Université de Lille
Faculté d'Ingénierie et de Management de la santé (ILIS)
Master Healthcare Business et Recherche Clinique

Valentine DEMOERSMAN

La personnalisation de la prévention grâce aux Big Data

Sous la direction de David DELERUE

Mémoire de fin d'étude de la 2e année de Master

Année scolaire 2018-2019

Master Healthcare Business et Recherche Clinique

Mémoire soutenu le 8 Octobre 2019

Composition du jury :

David DELERUE : Directeur de mémoire

Hélène GORGE : Présidente du jury

Régis LOGIER : Troisième membre du jury

Faculté d'Ingénierie et de Management de la santé (ILIS)

42 Rue Ambroise Paré

59120 LOOS

Remerciements

Je tiens à présenter des remerciements chaleureux à mon directeur de mémoire, Monsieur David DELERUE, pour sa patience, sa disponibilité, son encadrement, ses conseils et son expertise qui ont contribué à alimenter ma réflexion sur le sujet de ce mémoire.

J'adresse mes sincères remerciements à Madame Maya KHALIFE, Docteur Bruno LERNOUT et Docteur Jean-Marie GROSBOIS pour m'avoir permis de leur poser toutes mes questions au cours de mes recherches. Respectivement, votre expertise en tant qu'éditeur de logiciels, de médecin hospitalier et de médecin hospitalier intervenant à domicile m'ont permis d'avoir une vision plus exhaustive de la prévention en France et de faire avancer mon analyse.

Je désire aussi remercier Madame Hélène GORGE, Monsieur Régis LOGIER ainsi que toute l'équipe enseignante d'ILIS pour les connaissances qu'ils ont essayé de nous transmettre et pour nous avoir permis de les mettre en pratique en entreprise et lors de la rédaction de ce mémoire.

Je remercie également mes collègues de la société ALICANTE pour leur confiance et la bienveillance qu'ils m'ont témoigné.

Enfin, je tenais particulièrement à remercier Pierre ainsi que ma famille pour leur patience et leurs encouragements indéfectibles.

Sommaire

Remerciements **2**

Sommaire **3**

Table des illustrations **5**

Introduction **6**

La prévention grâce aux Big Data, une fiction… ancienne 6

La fiction qui inspire la réalité 6

La démocratisation du quantified self : la prévention basée Data pour tous ? 7

Présentation du sujet 7

Cadrage du sujet, limites de l'étude 7

Organisation du présent mémoire 8

I. Définitions préliminaires **9**

1.1 La prévention, un concept difficile à définir 9

1.2 Les Big Data 11

1.3 L'intelligence artificielle 11

II. Quelles utilisations de la Big Data et de l'intelligence artificielle dans la prévention ? **12**

2.1 La prévention dans l'histoire récente de la médecine et dans l'organisation de la santé, en France 12

2.2 Bibliographie synthétique : la prévention grâce aux Data et à l'intelligence artificielle, dans l'activité scientifique après les années 1950 15

2.3 Les Big Data : quelles données candidates dans le secteur santé, pour la prévention ? 19

2.4 En conclusion de cette partie 32

III. Quels impacts pourraient avoir ces outils de prévention sur l'économie globale de la santé ? **33**

3.1 La prévention en France 33

3.2 La prévention dans d'autres systèmes que le système français 40

IV. La prévention personnalisée constitue-t-elle un nouveau marché ? **51**

4.1 La demande 52

4.2 L'offre 55

Conclusion 61

Prévention santé, basée Data et intelligence artificielle : une activité en croissance 61

Prévention santé, basée Data et intelligence artificielle : de nombreux défis à relever 61

Les principales attentes du terrain : la Data et l'intelligence artificielle pour une prévention personnalisée au quotidien 64

Une évolution probable de l'organisation de la prévention 65

Bibliographie 66

Table des matières 72

Table des illustrations

Figure 1 Tricorder de Star Trek 6

Figure 2 Scout de la société Scanadu, inspiré du tricorder de Star Trek 6

Figure 3 Affiche de la Société Française de Prophylaxie Sanitaire et Morale 13

Figure 4 Nombre d'articles pour la requête 16

Figure 5 Mots clés "auteur" les plus fréquents dans le résultat de la requête 17

Figure 6 : représentation dans vosviewer des mots clés médicaux "auteur" principaux dans la sélection d'articles 18

