease 2019 : a model-based analysis. *The Lancet Infectious Diseases*, 0(0), March 2020. ISSN 1473-3099, 1474-4457. doi : 10.1016/S1473-3099(20)30243-7. URL [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30243-7/abstract](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30243-7/abstract). → pp. 18 et 109
- Vermeil, T., Peters, A., Kilpatrick, C., Pires, D., Allegranzi, B., and Pittet, D. Hand hygiene in hospitals : anatomy of a revolution. *The Journal of Hospital Infection*, 101(4) :383–392, April 2019. ISSN 1532-2939. doi : 10.1016/j.jhin.2018.09.003. → p. 8
- Virgin, H. W., Wherry, E. J., and Ahmed, R. Redefining Chronic Viral Infection. *Cell*, 138(1) :30–50, July 2009. ISSN 0092-8674, 1097-4172. doi : 10.1016/j.cell.2009.06.036. URL [https://www.cell.com/cell/abstract/S0092-8674\(09\)00783-1](https://www.cell.com/cell/abstract/S0092-8674(09)00783-1). Publisher : Elsevier. → p. 11
- Viterbi, A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13 (2) :260–269, April 1967. ISSN 0018-9448, 1557-9654. doi : 10.1109/TIT.1967.1054010. URL <http://ieeexplore.ieee.org/document/1054010/>. → p. 70
- Wallinga, J. and Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B : Biological Sciences*, 274(1609) :599–604, November 2006. doi : 10.1098/rspb.2006.3754. URL <https://royalsocietypublishing.org/doi/10.1098/rspb.2006.3754>. Publisher : Royal Society. → p. 103

- Wan, L., Lou, W., Abner, E., and Kryscio, R. J. A comparison of time-homogeneous Markov chain and Markov process multi-state models. *Communications in statistics. Case studies, data analysis and applications*, 2 (3-4) :92–100, 2016. ISSN 2373-7484. doi : 10.1080/23737484.2017.1361366. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5803756/>. → p. 42
- Wang, C. C., Prather, K. A., Sznitman, J., Jimenez, J. L., Lakdawala, S. S., Tufekci, Z., and Marr, L. C. Airborne transmission of respiratory viruses. *Science*, 373(6558) :eabd9149, August 2021. doi : 10.1126/science.abd9149. URL <https://www.science.org/doi/full/10.1126/science.abd9149>. Publisher : American Association for the Advancement of Science. → pp. 7 et 8
- Watson, H. W. and Galton, F. On the Probability of the Extinction of Families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4 :138–144, 1875. ISSN 0959-5295. doi : 10.2307/2841222. URL <https://www.jstor.org/stable/2841222>. Publisher : [Royal Anthropological Institute of Great Britain and Ireland, Wiley]. → p. 28
- Wienke, A. *Frailty Models in Survival Analysis*. Chapman and Hall/CRC, New York, September 2010. ISBN 978-0-429-13960-4. doi : 10.1201/9781420073911. → pp. 49 et 55
- Wilailak, S., Kengsakul, M., and Kehoe, S. Worldwide initiatives to eliminate cervical cancer. *International Journal of Gynaecology and Obstetrics*, 155 (Suppl 1) :102–106, October 2021. ISSN 0020-7292. doi : 10.1002/ijgo.13879. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9298014/>. → pp. 16, 122 et 123
- Winer, R. L., Hughes, J. P., Feng, Q., O'Reilly, S., Kiviat, N. B., Holmes, K. K., and Koutsky, L. A. Condom Use and the Risk of Genital Human Papillomavirus Infection in Young Women. *New England Journal of Medicine*, 354 (25) :2645–2654, June 2006. ISSN 0028-4793. doi : 10.1056/NEJMoa053284. URL <https://doi.org/10.1056/NEJMoa053284>. Publisher : Massachusetts Medical Society _eprint : <https://doi.org/10.1056/NEJMoa053284>. → pp. 8, 16 et 124
- Winer, R., Feng, Q., Hughes, J., O'Reilly, S., Kiviat, N., and Koutsky, L. Risk of Female Human Papillomavirus Acquisition Associated with First Male Sex Partner. *The Journal of Infectious Diseases*, 197(2) :279–282, January 2008. ISSN 0022-1899, 1537-6613. doi : 10.1086/524875. URL <https://academic.oup.com/jid/article-lookup/doi/10.1086/524875>. → p. 8
- Woodall, W. G., Zimet, G., Kong, A., Buller, D., Reither, J., Chilton, L., Myers, V., and Starling, R. Vacteens.org : A Mobile Web app to Improve HPV Vaccine Uptake. *Frontiers in Digital Health*, 3 :693688, August 2021. ISSN 2673-253X. doi : 10.3389/fdgth.2021.693688. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8521965/>. → p. 124

- Woolhouse, M., Scott, F., Hudson, Z., Howey, R., and Chase-Topping, M. Human viruses : discovery and emergence. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 367(1604) :2864–2871, October 2012. ISSN 0962-8436. doi : 10.1098/rstb.2011.0354. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3427559/>. → p. 3
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., and Zhang, Y.-Z. A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798) :265–269, March 2020. ISSN 1476-4687. doi : 10.1038/s41586-020-2008-3. URL <https://www.nature.com/articles/s41586-020-2008-3>. Number : 7798 Publisher : Nature Publishing Group. → p. 18
- Yang, H. and Rao, Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nature Reviews Microbiology*, 19(11) :685–700, November 2021. ISSN 1740-1534. doi : 10.1038/s41579-021-00630-8. URL <https://www.nature.com/articles/s41579-021-00630-8>. Number : 11 Publisher : Nature Publishing Group. → pp. 17 et 18
- Yates, C. A., Ford, M. J., and Mort, R. L. A Multi-stage Representation of Cell Proliferation as a Markov Process. *Bulletin of Mathematical Biology*, 79(12) :2905–2928, 2017. ISSN 0092-8240. doi : 10.1007/s11538-017-0356-4. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5709504/>. → p. 33
- Yoon, B.-J. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Current Genomics*, 10(6) :402–415, September 2009. doi : 10.2174/138920209789177575. → p. 28
- Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E., and Fouchier, R. A. M. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *The New England Journal of Medicine*, 367(19) :1814–1820, November 2012. ISSN 1533-4406. doi : 10.1056/NEJMoa1211721. → p. 18
- Zeng, B., Gao, L., Zhou, Q., Yu, K., and Sun, F. Effectiveness of COVID-19 vaccines against SARS-CoV-2 variants of concern : a systematic review and meta-analysis. *BMC medicine*, 20(1) :200, May 2022. ISSN 1741-7015. doi : 10.1186/s12916-022-02397-y. → pp. 7 et 19
- Zhang, P., Nouri, M., Brandsma, J., Iftner, T., and Steinberg, B. Induction of E6/E7 expression in cottontail rabbit papillomavirus latency following UV activation. *Virology*, 263(2), October 1999. ISSN 0042-6822. doi : 10.1006/viro.1999.9950. URL <https://pubmed.ncbi.nlm.nih.gov/10544111/>. Publisher : Virology. → pp. 64 et 125
- Zhang, T., Wu, Q., and Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Current biology : CB*, 30(7) : 1346–1351.e2, April 2020. ISSN 1879-0445. doi : 10.1016/j.cub.2020.03.022. → p. 18

- Zhang, Z. and Sun, J. Interval censoring. *Statistical methods in medical research*, 19(1) :53–70, February 2010. ISSN 0962-2802. doi : 10.1177/0962280209105023. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3684949/>. → p. 34
- Zhong, N. S., Zheng, B. J., Li, Y. M., Poon, n., Xie, Z. H., Chan, K. H., Li, P. H., Tan, S. Y., Chang, Q., Xie, J. P., Liu, X. Q., Xu, J., Li, D. X., Yuen, K. Y., Peiris, n., and Guan, Y. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People’s Republic of China, in February, 2003. *Lancet (London, England)*, 362(9393) :1353–1358, October 2003. ISSN 1474-547X. doi : 10.1016/s0140-6736(03)14630-2. → p. 18
- Zhou, M. *Empirical Likelihood Method in Survival Analysis*. Chapman and Hall/CRC, New York, June 2015. ISBN 978-0-429-09694-5. doi : 10.1201/b18598. → p. 38
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., and Shi, Z.-L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798) :270–273, March 2020. ISSN 1476-4687. doi : 10.1038/s41586-020-2012-7. URL <https://www.nature.com/articles/s41586-020-2012-7>. Number : 7798 Publisher : Nature Publishing Group. → p. 18
- zur Hausen, H. Papillomavirus infections—a major cause of human cancers. *Biochimica Et Biophysica Acta*, 1288(2) :F55–78, October 1996. ISSN 0006-3002. doi : 10.1016/0304-419x(96)00020-0. → pp. 15, 64 et 125

ANNEXE

A

Annexes et suppléments d'articles

A.1 Infections transitoires HPV

HPV detection patterns in young women from the PAPCLEAR longitudinal study : implications for HPV screening policies.

Ethics

The PAPCLEAR trial was promoted by the Centre Hospitalier Universitaire de Montpellier and approved by the *Comité de Protection des Personnes (CPP) Sud Méditerranée I* on 11 May 2016 (CPP number 16 42, reference number ID RCB 2016A00712-49); by the *Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé* on 12 July 2016 (reference number 16.504); and by the *Commission Nationale Informatique et Libertés* on 16 December 2016 (reference number MMS/ABD/AR1612278, decision number DR-2016488). This trial was authorised by the *Agence Nationale de Sécurité du Médicament et des Produits de Santé* on 20 July 2016 (reference 20160072000007). The ClinicalTrials.gov identifier is NCT02946346. All participants provided written informed consent.

Protocol of PAPCLEAR study

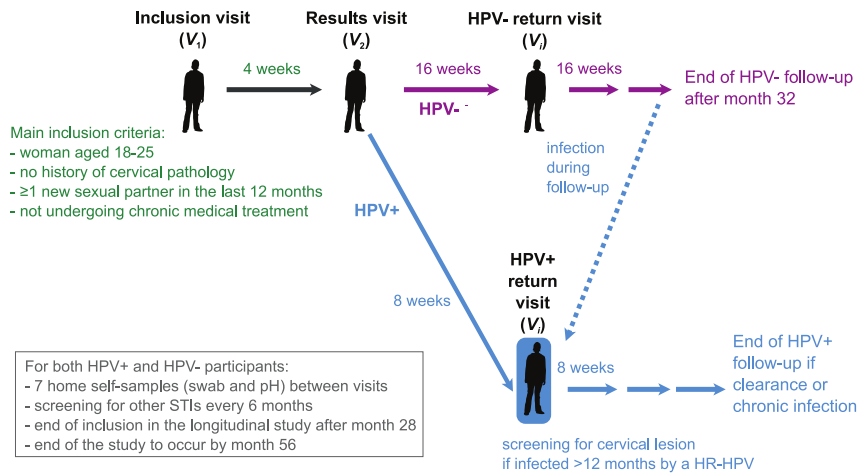


FIGURE A.1 – General structure of the PAPCLEAR study. [Murall et al., 2019b]

Defining events and durations

Results from the DEIA and LiPA25 assays yielded dated binary vectors. For each infectious event, we only know the intervals during which the infection started and ended, which means the data is ‘doubly interval censored’ and usually cumbersome to analyse [Bogaerts et al., 2017]. To simplify the problem, we computed duration using the conventional midpoint methodology. For this, we defined the start of an infection as the midpoint between the last negative test before and the first positive test of the infection. Likewise, we defined the end of an infection as the midpoint between the last positive test of the infection and the first negative test after the infection). For incomplete data, we assume the start to be at inclusion for left-censored observation and we assumed the end to be at last visit for right-censored observation. The bias associated with this simpler method is expected to be limited since our sampling scheme is regular and short-spaced [Law and Brookmeyer, 1992].

To study the time to HPV infection clearance, we defined as an ‘event’ or ‘episode’ a series of at least one positive LiPA25 detection for a given HPV type and a given participant. During the follow-up, we often detected several events per participant (sometimes even by the same genotype). We assumed that two consecutive episodes were independent even if only separated by one negative visit. Such patterns, also called intermittent [Malagón et al., 2023], are sometimes merged to form a longer episode instead of two separate entities [Paul et al., 2021].