Figure 7 Utilisation des mots clés 19

Figure 8 Représentation synthétique de la diversité des données disponibles à l'hôpital .. 21

Figure 9 Apprentissage décentralisé (25) 31

Figure 10 Répartition du financement de la prévention institutionnelle 35

Figure 11 Exemples de campagnes de prévention réalisées à l'échelle nationale via les médias 36

Figure 12 Présentation de la feuille de route "Accélérer le virage numérique" 38

Figure 13 Centres de soins Kaiser Permanente en Californie du Nord (56) 47

Figure 14 Un exemple de médecine préventive (57), Kaiser Permanente 49

Figure 15 Courbe de Kaplan Meyer montrant l'efficacité du programme “Collaborative Cardiac Care Service”(58) 51

Figure 16 Les dix principales maladies considérées dans la littérature sur l'intelligence artificielle 54

Figure 17 Copie d'écran de l'outil de visualisation et d'affichage de prédictions ©ARTERYS 59

Introduction

La prévention grâce aux Big Data, une fiction… ancienne

Dès 1966, *Star Trek(1)* faisait apparaitre dans ses épisodes le tricordeur, appareil mobile permettant de détecter, d'enregistrer et d'analyser grâce à des capteurs spécialisés dans la détection et l'analyse des signes vitaux, à des processus informatiques assurant la fonction de diagnostic et de comparaison à des bases de données de référence, qu'on imagine maintenant bâties sur des données massives. Le premier scanner portatif permettant de collecter de la donnée et de la comparer à des valeurs normales était imaginé.

De même, en 1987, on peut apercevoir, dans *Robocop (2)*, les premiers scanners thermiques et capteurs implantés permettant de monitorer le corps des individus.

L'exploitation des données vitales, en temps réel, bien avant la naissance est imagée dans *Bienvenue à Gattaca(3)* en 1997. Dans ce film, c'est l'Etat qui est propriétaire des données de ses citoyens qui sont soumis à des check-up quotidiens.

Plus récemment, en 2005, cela a également été imaginé dans *The Island (4)*, où l'urine et le sommeil sont des données enregistrées en temps réel et gardées en mémoire.

La fiction qui inspire la réalité

Cette fiction a parfois pu largement inspirer la réalité : le *Scout* de la société Scanadu avait été pensé pour permettre de mesurer la température corporelle, l'oxygénation du sang, la fréquence cardiaque et la pression sanguine.

Figure 1 Tricorder de Star Trek

Figure 2 Scout de la société Scanadu, inspiré du tricorder de Star Trek

La démocratisation du quantified self : la prévention basée Data pour tous ?

Ces outils sont classés dans la catégorie dite du « quantified self » (traduit par « automesure », « auto-quantification ») puisqu'ils permettent, en toute autonomie et sans examen médical, d'enregistrer, d'analyser voire de partager ses propres données. Ils sont le plus souvent reliés au smartphone ou à un ordinateur qui sert de relais, qui ensuite grâce en local ou grâce à des services Internet permettent de réaliser des graphiques et de comparer ses propres données aux valeurs dites normales (5).

Ces outils ont connu un déploiement commercial massif : il semble que la prévention grâce aux données, accessible à tous, promise par la fiction, devienne une réalité.

Mais est-ce réellement le cas ? L'approche « prévention personnalisée basée Data » promue notamment par ces outils a-t-elle prouvée son efficacité ? Quels sont les autres moyens, au domicile, en ville, à l'hôpital, de prévenir des maladies à partir des données ?

C'est à ces questions principales que ce mémoire souhaite apporter des éléments de réponse.

Présentation du sujet

Dans ce mémoire, nous tenterons d'objectiver et de comprendre l'intérêt de l'utilisation de l'intelligence artificiellef et des Big Data dans certains domaines de la prévention. Nous discuterons de la possibilité de personnaliser cette prévention grâce aux technologies informatiques actuelles, basées sur les données. Nous décrirons quelques acteurs du marché, base de l'identification de quelques modèles économiques candidats. Enfin, nous tenterons d'identifier les freins actuels au développement de la prévention basée sur l'intelligence artificielle. Pour cela, nous avons rencontré Madame Maya KHALIFE, Product Manager chez Arterys qui nous a donné son avis d'éditeurs de logiciels, Docteur Bruno LERNOUT en tant que médecin hospitalier et Docteur Jean-Marie GROSBOIS, médecin acteur de prévention secondaire et tertiaire en ville.

Cadrage du sujet, limites de l'étude

Comme nous pourrons le constater, la prévention, la data, l'intelligence artificielle, sont des notions vastes aux multiples origines, définitions et applications. Il semble impossible dans ce format d'aborder toutes les facettes et toutes les combinaisons.

C’est pourquoi, dans ce mémoire, nous nous intéresserons spécifiquement à la prévention collective, en santé humaine.

La zone géographique étudiée se limitera à la France. Une description rapide du système adopté dans d’autres pays sera cependant proposée, afin en contraste, de mieux identifier les freins qu’il peut y avoir en France, et les moteurs dans ces zones.

Un focus important sera fait sur le monde hospitalier : la prévention à l’hôpital, les Data à l’hôpital. De ce fait, la prévention s’appuyant sur des données autres que des données de santé, comme les données de pollution atmosphérique, de pollution de l’eau, etc., ne sera pas abordée.

Organisation du présent mémoire

Dans un premier temps, nous proposerons quelques définitions des notions principales traitées dans le mémoire : prévention ; big data.

Ensuite, nous ferons un état des lieux de l’utilisation de la Data et de l’intelligence artificielle dans les activités de prévention.

Dans une troisième partie, nous nous intéresserons sur l’impact des outils basés sur l’analyse des Big Data en prévention personnalisée dans le monde.

Enfin, dans une quatrième partie, nous nous concentrerons sur le nouveau marché commercial que peut constituer ce genre d’outil.

I. Définitions préliminaires

1.1 La prévention, un concept difficile à définir

La prévention en santé publique constitue l'ensemble des actions préventives, curatives, sociales et éducatives qui permettent d'aborder la santé de façon collective. On peut distinguer :

- La lutte contre les épidémies,
- La prévention des maladies, handicaps et accidents
- L'éducation à la santé
- La surveillance et l'amélioration de la santé au niveau populationnel
- L'identification et la réduction des risques sanitaires (travail, environnement etc.)
- La réduction des inégalités et l'accès aux soins.

En d'autres termes, la prévention en santé vise à limiter et éviter l'apparition et le développement des maladies, en agissant sur tout ou partie de la population et en particulier sur ses comportements et son environnement. La prévention peut s'exercer dans différents lieux de vie comme le domicile, l'école, le travail, les établissements de santé etc.

On distinguera la prévention incitative qui consiste à responsabiliser la population de la prévention coercitive qui vise à mettre en place des lois.(6)

Selon l'Organisation mondiale de la santé (OMS), la prévention peut être définie comme « tout acte destiné à éviter des phénomènes attendus ». Les définitions de la prévention seront différentes si l'on parle en termes d'économie et de santé publique. En économie, les travaux d'Ehrlich et Becker en 1972 (7) ont permis de distinguer deux types de prévention : l'autoprotection et l'auto-assurance. L'autoprotection concerne toutes les actions entreprises pour réduire la probabilité de survenue d'un sinistre alors que l'auto-assurance s'intéresse aux actions visant à limiter l'extension du sinistre s'il survient. Autrement dit, la prévention désigne l'ensemble des actions menées en vue de prévoir ou d'empêcher la

survenue de certains risques mais également de réduire les conséquences négatives induites par leur survenue.

En 1948, l'OMS défini la prévention comme « l'ensemble des mesures visant à éviter ou réduire le nombre ou la gravité des maladies, des accidents et des handicaps ». Cela a permis de distinguer trois catégories d'actions préventives. La prévention primaire représente l'ensemble des actions visant à diminuer l'occurrence ou l'incidence d'une maladie au sein d'une population et à réduire les risques d'apparition de nouveaux cas. La vaccination est par exemple une action de prévention primaire.

La prévention secondaire vise à diminuer la prévalence d'une maladie dans une population. Ce type d'action préventive intervient dès l'apparition du trouble ou de la maladie et vise à faire disparaitre les facteurs de risque pour ainsi réduire la morbidité ou les conséquences d'une maladie une fois celle-ci apparue. Ainsi, la prévention secondaire se concentre aujourd'hui sur le dépistage, le plus précocement possible, des maladies.

La prévention tertiaire représente les actions qui ont pour objectif de réduire les complications, invalidités ou rechutes suite à la maladie. La prévention tertiaire inclue également la réadaptation du malade d'un point de vue médical, psychologique et social.