To estimate the time of HPV detectability, i.e. the time of positive HPV detection, we computed the duration between the midpoint at the start of an infection and the midpoint at the end of the infection. If one or both of the endpoints were censored, we assumed the duration to be right-censored. We assumed the events to be independent [Dillner et al., 2011] and, therefore, defined the time between episodes to be independent

events. For the time to first incident infection, we excluded prevalent infection and computed the time from inclusion to the midpoint at start of first incident detection for a genotype and a participant. If the genotype is not detected during follow-up, we used a right-censored observation whose duration equals the time of follow-up of the participant. When analysing the time between positive episodes, we included all events and computed the time as the duration between the midpoint at start of expired episodes and the midpoint at start of the new episodes. There is in general, a lower number of data of redetection than expected because some participant were still positive for a genotype at end of follow-up, thus preventing us from computing a time of redetection. In both cases, the cumulative distribution functions (CDF) or survival functions were computed using the Nelson-Aalen estimator of the cumulative hazard rate function [Nelson, 1969, 1972; Aalen, 1978].

Model comparison

We compared the models using the corrected Akaike Information Criterion (AICc) as a metric for the penalised goodness of fit [Hurvich and Tsai, 1993]. Briefly, we first generated the maximum model with all the variables chosen for the Cox regression and then performed the model selection by subsetting all possible combinations from the maximum model and evaluating their respective AICc. We kept the models with an AICc smaller than the minimum AICc+2, following standard practice [Burnham and Anderson, 2004]. We then averaged the coefficients of the remaining models using a full averaging procedure to avoid artificial departure from 0. This was necessary because we averaged on all the selected models, not just on the ones with the variable whose coefficient was computed [Symonds and Moussalli, 2011]. Finally, we computed the hazard ratio by taking the exponential of these averaged coefficients.

Merging intermittent patterns

Following previous notations, intermittent patterns corresponds to successive positive HPV detection episodes separated only by one negative visit. Merging intermittent patterns modifies the data used for analysis, diminishing the number of events, making them last longer in average. In total, we detect 33 intermittent patterns. Merging the patterns decreased the number of positive detectable events by the same amount. However, it did not change much the results of the Cox regression. While the degree of significance varies between the two datasets, the same trends were observed between the two cases.

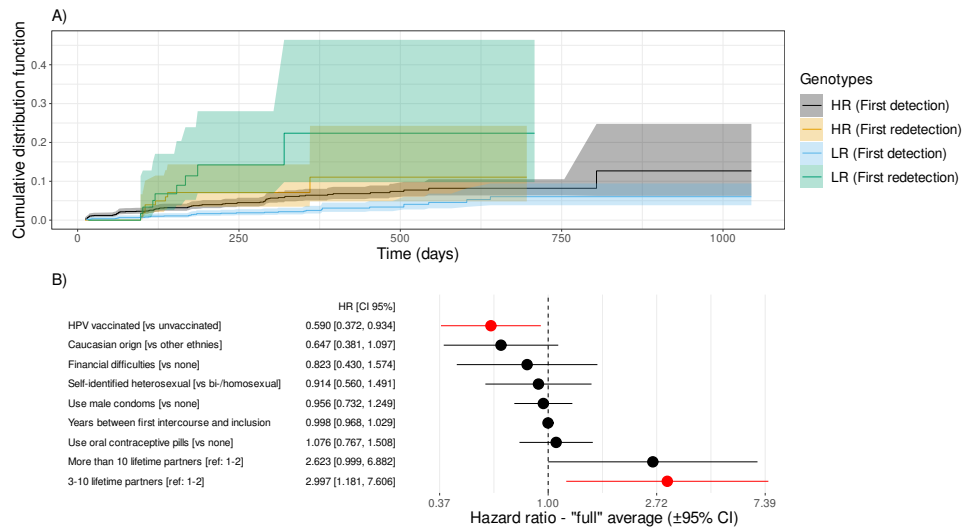


FIGURE A.2 – Cumulative distribution functions for the time to first incident detection and time to first genotype redetection stratified by HR/LR genotypes and effects of host covariates on these estimates for merged intermittent patterns. A) Cumulative distribution function (CDF) of the time to first incident HPV detection since inclusion and the time to the first redetection, stratified by HPV genotype status (HR and LR). B) Hazard ratio for the best models selected by Cox regression with frailty at the patient level. Significant covariates are in red and hazard ratios greater than 1 indicate the covariate is associated with a increased risk of detection, hence lower duration between consecutive episodes.

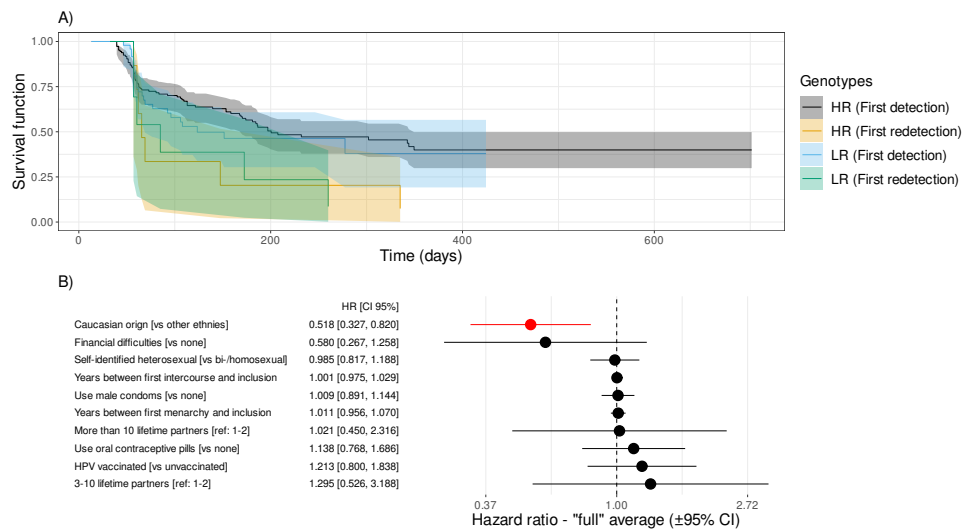


FIGURE A.3 – Survival function for the time of HPV detectability stratified by HR/LR genotypes and effects of host covariates on this estimate for merged intermittent patterns. A) Survival functions stratified by the genotype (HR/LR) for the time to loss of HPV detection. B) Hazard ratios for the host factors. Significant factors are in red and a hazard ratio lower than one indicates that the trait is associated with an decreased rate of loss of HPV DNA detection, hence longer survival functions. The reference level is indicated in the bracket for the qualitative variables (see Methods for details).

R packages

- SURVIVAL : non-parametric and parametric estimators of the survival function and Cox regression [Therneau, 2022b]; version 3.5-3.
- MUMIN : model selection and model averaging [Bartoń, 2022]; version 1.46.0.
- COXME : adding frailty effects to the hazard function as a centred Gaussian distribution [Therneau, 2022a]; version 2.2-17.

Graphical Weibull fit

Let $\lambda > 0$ be the scale parameter and $k > 0$ be the shape parameter, for all $t \in \mathbb{R}_+$, we can define the survival function of the Weibull distribution as :

$$S(t) = e^{-\left(\frac{t}{\lambda}\right)^k} \Leftrightarrow \log(-\log(S(t))) = k(\log(t) - \log(\lambda)) \quad (\text{A.1})$$

Thus using the $\log(-\log(\cdot))$ transformation of the survival function, estimated using non-parametric estimators like Nelson-Aalen or Kaplan-Meier, and plotting it versus the natural logarithm of the event times, we can assess if a Weibull distribution is an appropriate model to describe the data by evaluating the goodness of the fit as a linear model [Dilley et al., 2018]. Clearly, for the time to first incident detection pooled across all genotypes, and for both HR genotypes and LR genotypes grouping, the Weibull was relevant (panel A in Figure A.4). However, for the time to loss of HPV detection, we see a clear non-linear trend between the $\log(-\log(\cdot))$ transformation of the survival function and the $\log(\text{time})$ (panel B in Figure A.4), thus discouraging us for trying to fit a Weibull distributions to this data. The parameter estimates of the Weibull distribution for the time to first incident detection are displayed in Table A.1.

TABLE A.1 – **Estimates for the Weibull parameters (shape and scale) for the time to first incident detection.**

	scale (λ)	shape (k)
All genotypes	1.43 [0.79; 2.58] $\times 10^4$	0.832 [0.713; 0.970]
HR genotypes	1.12 [0.63; 2.01] $\times 10^4$	0.832 [0.714; 0.971]
LR genotypes	1.93 [1.27; 2.94] $\times 10^4$	0.832 [0.714; 0.971]

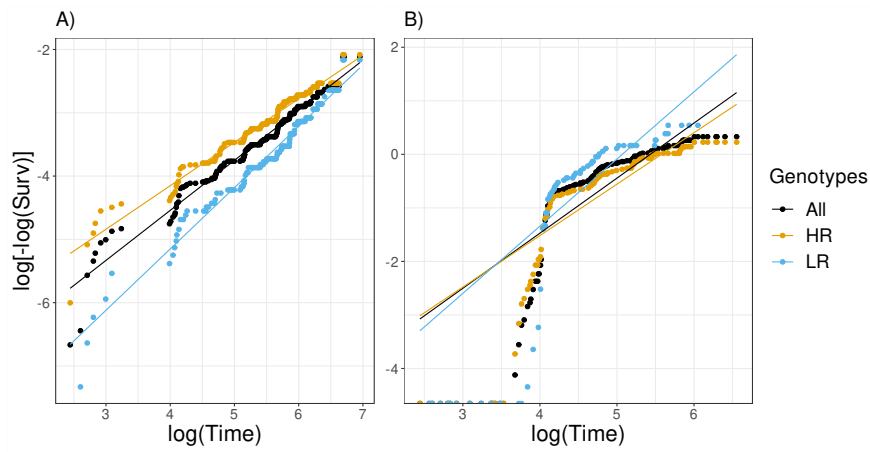


FIGURE A.4 – **Graphical assessment for the goodness of Weibull fit.** On panel A) we displayed the $\log(-\log(\cdot))$ transformation of the survival function for the time to first incident detection versus the $\log(\text{Time})$ and on panel B) the same transformation for the survival functions of the time to loss of HPV detection. For panel A) the linear fit is acceptable while for panel B) there is a clear non-linear trend.

A.2 Modèle de Markov caché pour les infections HPV

Organisation des différentes études cliniques

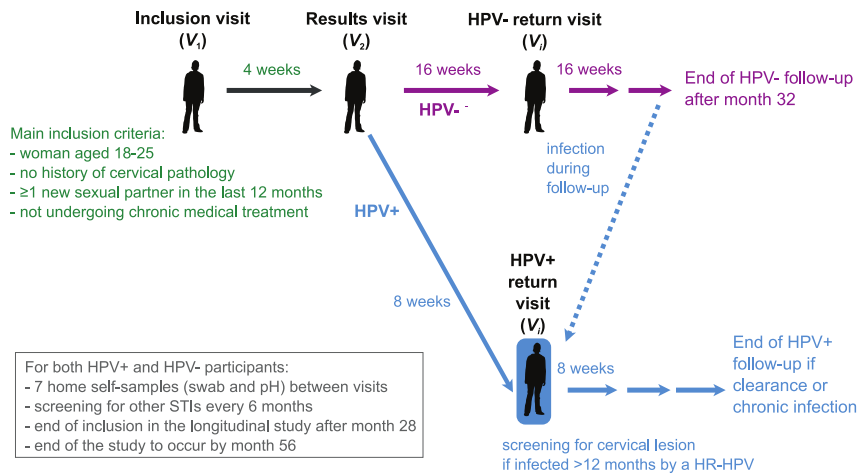


FIGURE A.5 – Organisation de la partie longitudinale de l'étude clinique PASCLEAR (NCT02946346).