Il existe également une autre façon de définir la prévention. En effet, en 1928, RS. Gordon présenta une classification de la prévention en 3 parties :

La prévention universelle est destinée à l'ensemble de la population, quel qu'en soit son état de santé.

La prévention sélective s'adresse à une population spécifique comme les enfants, les travailleurs, les automobilistes etc.

La prévention ciblée concerne des groupes de personnes dont les facteurs de risques sont plus élevés que la normale. Par exemple la toxoplasmose chez les femmes enceintes.(8)

Enfin, nous pouvons distinguer la prévention dite institutionnelle de la prévention contenue dans la « consommation de soins et de biens médicaux ».

Dans le premier cas, la prévention institutionnelle est organisée dans le cadre de programme nationaux ou départementaux, avec des financements très divers (Sécurité Sociale, Etat et collectivités locales, secteur privé). Elle correspond à des choix collectifs, à une vision socialisée de la protection de santé.

Dans le second cas, il s'agit d'une prévention individuelle, à l'initiative du citoyen qui consulte lui-même des institutions de soins et procède à des achats de produits pharmaceutiques ou de mieux-être mais qui ne relève pas de programmes institutionnalisés.(9) (10)

1.2 Les Big Data

« Big Data » est un concept qui n'a pas de définition scientifique formelle, ou alors celles-ci ne sont pas consensuelles. Les Big Data résultent de la massification des données numériques. Ce concept regroupe une famille d'outils qui répondent à la règle de 4 V : elles représentent un volume de données considérable à traiter, une grande variété d'information provenant de sources diverses, structurées ou non, à une vélocité importante puisqu'elles sont maintenant collectées sous forme de flux et dont la véracité est à déterminer. En effet, le principal défi du Big Data est l'exactitude de ces méga données, dont l'enjeu réside sur le plan économique mais également en tant que moteur d'innovation.

1.3 L'intelligence artificielle

L'Intelligence Artificielle n'a pas non plus de définition consensuelle. Certains scientifiques comme Luc Julia estiment même qu'elle n'existe pas. Dans ce mémoire, nous considèrerons que l'intelligence artificielle est la science dont le but est de faire faire par une machine des tâches que l'homme accomplit en utilisant son intelligence. Le terme d'intelligence artificielle est apparu en 1956. (11)

Jean-François Goglin, Conseiller Système d'Informations de santé à la FEHAP, définit l'intelligence artificielle comme étant un système numérique capable de résoudre des problèmes complexes selon des raisonnements similaires à la logique et au raisonnement humain. L'intelligence artificielle perçoit l'environnement, l'analyse et prend une décision en conséquence mais surtout apprend. (12)

L'intelligence artificielle est presque toujours associée à des notions d'apprentissage. L'apprentissage permet à la machine d'améliorer ses performances, d'apprendre de

nouvelles tâches et d'acquérir de nouvelles compétences. Il peut être supervisé, dans ce cas l'algorithme apprend à fournir la bonne réponse à partir de données caractérisant des exemples. Cet apprentissage peut également être non supervisé : l'algorithme apprend seul à partir de données brutes. Il peut enfin être renforcé, cela implique que l'algorithme réagit à son environnement.

II. Quelles utilisations de la Big Data et de l'intelligence artificielle dans la prévention ?

2.1 La prévention dans l'histoire récente de la médecine et dans l'organisation de la santé, en France

2.1.1 L'apparition du « Mieux vaut prévenir que guérir » dans les politiques publiques

La notion de santé publique date de la fin du 18e siècle et apparait comme une prérogative pour l'Etat. A cette époque, on ne sait pas encore soigner les maladies, bien que leurs origines et mécanismes soient de plus en plus étudiées. C'est à ce moment-là que les notions de prévention voient le jour. Quelques vaccins comme celui contre la variole sont déjà utilisés mais ils ne sont pas nombreux. La prévention concerne ainsi les conditions de travail et les comportements. Au siècle des Lumières, la santé publique laisse penser que la réduction voire la disparition des maladies serait possible grâce à une modification des comportements de la société. La Révolution remettra toute ces croyances en cause au profit d'une modification des comportements individuels (13).

2.1.2 L'avant-guerre

Pour connaitre les vraies avancées médicales en termes de prévention, il faut attendre la fin du XVIII siècle où, Edward Jenner fut le premier à pratiquer la vaccination. En effet il mit en place un vaccin contre la variole, épidémie qui a déjà décimé des milliers de personnes. Une centaine d'années plus tard, Louis Pasteur créa le premier vaccin atténué en injectant des virus affaiblis à ses patients.

Le développement de la politique de prévention a également été mené par Alfred Fournier, médecin spécialisé en vénérologie, notamment sur la syphilis. En 1886, cette

maladie ne connait aucun traitement efficace et est de plus en plus répandue. La politique de prévention pensée par Fournier avait pour but d'alerter au mieux le public mais surtout les médecins sur le danger de la syphilis. En 1901 fut ainsi créée la Société Française de prophylaxie sanitaire et morale dont l'un des objectifs principaux était la mise en place d'actions de prévention. Ces actions de prévention étaient d'une part destinée à informer la population sur la transmission de cette maladie, par l'utilisation d'images choquantes montrant des patients atteints de syphilis, de mettre en lumière des œuvres comme « Le baiser qui tue » grâce au cinéma ou de pièces de théâtres comme « les avariés » d'Eugène Brieux qui mettaient en scène le concept d'hérédosyphilis selon lequel la syphilis du père était due à des relations sexuelles fautives avec des prostituées et avait des conséquences sur la descendance de cet homme.

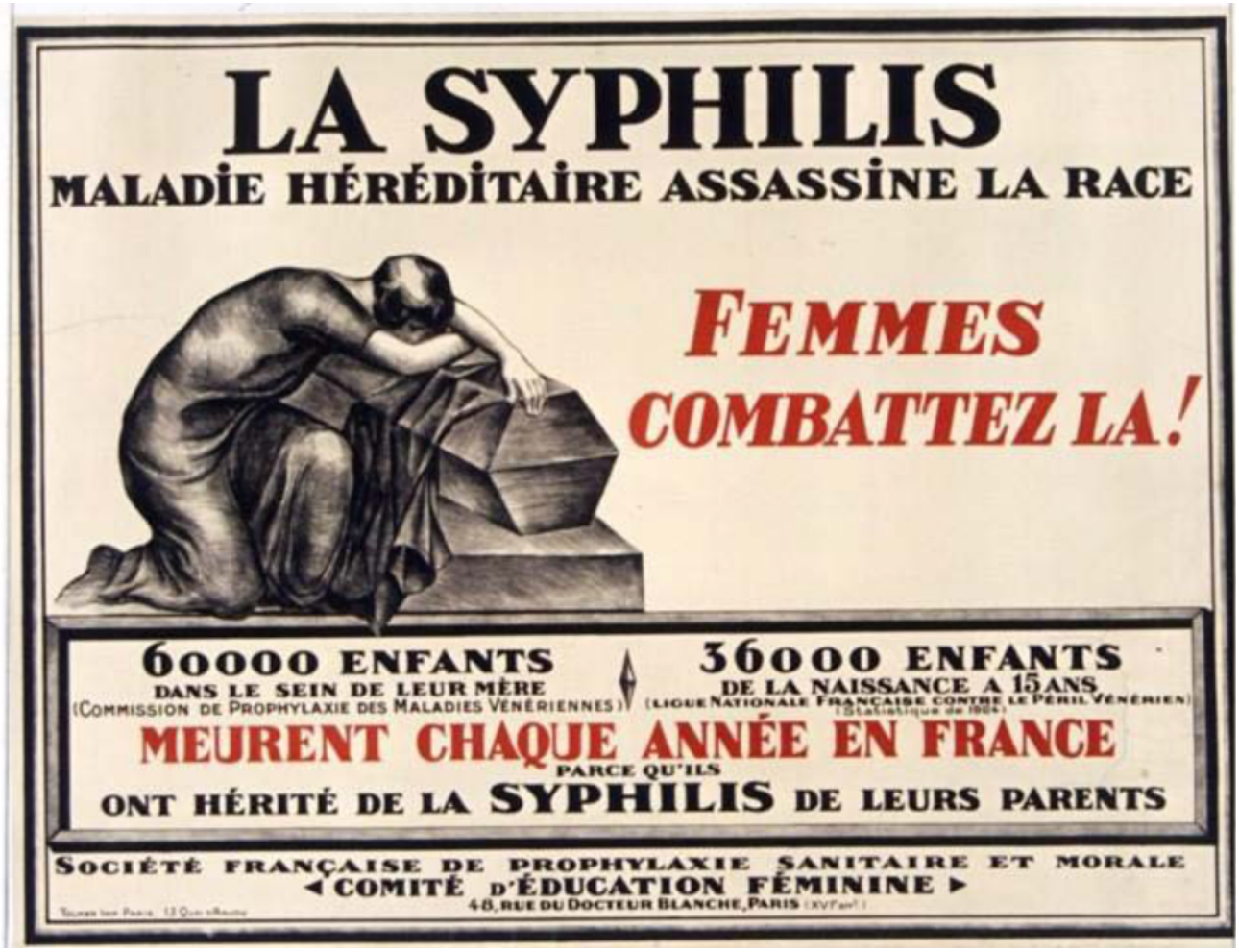

Figure 3 Affiche de la Société Française de Prophylaxie Sanitaire et Morale

Pour informer les médecins sur cette maladie, il demanda que chaque étudiant soit formé à la syphiligraphie par un stage obligatoire dans le service des maladies vénériennes.

Le financement des actions de prévention se fit, en partie, par l'instauration de surtaxe sur certains produits comme les timbres.

2.1.3 L'entre-deux guerres

Au début des années 1920, la population sort d'un conflit qui a affaibli le pays, ses habitants et le corps médical. Pour pallier ce contexte très difficile, l'Etat a décidé d'accentuer son engagement en santé publique notamment dans la lutte contre l'alcoolisme, la syphilis, la tuberculose et le cancer. En janvier 1920, une nouvelle politique de santé publique a été mise en place avec la création du ministère de l'Hygiène, de l'Assistance et de la Prévoyance sociale.

2.1.4 Le développement de la santé publique dès la fin de la seconde guerre mondiale

Au lendemain de la seconde guerre mondiale, deux instances sont créées : la sécurité sociale voit le jour en 1944 et la médecine du travail nait en 1945. Toutefois, la prévention a encore des difficultés pour être mise au premier plan du système de santé. Des campagnes sont menées, contre l'alcoolisme par exemple mais elles seront jugées, a posteriori, inefficaces. La prévention est surtout réalisée au travail où l'on constate que prévenir les accidents de travail et les maladies liées à l'activité professionnelle permet de disposer de ses salariés dans de meilleures conditions puisqu'ils sont moins absents et plus heureux.