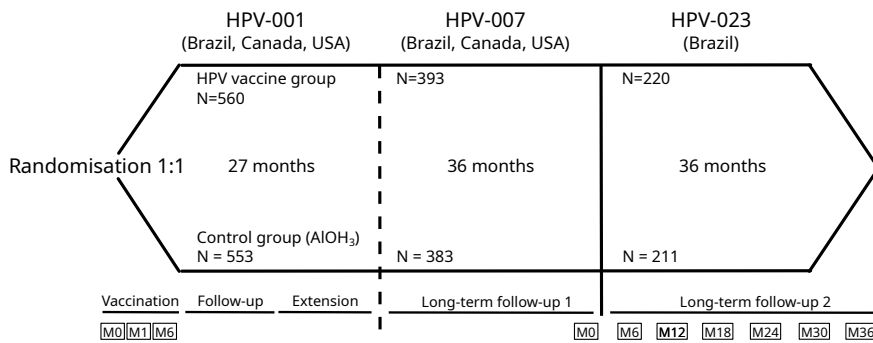


FIGURE A.6 – Organisation des études cliniques HPV-001 (NCT00689741), HPV-007 (NCT00120848) et HPV-023 (NCT00518336). Ces trois études cliniques ont été menées par l'entreprise GSK entre 2001 et 2010 pour évaluer de l'efficacité et de la sécurité du vaccin Cervarix[™] sur une cohorte nord-Américaine et brésilienne.

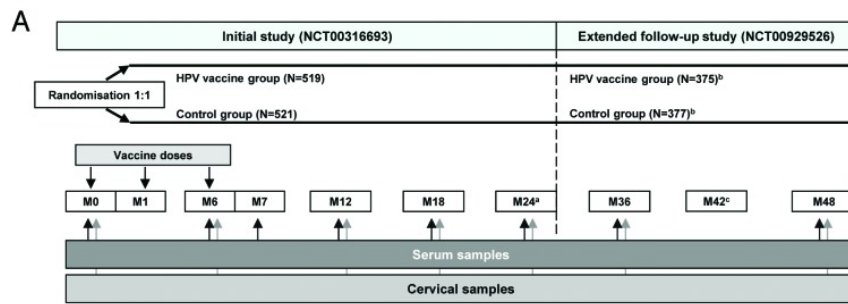


FIGURE A.7 – Organisation des études cliniques HPV-032 (NCT00316693) et HPV-063 (NCT00929526). Ces deux études cliniques ont été menées par l'entreprise GSK entre 2006 et 2011 pour évaluer de l'efficacité et de la sécurité du vaccin Cervarix™ sur une cohorte japonaise. Tous les crédits pour le graphique reviennent à [Konno et al. \[2014\]](#).

Génotypes détectés

TABLE A.2 – Nombre de participantes détectées positives à un génotype donné.

Cohortes	Cohorte brésilienne	Cohorte japonaise	PAPCLEAR	Global
HPV6	11	55	5	71
HPV11	9	12	0	21
HPV16	15	119	12	146
HPV18	5	90	9	104
HPV31	16	70	17	103
HPV33	7	16	1	24
HPV34	3	39	2	44
HPV35	14	28	15	57
HPV39	15	95	15	125
HPV40	9	40	2	51
HPV42	5	31	5	41
HPV43	17	40	3	60
HPV44	12	11	8	31
HPV45	10	8	3	21
HPV51	41	134	32	207
HPV52	43	206	23	272
HPV53	33	106	27	166
HPV54	22	57	15	94
HPV56	13	99	15	127
HPV58	20	95	6	121
HPV59	9	30	8	47
HPV66	20	81	26	127
HPV68	27	56	0 [†]	83
HPV70	20	11	3	34
HPV74	18	22	1	41

[†] Pour plusieurs participantes, il n'a pas été possible de discriminer entre une infection par HPV39, HPV68 ou HPV73 sur le kit de détection.

Estimation des taux de transitions entre état

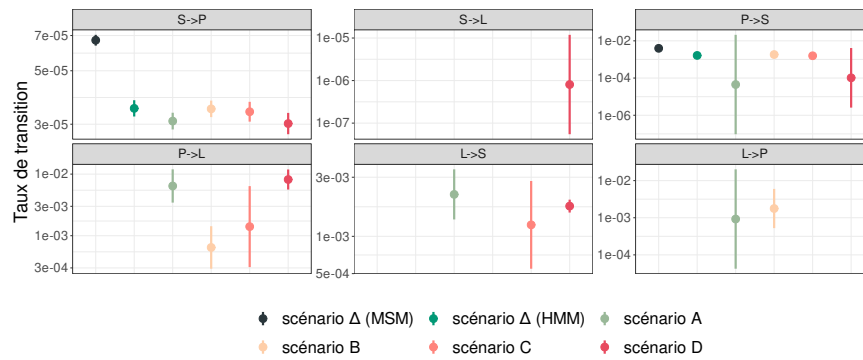


FIGURE A.8 – Générateur du processus caché pour chaque scénario.

Zoom sur les probabilités d'émission de l'état S

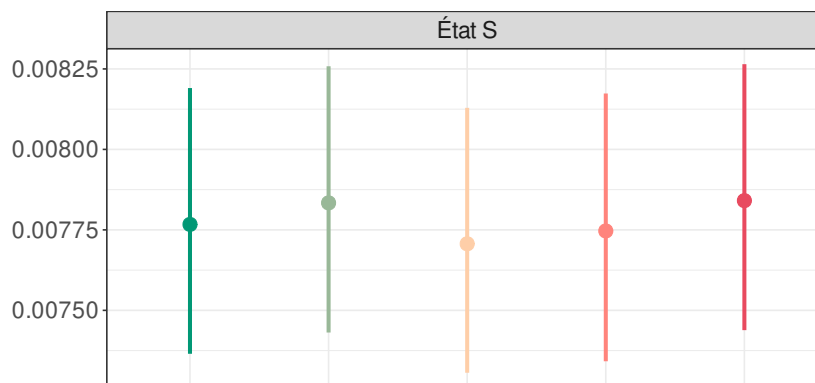


FIGURE A.9 – Probabilité d'émission depuis l'état S . Pour le scénario Δ MSM, la probabilité d'émission depuis l'état S est fixé à 0.

Fonction de survie du temps de retour à l'état sain

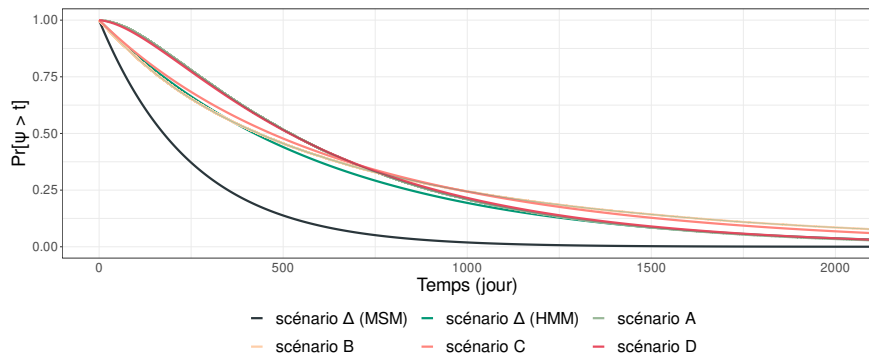


FIGURE A.10 – Fonction de survie du temps de retour à l'état sain.

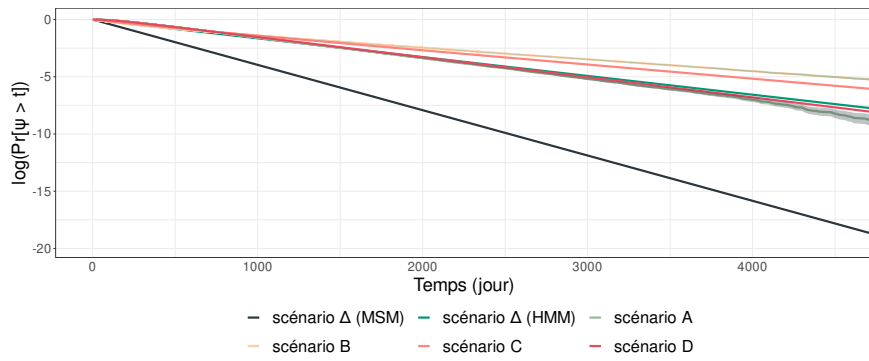


FIGURE A.11 – Logarithme de la fonction de survie du temps de retour à l'état sain.

A.3 Modélisation du cycle de vie HPV

Episome partitioning and symmetric cell divisions :
quantifying the role of random events in the
persistence of HPV infections

Details on intracellular branching processes

We distinguish two cases, constant or Poisson amplification, for the two processes are slightly different. We remind that X_n is a random variable following the total number of episomes in all infected stem cells. Conditioned on the absence of symmetric cell divisions, we model the variation of X_n over time as a Bienaymé-Galton-Watson (BGW) branching process defined by the following recurrence relation, for all $n \in \mathbb{N}$

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_i \quad (\text{A.2})$$

where $\xi_i = \xi$ are independent and identically distributed random variables modeling the offspring number of a single individual within a generation.

Dirac episome multiplication In this case, each episomes distributed to the daughter stem cell gives exactly λ new episomes then dies, so the random variable ξ can be written as

$$\xi = \begin{cases} 0 & \text{with probability } 1 - p \\ \lambda & \text{with probability } p \end{cases} \quad (\text{A.3})$$

Hence we deduce the associated probability generating function (PGF) $g(z)$, $z \in [0, 1]$:

$$g(z) = \sum_{k=0}^{\infty} \mathbb{P}[\xi = k] z^k = 1 - p + p z^\lambda \quad (\text{A.4})$$

Following classical results on BGW branching processes [[Athreya and Ney, 1972](#)], $\mathbb{P}[X_\infty = 0 | X_0 = 1] = p_{\text{ext}}$ is the smallest fixed point of the PGF on the interval $z \in [0, 1]$:

$$p_{\text{ext}} = \inf\{z \in [0, 1], g(z) = z\}$$

One can show that if $m = g'(1) = \lambda p \leq 1$, the average number of viral copies produced by one episome between each iteration, then $p_{\text{ext}} = 1$. If $m > 1$ one must solve the following polynomial equation to find p_{ext}

$$1 - p - z + pz^\lambda = 0 \quad (\text{A.5})$$

For $\lambda \in \{2, 3\}$, such solutions are trivial but for higher value of λ , the solution becomes intractable. The Figure A.12 shows the variation of extinction probability with the parameters λ and p .

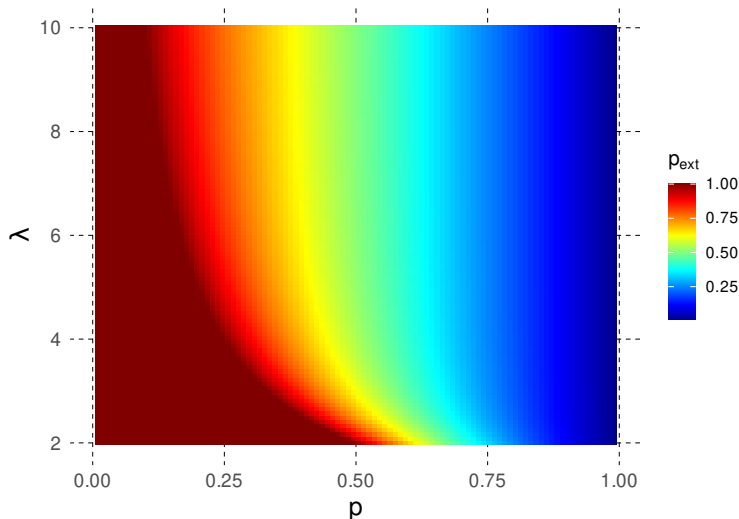


FIGURE A.12 – **Effect of episome replication (λ) and asymmetry in episome partitioning (p) on the cumulative probability of extinction (p_{ext}).** We here assume a Dirac scenario. p_{ext} decreases with both λ and p , and reaches a threshold $p_{\text{ext}} = 1 - p$ for a given p when λ increases. This is consistent with the fact that if λ is sufficiently high, the main source of extinction is the first cell division of a stem cell containing 1 episome.