À partir de 1968, date de création de l'INRS (Institut national de recherche et de sécurité), les études et recherches sont intégrées dans les missions de l'Institut. Ces travaux de recherche appliquée concernent des domaines aussi variés que la physique, la mécanique, l'ergonomie, la chimie ou la toxicologie… Ils permettent à l'INRS de proposer des solutions de prévention répondant mieux aux besoins du monde du travail. Aux activités de formation, de recueil et de diffusion de l'information s'ajoutent donc la conduite de recherches et la valorisation et la diffusion des résultats de ces travaux.

Depuis les années 80, l'activité de recherche a considérablement évolué, avec comme défi permanent, le transfert de ces résultats vers les entreprises et les spécialistes de la prévention (réseau de l'Assurance maladie, médecine du travail, CHSCT…). L'INRS peut ainsi diffuser des connaissances validées et un savoir-faire adapté pour que les acteurs ou

relais de la prévention mettent en place des actions efficaces. Il s'agit par exemple d'actions de formation, ou de diffusion aussi large que possible de supports et d'outils. C'est également à cette époque que la prévention prend un nouvel essor. Elle s'inscrit désormais le plus en amont possible, dès la conception des locaux ou des équipements de travail.

A travers les époques, nous avons pu voir que la prévention a essayé, timidement, de faire partie intégrante du système de santé. Nous avons également pu constater que la notion de prévention a elle-même évoluée au fil du temps.

Par contre, l'utilisation systématique de données par ces divers instituts français, créés depuis le XIXème siècle, n'est pas documentée.

2.2 Bibliographie synthétique : la prévention grâce aux Data et à l'intelligence artificielle, dans l'activité scientifique après les années 1950

L'objectif de cette partie est de voir s'il est possible, au travers de recherches bibliographiques, de dégager des tendances dans l'utilisation récente des techniques d'intelligence artificielle et/ou d'utilisation de données massives dans le cadre de la prévention.

2.2.1 Méthodologie : construction des requêtes

2.2.1.1 Outils utilisés pour construire la revue des publications scientifiques

Les objectifs de cette revue sont, dans le cadre du mémoire, de décrire rétrospectivement et macroscopiquement l'activité de publication scientifique liée aux sujets de la prévention (primaire, secondaire et tertiaire), utilisant les données informatisées en général, et les données des dossiers médicaux des patients en particulier.