Poisson episome multiplication We now assume that upon division, each episome dies and gives ρ descendants, where ρ is a random variable following a Poisson distribution, $Poisson(\lambda)$. Hence the distribution of ξ follows :

$$\mathbb{P}[\xi = k] = \mathbb{1}_{\{k=0\}}(1 - p) + p \mathcal{P}^\lambda(k) \quad (\text{A.6})$$

where $\mathcal{P}^\lambda(k)$ is the probability mass function of the Poisson distribution $Poisson(\lambda)$. The PGF associated with such process is

$$g(z) = 1 - p + p e^{\lambda(z-1)} \quad (\text{A.7})$$

Similarly to the previous case, if $m = \lambda p \leq 1$, the probability of extinction is equal to 1, else this probability is lower than 1. Let W be

the Lambert function defined as the inverse function of $z \mapsto ze^z$. If $m > 1$ we have

$$p_{\text{ext}} = \frac{-W(-\lambda p e^{-\lambda p}) + \lambda(1-p)}{\lambda} \quad (\text{A.8})$$

The sensitivity of p_{ext} with the parameters λ and p are displayed in Fig 4.2. Between the two cases, there are few differences, except for the low value of λ where $\mathcal{P}^\lambda(0)$ is non-negligible, hence favoring extinction compared to the Dirac case. We numerically checked our model converges to theoretical predictions. We display some results in Figure A.14. For the majority of the sets, the probability of extinction in our model does converge to numerical predictions. For some values, our model show higher values of p_{ext} , this notably happens when the intrahost capacity (C) and/or the regime is slightly supercritical. Theoretical predictions rely on the assumption the process might reach infinity. We diverge from this assumption when we constrain the intrahost capacity. Such constrain assure an asymptomatic probability of extinction equal to 1 even if the regime is supercritical. Yet, the time to extinction is increasing rapidly with C . Hence on our time interval of 3 years, in most cases, this phenomenon plays a negligible role.

Intracellular dynamics : amplification step prior to distribution step

Assume we reverse the two steps of the intracellular described in the « Methods » subsection*. Upon division, the episomes are first amplified before being distributed in the two daughter cells. If we put the time origin just after the first amplification phase, we notice we are back in the scenario where the distribution of the viral copies occurs before the amplification phase. Therefore we deduce that a process starting with the amplification phase is equivalent to the same process starting with the distribution step and extra initial viral copies. From classical results on BGW branching process, we have the following results, for all $\nu \in \mathbb{N}^*$:

$$\mathbb{P}[X_\infty = 0 | X_0 = \nu] = (\mathbb{P}[X_\infty = 0 | X_0 = 1])^\nu \quad (\text{A.9})$$

Let p_{ext}^D be the probability of extinction when the intracellular process starts with the distribution step and p_{ext}^A the probability of extinction when it begins with the amplification phase. For all $\lambda \in \mathbb{N}$ we hence have :

$$(p_{\text{ext}}^D)^\lambda = p_{\text{ext}}^A \quad (\text{A.10})$$

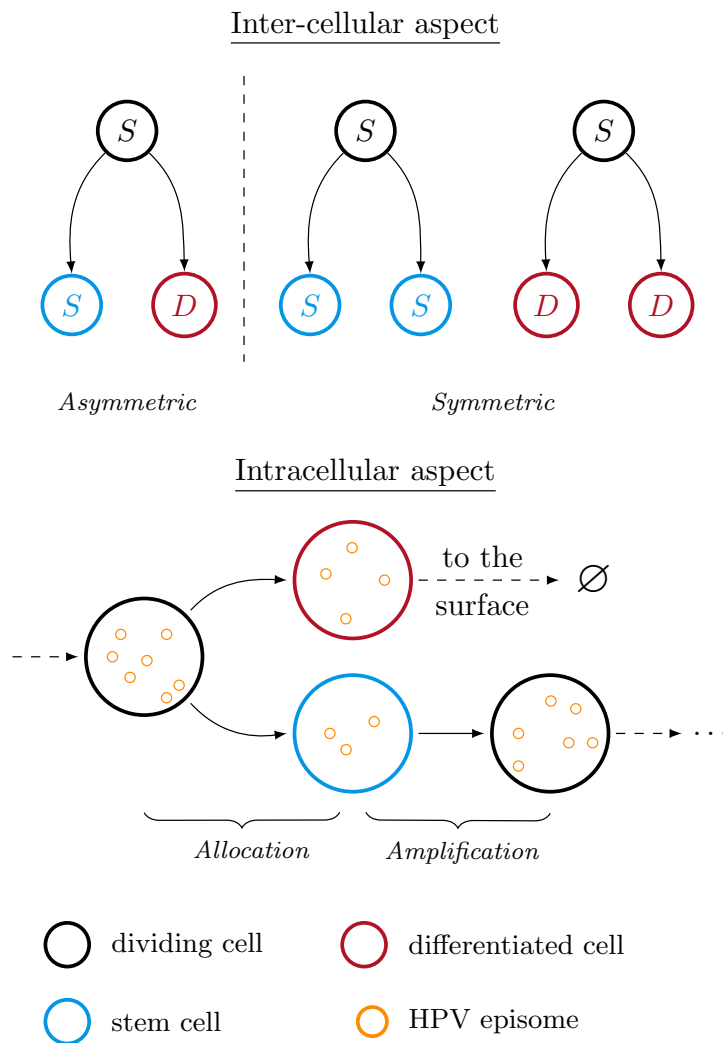


FIGURE A.13 – **Diagram of the stochastic model of HPV-infected cells division.** Stem cells in the basal layer divide mostly asymmetrically. Rare symmetric divisions occur that either yield two stem cells (with probability r) or two differentiated cells (with probability s). These can drastically impact the dynamic of the infection. Upon asymmetric division, viral copies are allocated randomly in the two daughter cells. Each episome is allocated independently to the daughter stem cell with probability p or else goes into the differentiated cell that migrates towards the epithelium surface. These cells do not participate in the persistence of the infection. Copies allocated to the stem cell are then amplified in λ copies on average. This random process is repeated every asymmetric division. The inter-cellular diagram is largely inspired from Fig 1 in reference [Ryser et al., 2015]

Detail on the probability of extinction in the symmetric division case.

We remind the inter-cellular process is modeled using the following discrete BGW branching process. Let Γ_n be the random variable descri-

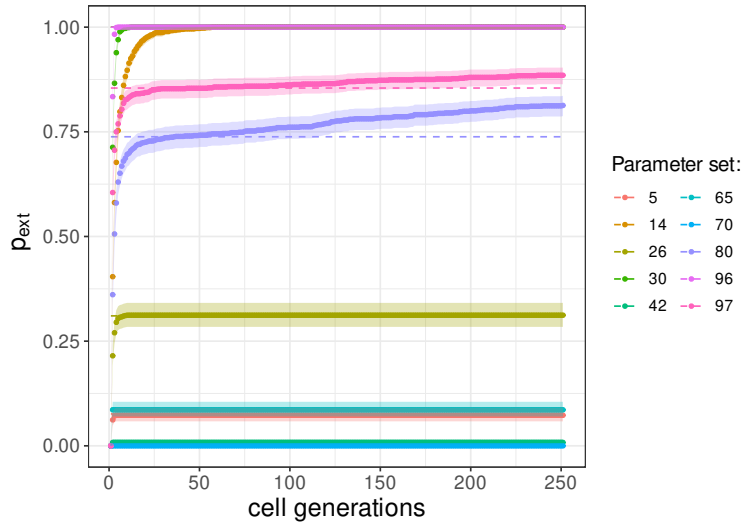


FIGURE A.14 – **Estimated cumulative probability of extinction** (p_{ext} with finite intra-host capacity (plain dots, ribbons correspond to 95% confidence intervals) and according to theory (dashed lines). In most cases, the estimations of p_{ext} under constrained host capacity do not diverge from theoretical predictions on the time window considered. Deviations emerge when the intra-host regime is slightly supercritical and/or the intra-host capacity is limited (< 50). The colors indicate the row of the parameter matrix `LHS_intra` used to obtain the cumulative probability of extinction.

bing the total number of infected stem cells at time $n \in \mathbb{N}$. We note by $\tilde{\Gamma}_n$ the random variable describing the dynamic of infected stem cells when the number of episomes in each stem cell is equal to ∞ (i.e. when omitting intracellular stochasticity). This variable is defined by the following recurrence relation :

$$\tilde{\Gamma}_{n+1} = \sum_{i=1}^{\tilde{\Gamma}_n} \phi_i \quad (\text{A.11})$$

where $\phi_i = \phi$ are independent and identically distributed random variables modeling the offspring number of a single infected cell within a generation. From equation 4.2 we can characterize ϕ as follow

$$\phi = \begin{cases} 2 & \text{with probability } s \\ 0 & \text{with probability } r \\ 1 & \text{with probability } 1 - r - s \end{cases} \quad (\text{A.12})$$

For all $z \in [0, 1]$, the PGF $h(z)$ associated to this reproduction law is defined as :

$$h : z \rightarrow r + (1 - r - s)z + sz^2 \quad (\text{A.13})$$

We deduced that the average number of stem cell produced by each dividing stem cell is equal to $m = 1 - r + s$. The probability of extinction p_{ext} follows

$$p_{\text{ext}} = \inf\{z \in [0, 1], h(z) = z\} \quad (\text{A.14})$$

In this case the results is straightforward and p_{ext} is the smaller roots of the polynomial equation

$$r - (r + s)z + sz^2 = 0, z \in [0, 1] \quad (\text{A.15})$$

Hence we deduced $p_{\text{ext}} = \min\{\frac{r}{s}, 1\}$. We checked numerically our model converges the theoretical predictions and displayed some results in Figure A.15.

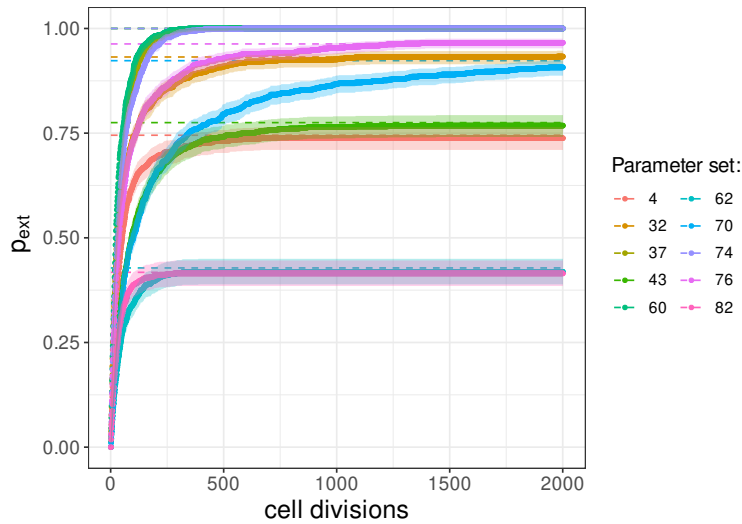


FIGURE A.15 – **Estimated cumulative probability of extinction** (p_{ext} with finite intrahost capacity (plain dots, ribbons correspond to 95% confidence intervals) and according to theory (dashed lines). We observe no significant differences between estimations and theoretical predictions. The colors indicate the row of the parameter matrix LHS_inter used to obtain the cumulative probability of extinction.

Lower bound of the probability of extinction

We remind that Γ_n (resp. $\tilde{\Gamma}_n$) is the random variable describing the dynamic of infected stem cells over time in the general framework (resp. in

the absence of intracellular stochasticity) at time n and $X_n^{(k)}$ the random variable characterizing the number of viral copies in lineage k at time n . We can write that :

$$\begin{aligned} \mathbb{P}[\Gamma_{n+1} = 0 | \Gamma_n = \alpha] &= \mathbb{P}[\tilde{\Gamma}_{n+1} = 0 | \tilde{\Gamma}_n = \alpha] + \\ &\quad \sum_{i=1}^{\alpha} \mathbb{P}[\tilde{\Gamma}_{n+1} = i | \tilde{\Gamma}_n = \alpha] \times \sum_{k=1}^i \mathbb{P}[X_{n+1}^{(k)} = 0 | X_n^{(k)} > 0] \\ \mathbb{P}[\Gamma_{n+1} = 0 | \Gamma_n = \alpha] &\geq \mathbb{P}[\tilde{\Gamma}_{n+1} = 0 | \tilde{\Gamma}_n = \alpha] \end{aligned}$$

This relation holds for all $n \in \mathbb{N}$, thus we deduce that

$$\mathbb{P}[X_\infty = 0 | \Gamma_0 = 1] \geq \min\left(\frac{r}{s}, 1\right) \quad (\text{A.16})$$

Parameters generations using LHS

We generate two matrices of parameters using R package « LHS » [Carnell, 2020]. This package generates matrices $M = (m_{i,j})_{(1 \leq i \leq n, 1 \leq j \leq k)}$ where

$$0 \leq m_{i,j} \leq 1 \quad ; \quad 1 \leq i \leq n, 1 \leq j \leq k$$

As most of our parameters are defined outside of $I = [0, 1]$, we modify $(m_{i,j})$ so that we get meaningful results regarding the parameter support. Let $[a_j, b_j]$ be the definition domain of variable j , $a_j \leq b_j$, then the value of parameter j for the i -th set, $S_{i,j}$ is

$$S_{i,j} = a_j + m_{i,j}(b_j - a_j) \quad (\text{A.17})$$

When variable j is an integer we rounded up $S_{i,j}$. For the categorical variable describing the intracellular amplification regime (Dirac or Poisson distributed), we split the sampling space in two half : we assign the Fixed multiplication regime when $(m_{i,j}) < 0.5$ and the Poisson regime otherwise. See Table 4.1 for more details on the domain of definition of numerical variables. Additionally we set the upper bound for λ and N_0 to 10.