La base de données sur laquelle la recherche s'est menée est Pubmed (14). L'outil public français de construction de requêtes bibliographiques médicales, conçu et développé par le CHU de Rouen, a ponctuellement été utilisé pour affiner les requêtes (15).

Toutes les langues de publication présentes dans le stock de publication ont été retenues.

La période couverte est celle du fonds auquel Pubmed donne accès.

Les directives proposées par le cadre méthodologique Prisma (16) lorsqu'applicables, ont été suivies.

2.2.1.2 Les mots clés recherchés et construction de la requête

La première requête suivante a été construite progressivement :

((Computing Methodologies[MeSH Terms]) or (primary prevention[MeSH Terms]) or (secondary prevention[MeSH Terms]) or (tertiary prevention[MeSH Terms]) or (Infection control[MeSH Terms]) or (Preventive medicine[MeSH Terms])) **and** ((data mining[MeSH Terms]) or (big data[MeSH Terms]) or (artificial intelligence[MeSH Terms]) or (Decision support systems, Clinical[MeSH Terms]))

La partie bleue correspond aux termes MESH exprimant la prévention.

La partie verte correspond aux termes MESH représentant l'analyse de données.

La requête renvoie plus de 92.000 résultats en près de 70 ans, la très grande majorité dans les 20 dernières années.

Ce qui donne l'évolution quantitative dans le temps représentée dans la figure suivante (l'évolution qualitative devrait impliquer d'autres métriques comme l'Impact Factor, qui n'est pas disponible dans Pubmed).

Figure 4 Nombre d'articles pour la requête

2.2.1.3 Discussion

Dans les limites exposées, cela conduit à penser que le sujet conjoint de l'utilisation des données informatisées pour les différents types de prévention est un sujet en forte croissance, en nombre de publications.

Une analyse plus fine avec l'outil VOSVIEWER (17) permet d'isoler et de représenter les mots clés fréquemment utilisés dans cette sélection d'articles.

Afin de repérer les principales spécialités, les 92 termes présents dans au moins 50 articles (seuil arbitrairement choisi) ont été retenus (sur 50 376 mots clés au total).

Les domaines médicaux ou médicotechniques saillants isolés par cette méthode sont représentés dans le graphe suivant :