Annotation of infected cell lineages

Similarly to phylogeny, our process viewed from the infected stem cell perspective, can be visualized as a tree whose origin is the first infected cell. As long as the cells descending from that first infected cell divide asymmetrically, we remain on the same branch : we group such cells in a cell lineage. The moment, a stem cell divides symmetrically in two stem

cells, the tree is split into two new branches that follow the same process independently. When a symmetric divisions into two differentiated cells occurs, we stop the branch. To track each branch we use the following notations.

- i) The first cell lineage is labeled **O**.
- ii) The two daughter stem cells are labeled **L** and **R** from its parent cell point of view. To distinguish them from other stem cells in the tree, they are named as follow :
 - Take the parent's name (usually the letter **O** followed by a succession of **L** or **R**).
 - Add the letter **L** or **R** to the parent's name.

For instance the two new stem cells lineage descending from the initial infected stem cell lineage are labeled **OL** and **OR**. Or, upon division, the stem cell lineage **OLLR** gives two new stem cell lineages **OLLRL** and **OLLRR**. The choice of **L** or **R** is random and have no influence on the rest of the process.

Simulation of our process

As the intracellular branching process is nested inside the inter-cellular branching process, we first need to compute the underlying tree of infected cells, then simulate the trajectories of the viral copies for each cell lineage (i.e. each branch) starting from the origin. We simulated the tree and the different trajectories using Python v3.9.1+.

Determination of the inter-cellular tree

The branching process for the inter-cellular dynamic can be viewed as a discrete birth-death process. Each individual, upon division, randomly takes one of the following three paths : i) divide symmetrically in two differentiated cells (death) with probability r ii) divide symmetrically in two stem cells (birth) with probability s iii) divide asymmetrically (pursue its life) with probability $1 - r - s$. Thus for each individual we can compute the time of symmetric divisions as it follows geometric distribution $Geometric(r + s)$. Once the time of symmetric division is determined we randomly assign of the two outcomes by drawing a random number from uniform distribution $Uniform([0, 1])$: if this number is lower than $r/(r + s)$ then the division gives two differentiated cells, otherwise the cell divides into two stem cells. We repeat this method until all cells have divided into two differentiated cells or when the remaining cell divisions time occur after a given threshold.

Intracellular dynamic for each branch

The intracellular branching process can also be consider as a Markov-chain. To simulate the dynamic of a Markov-chain we rely on its associa-

ted stochastic Matrix $P = (P_{i,j})_{(i,j) \in \mathbb{N}^2}$, where $P_{i,j} = \mathbb{P}[X_1 = j | X_0 = i]$. Depending on the episodes amplification regime (Fixed of Poisson - amplification or allocation phase first), the matrix varies. We detail the matrix P for the Fixed and Poisson cases when allocation occurs before amplification. We remind that when amplification occurs before allocation, the results on the probabilities of extinction are similar to the result obtained when the two phases are reversed raised to the power λ . First we explicit the stochastic matrix in the Fixed case :

$$P_{i,j} = \begin{cases} \mathcal{B}_p^i(j/\lambda) & \text{if } j \in \lambda\mathbb{N} ; j \leq \lambda i < C \\ 1 - \sum_{k=0}^{C-1} P_{i,k} & \text{if } j = C \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.18})$$

We note by $\mathcal{B}_p^n(x)$ the probability of drawing x individuals from a binomial distribution $Binomial(n, p)$. If we now assume a Poissonian amplification regime with distribution $Poisson(\lambda)$, then for all $(i, j) \in \mathbb{N}^2$:

$$P_{i,j} = \begin{cases} \sum_{k=0}^i \mathcal{B}_p^i(k) \mathcal{P}^{\lambda k}(j) & \text{if } j < C \\ 1 - \sum_{k=0}^{C-1} P_{i,k} & \text{if } j = C \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.19})$$

We note by $\mathcal{P}^n(k)$ the probability of drawing k individuals from a Poisson distribution $Poisson(n)$. As the state 0 is included, P is squared matrix of size $C + 1 \times C + 1$. Given the stochastic matrix P for all iteration $n \in \mathbb{N}$ and $i \in \llbracket 0, C \rrbracket$, we iteratively determine the trajectory between two events of symmetric divisions by randomly choosing the next value for iteration $n + 1$. We display an example of a random trajectory in Figure A.16.

Effects of initial conditions

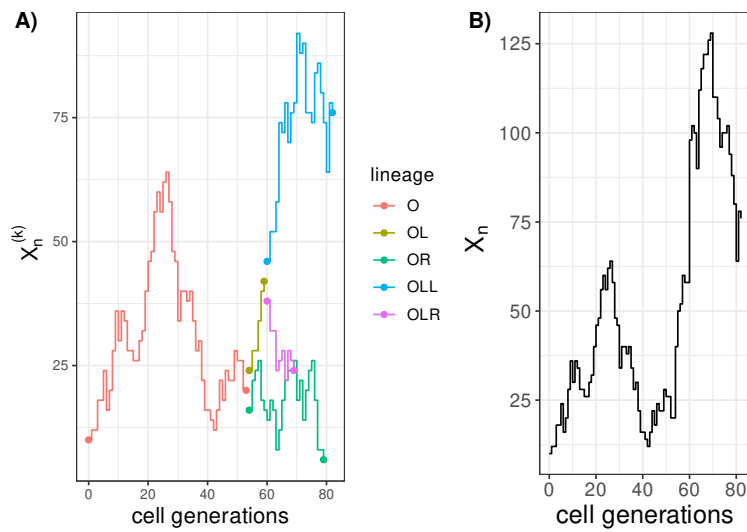


FIGURE A.16 – **Stochastic trajectory of the dynamic of the number of episomes over time** On the left panel (A) we displayed the dynamics inside each cell lineage over time. On the right panel (B) we plot the variation in the total number of episomes. We fixed the value of the parameters as follows : $p = 0.5$, $\lambda = 2$, $s = r = 0.02$, $N_0 = 10$ and $C = 200$. The episomes amplification is fixed and happens after cell divisions. On panel (A), each color represents a cell lineage.

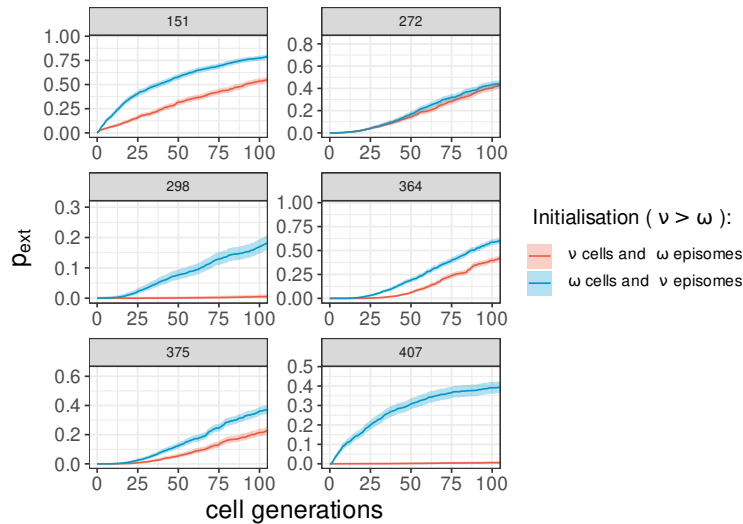


FIGURE A.17 – **Estimations of the cumulative probability of extinction (p_{ext}) for two sets of parameters that only differ by the allocation of the viral copies at the start of infection.** We display in red the estimations of p_{ext} in the scenario where the initial few viral copies are spread in more cells and in blue the scenario where more viral copies are spread in less cells. The solid lines indicates the estimations of p_{ext} and the ribbon the 95% confidence intervals. The cumulative probability of extinction is generally lower in the first scenario compared to the latter. Thus it is more beneficial for the virus to spread its copies in the maximum number of cells at the the start of the infection. On each facet, the title indicates the row of the parameter matrix LHS_A.

A.4 Modélisation de la première vague du COVID-19

Models parameters

TABLE A.3 – Models parameters and constraints.

Notation	Signification	Model	Constraint/value	Reference
R_0	Basic reproduction number	all	3.02	[Sofonea et al., 2020]
p	Infection fatality ratio	all	0.874 %	[Sofonea et al., 2020]
η	Contact rate in the population	all	0.243	inferred
$(\omega_i)_{i \in \mathbb{N}}$	Serial interval distribution	stochastic	discretised <i>Weibull</i> (mean : 4.8 d., sd : 2.3 d.)	[Nishiura et al., 2020a]
θ	Infection to death delay	stochastic	<i>LogNormal</i> (mean : 23.3 d., sd : 9.7 d.)	[Sofonea et al., 2020]
k	Shape parameter (Gamma distribution of the infectivity)	stochastic	0.16	[Lloyd-Smith et al., 2005]
t_0	Time of epidemic wave initiation	SEAIRHD	2020/01/22	inferred
β	<i>Per capita</i> infectious rate	SEAIRH	$\frac{R_0 \gamma \sigma}{S_0 (\gamma + \sigma)}$	calculated
ϵ	Rate of end of latency	SEAIRHD	0.415	inferred
σ	Rate of symptoms onset	SEAIRH	$\frac{1}{5.6 - 1/\epsilon}$	[Linton et al., 2020]
γ	End of infectivity rate	SEAIRHD	0.653	inferred
α	Death rate	SEAIRH	$\frac{1}{17.7 - 1/\gamma}$	[Sofonea et al., 2020]
\mathcal{N}	total population	SEAIRHD	64171900	INSEE

Contact rate variation during and post lock-down period

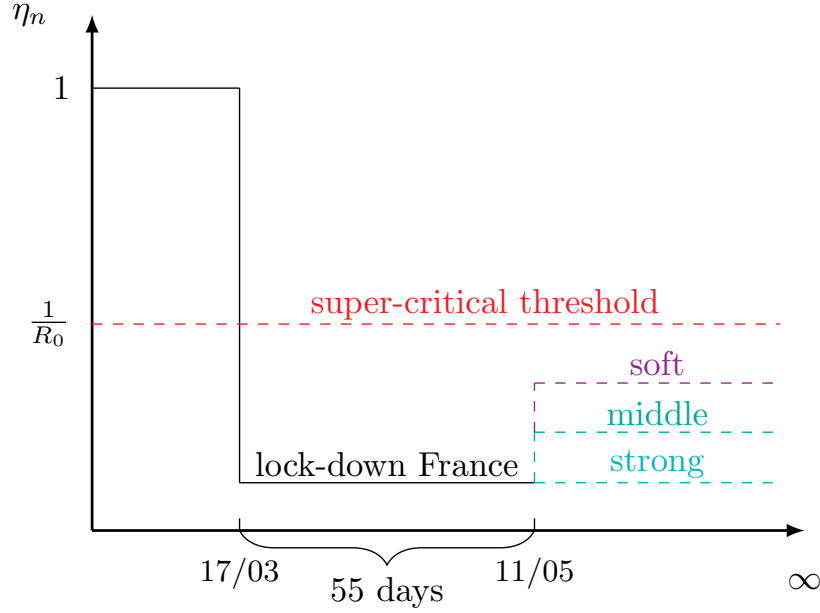


FIGURE A.18 – **Variation of the contact rate (η_t) for the estimation of the time to extinction (τ).** On Mar 17 a lock-down was instated in France. We estimated its efficacy to be around $1-\eta_t = 0.76$ [Sofonea et al., 2020]. The lock-down officially ended on May 11, for a total of 55 days under full lock-down. We evaluated the time to reach extinction under various extension constraints by assuming an infinite lock-down prolongation with fixed intensity, potentially different from the value estimated during the Mar 17 and May 11. The intensity of the lock-down/extension is inversely proportional to the contact rate (η_t).

Markovian SEAIRHD model

The model is captured by the following set of differential equations :

$$\begin{aligned}
 dS/dt &= -\eta_t\beta(A+I)S & dR/dt &= (1-p)\gamma I \\
 dE/dt &= \eta_t\beta(A+I)S - \epsilon E & dH/dt &= p\gamma I - \alpha H \\
 dA/dt &= \epsilon E - \sigma A & dD/dt &= \alpha H \\
 dI/dt &= \sigma A - \gamma I
 \end{aligned} \tag{A.20}$$

where η_t measures the public health intervention impacts on the disease spread at day t , β is the per capita transmission rate of asymptomatic and symptomatic hosts, ϵ is the rate of end of latency, σ is the rate of symptoms onset, γ is the recovery rate, p is the infection fatality ratio (IFR), and α is the rate at which hospitalised patients die. Our

goal is to capture the key features of the infection life-history, especially the incubation period, the asymptomatic transmission, and the delay to hospitalized deaths, but not to fit the epidemic in details.

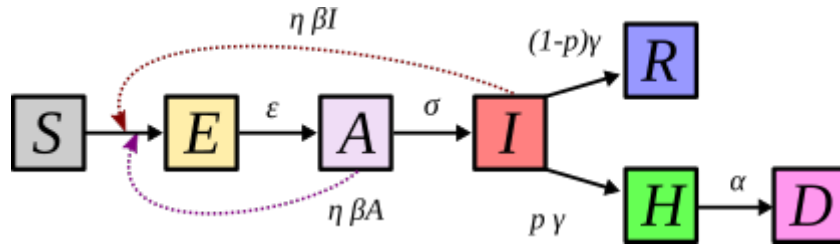


FIGURE A.19 – **Flow diagram of the SEAIRHD model.** This SEAIRHD model corresponds to the ODE model detailed in eq. S1a to S1g. Susceptible individuals (S) can be infected and become asymptotically infected, but noninfectious (E). After a delay, they become asymptotically infectious (A). After symptoms onset, they are symptomatic and infectious (I). A proportion will recover (R), with or without hospitalisation, but some, after a delay during which they are not infectious anymore (H compartment), will die (D). Parameter notations are shown in the main text.

SEAIRH4D model : An alternative model is the SEAIRH4D : in this case, memory is introduced on the delay from infection to death, *ie* it follows an Erlang with shape parameter 4 and the same mean. This is biologically more realistic than the SEAIRHD model where this delay is exponentially distributed.

Deterministic implementation : The set of ODE shown in the previous paragraph is solved using 'odeint' function from Numpy on Python 3.8.3. We then applied a moving average, with a window of 7 days, as done with the real data. We estimated the following parameters for the SEAIRHD and SEAIRH4D models using a maximum likelihood procedure : t_0 , γ , σ and $\eta_{lockdown}$. The maximum-likelihood was found using the Nelder-Mead procedure implemented in the 'minimize' function from Scipy. We computed the likelihood assuming that the daily mortality incidence is Poisson distributed, and independence between daily incidences. We used the data of daily hospital mortality in France from January 1st to May 11 (end of national lockdown), on which a moving average of 7 days is applied to avoid "week-end effects". The code is available on the gitlab repository (see main text Methods).

We compared those two models to the discrete time non-markovian model (Fig. A.20).

Stochastic implementation : Using the same parameters, we simulated 1,000 times a stochastic version of this model, using a Gillespie

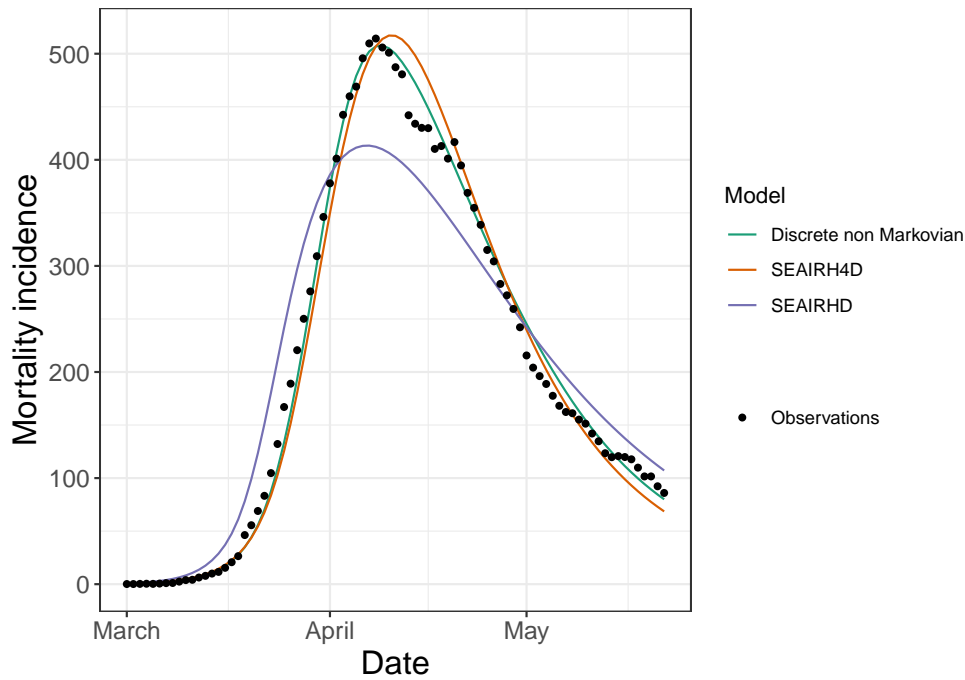


FIGURE A.20 – **Best fits to the observed mortality incidence for several models.** We estimated the best fits to the observed data by minimising the Akaike Information Criterion (AIC). The SEAIRHD shows mediocre fit to the data (AIC : 1324) but taking into account a more realistic delay from infection to death, SEAIRH4D model, improves the estimations drastically (AIC : 730). The discrete time model with memory effects provides a better fit (AIC : 461).

algorithm with the package *TiPS* [Danesh et al., 2023b] on R v.3.6.3 [R Core Team, 2020a]. We stored the time to 100 daily deaths, after applying a moving average with a window of 7 days on the simulations.

DS model sensitivity analysis

Effect of the serial interval distribution

The time elapsed between infection events in an infector-infectee pair is called the generation time. It is a key epidemiological parameter in our model, though almost impossible to measure directly. However, the serial interval distribution - i.e. the time between symptom onsets in an infector-infectee pair- has theoretically the same expectation, but a higher variance.

To our knowledge, the only available data to estimate this parameter come from Asia [Nishiura et al., 2020b; Li et al., 2020]. In our results, we use the distribution inferred by [Nishiura et al., 2020b], *i.e.* a Weibull distribution with mean 4.8 days and standard deviation 2.3 days. This estimate could change, e.g. through behavioral or contact structure



FIGURE A.21 – **Probability mass function for Weibull distributions with mean 4.8 and various standard deviation.** Serial interval distributions used hereafter to analyze the sensitivity to the serial interval standard deviation.

between the countries where the data come from and France. Intuitively, the epidemic growth is very sensitive to the generation time expectation, hence the epidemic starting date would be shifted significantly. Here, we focus on the generation time variance to see to what extent the generation time distribution can affect the epidemiological dynamic (fig. A.21).

We observe that the higher the standard deviation, the later the starting date is inferred (Fig. A.22). The median value is shifted from January 11th to 19th when the standard deviation is doubled, under the homogeneous infectivity model. However, this variation remains within the interval containing 95 % of the variation, using the serial interval determined by [Nishiura et al., 2020b].

Effect of the initial number of imported cases

We made the assumption that only one imported case was responsible for the whole outbreak wave. To note, this is not incompatible with the occurrence of earlier cases, such as the one that occurred on December 27th in France, because in case of individual R_0 heterogeneity, most of those early cases could have died out without contributing to the main outbreak wave.

However, it can be argued that several imported cases may have contributed to the outbreak wave, and consequently may have accelera-

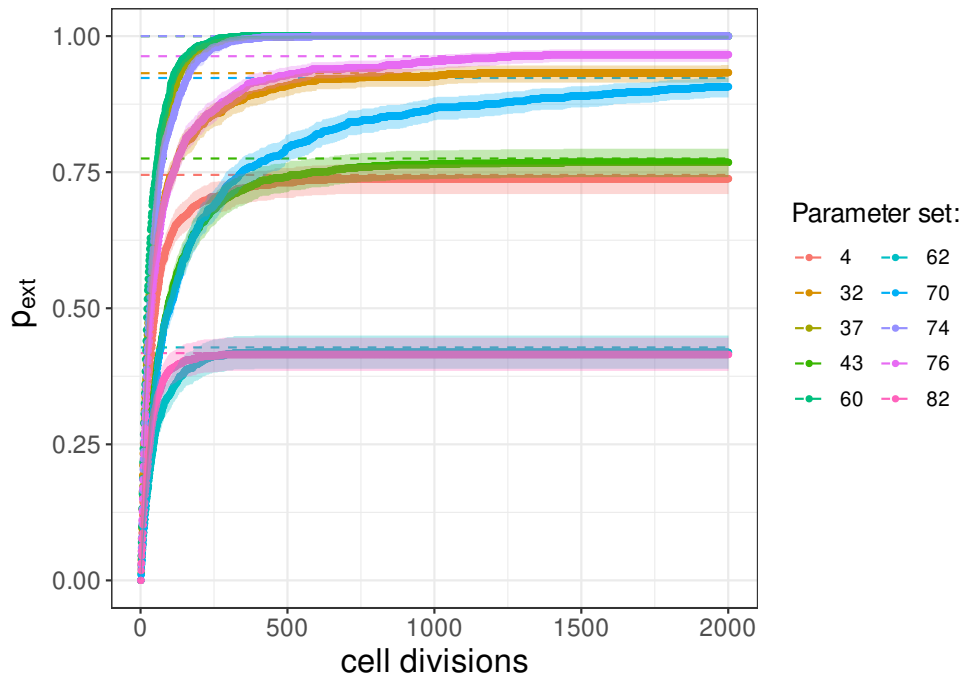


FIGURE A.22 – **Time to 100 deaths distributions for various standard deviation of the serial interval distribution.** We remind that all serial intervals share similar mean following [Nishiura et al., 2020b] estimations. The boxplots show the following quantiles : 2.5%, 25%, 50 %, 75%, 97.5 %. An increase in the standard deviation of the serial interval distribution shorten the time to 100 deaths.

ted the dynamic. This would imply that the outbreak would have started later.

As expected, the total number and intensity of imported cases responsible for the outbreak wave in France affects the starting date of the wave in the order of 4 to 9 days (fig. A.23), which is similar to the sensitivity to the individual infectivity heterogeneity. Increasing the number of daily imported cases decreases the time to 100 deaths, even if the total number of imported cases is equivalent. Moreover, at a given number of daily imported cases, we can see that the difference between an importation during the first 5 days or during the first 30 days of the epidemic is very small. Therefore, only the importation of new infected individuals during the first days of the outbreak has an impact on the epidemic dynamic.

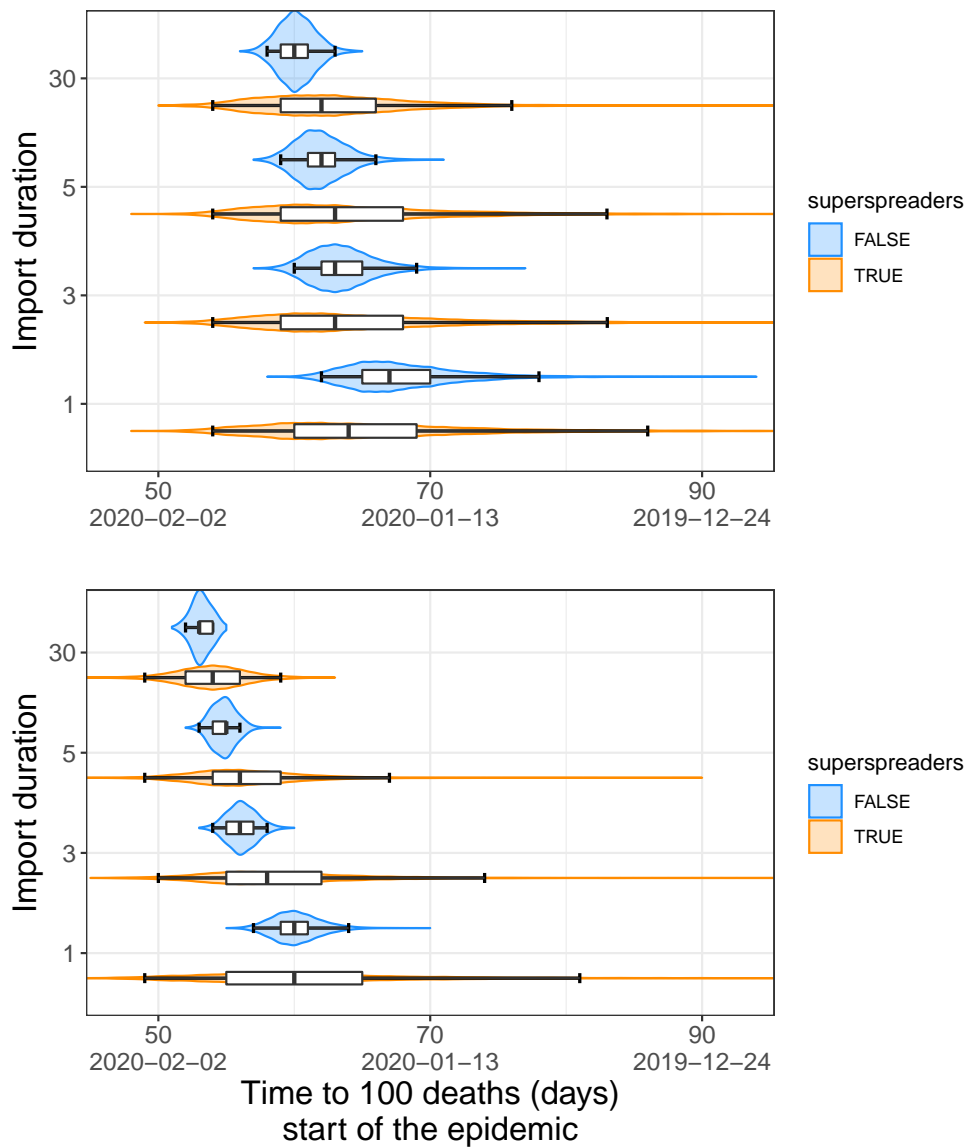


FIGURE A.23 – **Time to 100 deaths distributions for various importation scenarios.** Both panel display the effect of an increase in the period of first cases import on the distribution of the time to 100 deaths. In the upper panel, 1 cases was imported whereas in the lower panel 5 case were imported. The boxplots show the following quantiles : 2.5%, 25%, 50 %, 75%, 97.5 %.

Eradication and rebound risk with superspreading events

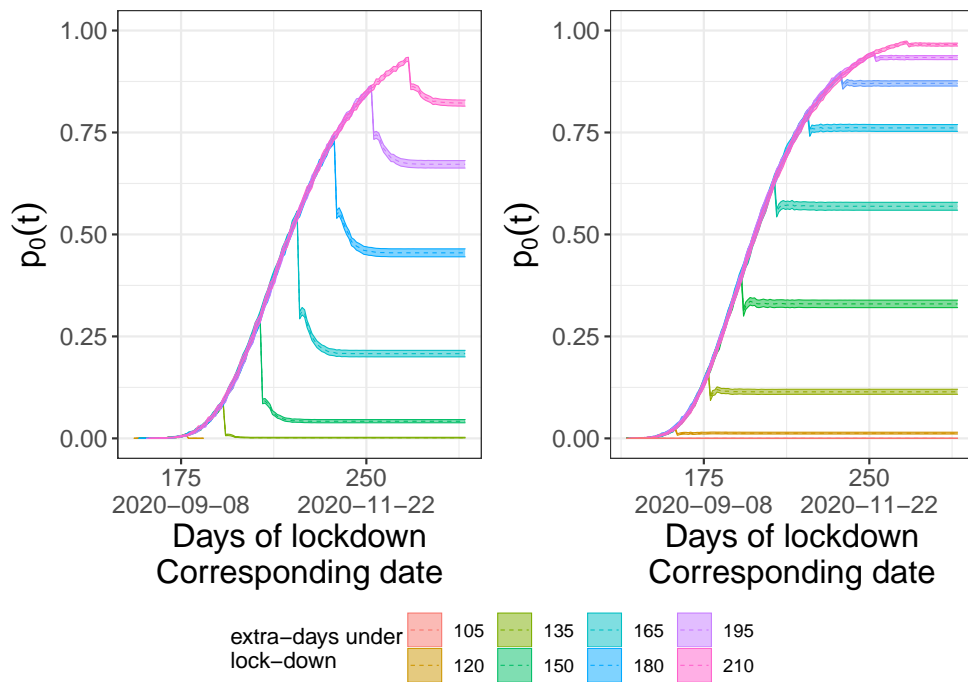


FIGURE A.24 – **Variation of the estimated probability of having no new cases ($p_0(t)$) with finite lock-down extension post 11 May** : Here the x-axis corresponds to the number of days since the start of the lock-down (Mar 17) and to the corresponding date and the y-axis to estimation of the estimated probability of having no new cases. For the first 55 days we set the contact rate to its estimated value in France ($\eta_t = 0.24$), after this period we increase this rate to $\eta_{t>55} = 0.267$ for a fixed duration (d) comprised between 15 days to 210 days. For the rest of the simulation we release all contact rate restrictions ($\eta_{t>55+d} = 1$). On the left panel we displayed the case without superspreading events and on the right panel when accounting for transmission heterogeneity. We plotted the estimation of the probability of epidemic extinction as dashed lines and the confidence interval as solid lines.

Eradication and lock-down initiation date

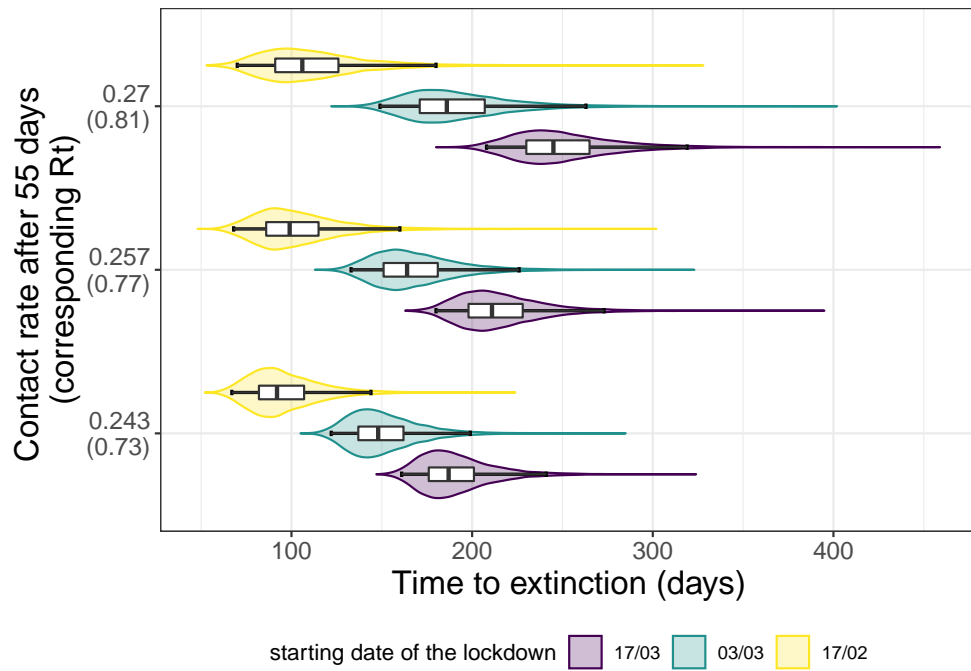


FIGURE A.25 – **Effect of the lock-down intensity, stochasticity, and initiation date on the time to extinction (τ) without individual spreading heterogeneity assumption.** The distributions of the time to extinction (in days since the start of the lock-down) for several contact rate restrictions post 55 first days are plotted on the Y-axis using violin plots and boxplots. In this graph we assume there is no individual spreading heterogeneity. The colors indicate the different initiation date of the lock-down : in purple it starts on Feb 17, green Mar 03 and yellow on Mar 17 (official start). The box extends from the lower to upper quartiles of the data. The whiskers expand from the 2.5% to the 97.5% quantiles.

Time to 20 new cases

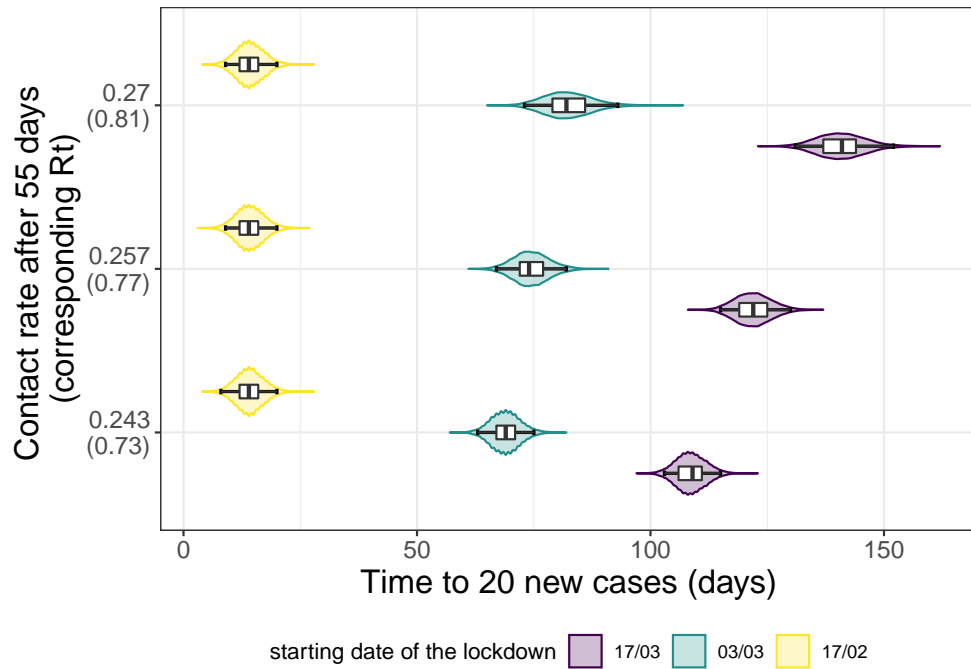


FIGURE A.26 – **Effect of the lock-down intensity, stochasticity, and initiation date on the time to 20 new cases without super-spreading events.** The distributions of the time to 20 new cases (in days since the start of the lock-down) for several contact rate restrictions post 55 first days are plotted on the Y-axis using violin plots and box-plots. In this graph we assume no individual spreading heterogeneity. The colors indicate the different initiation date of the lock-down : in purple it starts on Feb 17, green Mar 03 and yellow on Mar 17 (official start). The box extends from the lower to upper quartiles of the data. The whiskers expand from the 2.5% to the 97.5% quantiles.

ANNEXE

B

Abréviations

- ADN : Acide désoxyribonucléique
- ADNccc : ADN circulaire fermé de façon covalente (*Covalently Closed Circular DNA ; cccDNA*)
- ADNrc : ADN circulaire relâché (*Relaxed circular DNA ; rcDNA*)
- ALT : Alanine AminoTransférase
- Anti-HBe :
- Anti-HBs :
- ARN : Acide ribonucléique
- ARNm : ARN messenger
- ARNpg : ARN pré génomique
- CDC : Centre de contrôle et de prévention des maladies (*Center for Disease Control and prevention*)
- Covid-19 : Maladie à coronavirus 2019 (*Coronavirus disease 2019*)
- DOI : Identifiant numérique d'objet (*Digital Object Identifier*)
- EDO : équation différentielle ordinaire (*Ordinary differential equation ; ODE*)
- IQR : Écart interquartile (*Interquartile ranger*)
- GSK : GlaxoSmithKline
- HBV : Virus de l'Hépatite B (*Hepatitis B virus*)
- HMM : Modèle de Markov caché (*Hidden Markov Model*)
- HPV : Papillomavirus Humain (*Human Papillomavirus*)
- HSV : Virus de l'herpès (*Herpes Simplex Virus*)
- IARC : Centre International de la Recherche sur le Cancer (*International Agency for Research on Cancer*)

- IST : Infection Sexuellement Transmissible
- LUCA : Dernier Ancêtre Commun Universel, parfois raccourci en DACU (*Last Universal Common Ancestor*)
- MERS : Syndrome respiratoire du Moyen-Orient (*Middle-East Respiratory Syndrome*)
- MERS-CoV : Coronavirus du syndrome respiratoire du Moyen-Orient (*Middle-East Respiratory Syndrome-related Coronavirus*)
- MLE : Estimateur du maximum de vraisemblance (*Maximum Likelihood Estimator*)
- MSM : Modèle multi-états (*Multi-state model*)
- ORF : Cadre de lecture ouverte (*Open Reading Frames*)
- PCR : Amplification en chaîne par polymérase (*Polymerase Chain Reaction*)
- PV : Papillomavirus
- RT-PCR : Transcription inverse - amplification en chaîne par polymérase (*Reverse Transcription-Polymerase Chain Reaction*)
- SARS-CoV-2 : Coronavirus 2 du syndrome respiratoire aigu sévère (*Severe Acute Respiratory Syndrome coronavirus 2*)

ANNEXE

C

Liste des figures et tableaux

C.1 Figures

1.1	Génome de HPV16 et organisation des cadres de lecture ouverte et des protéines qu'ils codent. Tous les crédits pour la production de l'image reviennent à de Sanjosé et al. [2018]	13
1.2	Modèles SIS et SIR. Le diagramme du haut correspond au modèle SIS et celui du bas au modèle SIR. La catégorie S correspond aux individus susceptibles (personnes à risque de contracter la maladie), le groupe I aux personnes infectées qui peuvent transmettre le pathogène et l'ensemble R aux personnes immunisées ou décédées (qui ne peuvent plus être contagieux).	23
1.3	Diagrammes de flux pour des modèles SEAIR et $S_1S_2I_1I_2V$. Le diagramme du haut décrit le modèle SEAIR et celui du bas le modèle $S_1S_2I_1I_2V$. Le premier modèle peut être utilisé pour décrire la dynamique d'un pathogène présentant une phase d'incubation non-infectieuse et une phase asymptomatique infectieuse (par exemple dans le cas du COVID-19, voir chapitre 5). Le second cas peut être employé pour décrire l'infection d'un tissu contenant deux lignées cellulaires qui peuvent être infectées par une même particule virale.	27
1.4	Diagramme de l'histoire naturelle des infections HPV et des cancers du col de l'utérus. Tous les crédits pour la production du graphique reviennent à Gravitt [2011]	41

2.1	Cumulative distribution functions for the time to first incident detection and time to first genotype redetection stratified by HR/LR genotypes and effects of host covariates on these estimates. A) Cumulative distribution function (CDF) of the time to fist incident HPV detection since inclusion and the time to the first redetection, stratified by HPV genotype status (HR and LR). B) Hazard ratio for the best models selected by Cox regression with frailty at the patient level. Significant covariates are in red and hazard ratios greater than 1 indicate the covariate is associated with an increased risk of detection, hence lower duration between consecutive episodes.	58
2.2	Survival function for the time of HPV detectability stratified by HR/LR genotypes and effects of host covariates on this estimate. A) Survival functions stratified by the genotype (HR/LR) for the time to loss of HPV detection. B) Hazard ratios for the host factors. Significant factors are in red and a hazard ratio lower than one indicates that the trait is associated with a decreased rate of loss of HPV DNA detection, hence longer survival functions. The reference level is indicated in the bracket for the qualitative variables (see Methods for details). . .	59
3.1	Cinq variantes d'une chaîne de Markov à trois états cachés et à loi d'émission binomiale. Ces cinq modèles supposent l'absence d'un état absorbant, la communication réciproque entre les compartiments S et P et l'impossibilité pour le compartiment L de précéder l'état P sans passage intermédiaire par le compartiment S . Les compartiments dans les rectangles bleus sont les états cachés du processus et les cercles oranges correspondent aux états observés.	71
3.2	Probabilité de détection du virus selon l'état caché. Les probabilités de détection depuis l'état S sont écrasées par la perspective mais un zoom sur les valeurs de p_S selon les scénarios est disponible en annexe (Figure A.9). . . .	75

3.3 **Effet de la vaccination sur les taux de transitions entre états cachés.** Pour toutes les analyses, la référence est la non-vaccination. Un *hazard ratio* inférieur à 1 (sous la ligne tiretée rouge) indique que le temps de transition est plus long pour les personnes vaccinées comparativement aux personnes non-vaccinées, et inversement si le *hazard ratio* est supérieur à 1. Les points en diamants soulignent les *hazard ratios* significatifs et les barres indiquent les intervalles de confiance à 95 %. Par soucis de lisibilité, l'axe des ordonnées a été restreint à l'intervalle [0,12 - 7,5]. 78

Modélisation du cycle de vie HPV 83

4.1 Difference in the CDF of p_{ext} for two sets of parameters that only differ by one value 93

4.2 Effect of λ and p on the CDF of p_{ext} (Poisson) 94

4.3 Variance-based sensitivity analysis for CDF of p_{ext} 95

4.4 Variations in the CDF of p_{ext} with initial values 97

Modélisation de la première vague de COVID-19 en France 102

5.1 Estimated distribution of the number of days until daily mortality incidence reaches 100 cases 113

5.2 Effect of lockdown intensity, stochasticity, and superspreading events on the time to extinction 115

5.3 Effect of the lockdown intensity, stochasticity, and initiation date on the time to extinction under individual spreading heterogeneity assumption. 116

5.4 Effect of the lockdown intensity, stochasticity, and initiation date on the time to 20 new cases under individual spreading heterogeneity assumption. 117

Suppléments article 177

Infections transitoires HPV 177

A.1 **General structure of the PAPCLEAR study.**[Murall et al., 2019b] 178

A.2	Cumulative distribution functions for the time to first incident detection and time to first genotype redetection stratified by HR/LR genotypes and effects of host covariates on these estimates for merged intermittent patterns. A) Cumulative distribution function (CDF) of the time to fist incident HPV detection since inclusion and the time to the first redetection, stratified by HPV genotype status (HR and LR). B) Hazard ratio for the best models selected by Cox regression with frailty at the patient level. Significant covariates are in red and hazard ratios greater than 1 indicate the covariate is associated with a increased risk of detection, hence lower duration between consecutive episodes.	180
A.3	Survival function for the time of HPV detectability stratified by HR/LR genotypes and effects of host covariates on this estimate for merged intermittent patterns. A) Survival functions stratified by the genotype (HR/LR) for the time to loss of HPV detection. B) Hazard ratios for the host factors. Significant factors are in red and a hazard ratio lower than one indicates that the trait is associated with an decreased rate of loss of HPV DNA detection, hence longer survival functions. The reference level is indicated in the bracket for the qualitative variables (see Methods for details).	181
A.4	Graphical assessment for the goodness of Weibull fit. On panel A) we displayed the $\log(-\log(\cdot))$ transformation of the survival function for the time to first incident detection versus the $\log(\text{Time})$ and on panel B) the same transformation for the survival functions of the time to loss of HPV detection. For panel A) the linear fit is acceptable while for panel B) there is a clear non-linear trend. . . .	183
	<i>Modèle de Markov caché pour les infections HPV</i>	184
A.5	Organisation de la partie longitudinale de l'étude clinique PAPCLEAR (NCT02946346).	184
A.6	Organisation des études cliniques HPV-001 (NCT00689741), HPV-007 (NCT00120848) et HPV-023 (NCT00518336). Ces trois études cliniques ont été menées par l'entreprise GSK entre 2001 et 2010 pour évaluer de l'efficacité et de la sécurité du vaccin Cervarix™ sur une cohorte nord-Américaine et brésilienne.	184

A.7	Organisation des études cliniques HPV-032 (NCT00316693) et HPV-063 (NCT00929526). Ces deux études cliniques ont été menées par l'entreprise GSK entre 2006 et 2011 pour évaluer de l'efficacité et de la sécurité du vaccin Cervarix™ sur une cohorte japonaise. Tous les crédits pour le graphique reviennent à Konno et al. [2014].	185
A.8	Générateur du processus caché pour chaque scénario.	187
A.9	Probabilité d'émission depuis l'état S. Pour le scénario Δ MSM, la probabilité d'émission depuis l'état S est fixé à 0.	187
A.10	Fonction de survie du temps de retour à l'état sain.	188
A.11	Logarithme de la fonction de survie du temps de retour à l'état sain.	188
	<i>Modélisation du cycle de vie HPV</i>	189
A.12	Effect of λ and p on the CDF of p_{ext} (Poisson)	190
A.13	Diagram of the stochastic model	192
A.14	Theoretical CDF of p_{ext} with infinite intra-host capacity vs simulated estimations with finite intra-host capacity .	193
A.15	Theoretical CDF of p_{ext} vs simulated estimations without intra-host dynamic	194
A.16	Illustration of an output of the stochastic model	198
A.17	Variation of the CDF of p_{ext} with initial episome allocation	199
	<i>Modélisation de la première vague du COVID-19</i>	200
A.18	Diagram on the modelling of the contact rate variation during lock-down	201
A.19	Flow diagram of the SEAIRHD model	202
A.20	Best fits to the observed mortality incidence for several models.	203
A.21	PMf of the Weibull distributions with mean 4.8 and various SD	204
A.22	Effect of serial interval distribution on time to 100 deaths	205
A.23	Effect of importation scenarios on time to 100 deaths . .	206
A.24	Eradication probability w/-w/o superspreading events . .	207
A.25	Effect of choice of lockdown on eradiction time	208
A.26	Effect of choice of lockdown on time to 20 new cases . . .	209

C.2 Tableaux

2.1	Baseline description of the PAPCLEAR cohort for participants included in the analysis (> 2 visits). Except for the vaccine used not included in the analysis, the first level display for all categorical variables is the reference level used in the Cox analysis. Missing observations were removed from the analysis.	57
3.1	Récapitulatif des participantes incluses dans l'analyse.	73
3.2	Temps de séjour dans les états cachés. L'unité de temps est en nombre de jour.	75
3.3	Temps de retour à l'état sain. L'unité de temps est en nombre de jours. Pour les scénarios A et B, 100.000 trajectoires ont été générées pour estimer la moyenne et les quantiles.	77
	Modélisation du cycle de vie HPV	83
4.1	Parameters and notation of the stochastic model	91
	Suppléments article	177
	<i>Infections transitoires HPV</i>	177
A.1	Estimates for the Weibull parameters (shape and scale) for the time to first incident detection. . . .	182
A.2	Nombre de participantes détectées positives à un génotype donné.	186
	Modélisation de la première vague du COVID-19	200
A.3	Models parameters and constraints.	200

C.3 Prompts midjourney

Les images introduisant les chapitres ont été générées par Midjourney à la version 5.2 :

- **Chapitre 1.** Variations de « *poster of a hand holding a vaccine, promoting vaccination, a virus in the background, in the style of 1950 american commercials , dawn colors, –seed 26011996* »
- **Chapitre 2.** Variations de « *Painting of man photographing big viruses floating in the air with an old Dagerreotype machine, countryside in the 1900 provence, in the style of Monet –seed 26011995* »
- **Chapitre 3.** « *Colored etching of the shadow of a big virus on sliding door, in the style of Ohara Koson, dim light –seed 26011996*
fg
- **Chapitre 4.** Variations de « *Hieronymus Bosch’s painting of lab of math scientists working, poster of viruses in the back, blackboard with equation behind –seed 26011994* »
- **Chapitre 5.** Variations de « *Colored woodcarving of viruses floating around in the air, in an empty street in Paris, in the style of Ohara Koson, dim light, plunging view from the top of a building –seed 26011996* »
- **Chapitre 6.** Variations de « *Edward Hopper’s painting of a 1950 big computer processing punched card, dawn light, view from the back, –seed 25041965* »