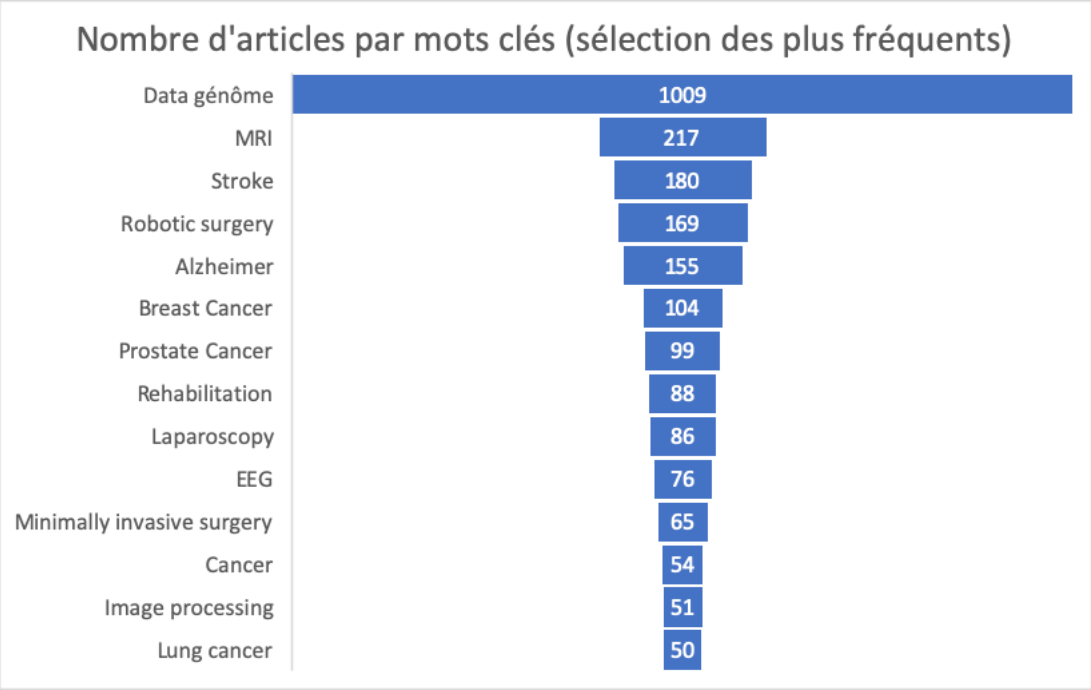

Figure 5 Mots clés "auteur" les plus fréquents dans le résultat de la requête

La représentation en graphe des cooccurrences de ces termes est l'objet des figures suivantes :

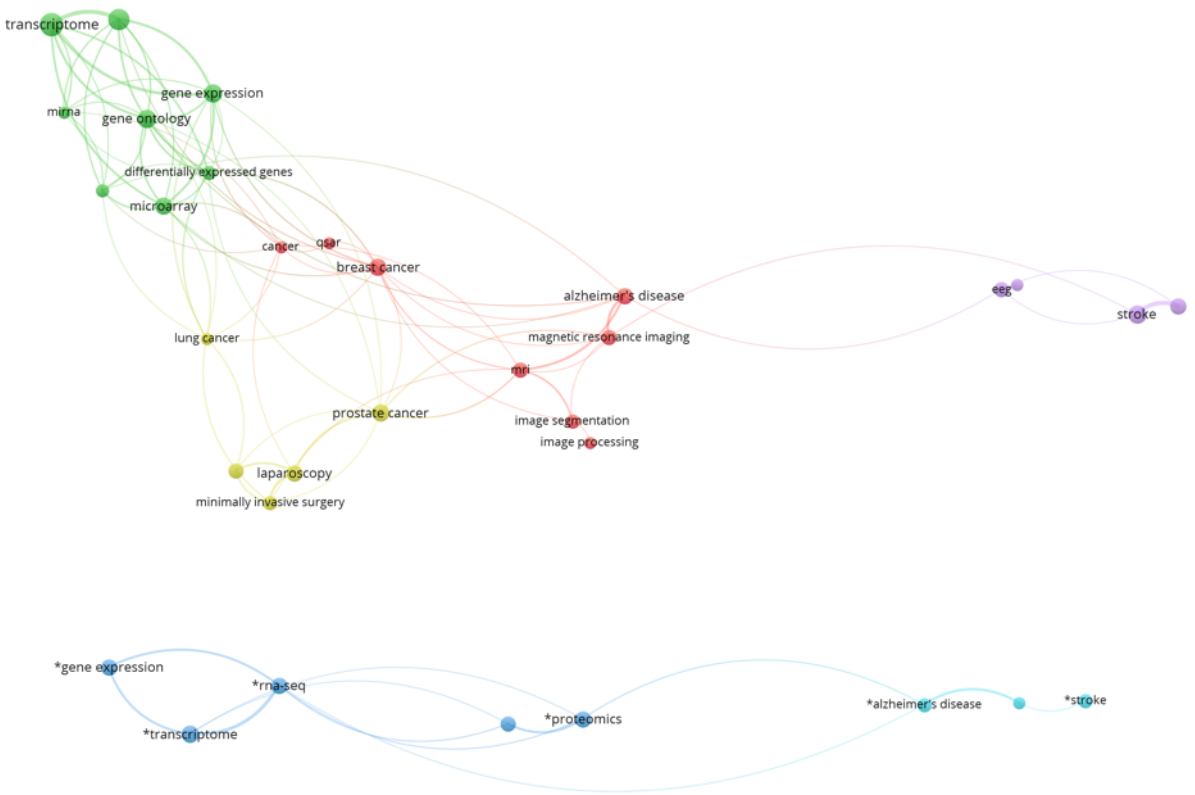

Figure 6 : représentation dans vosviewer des mots clés médicaux "auteur" principaux dans la sélection d'articles

Ce comptage et cette représentation montrent bien la spécificité des travaux sur les données du génome, qui forment un ensemble distinct de clusters. Ces travaux sur les données du génome sont très nombreux sur la période : la prévention utilisant les techniques d'analyse de données (dont l'intelligence artificielle) sur l'analyse des données massives issues du décodage du génome semblent via ce prisme avoir été, et être encore, une activité très importante dans la recherche médicale. A noter aussi la présence de mots clés qui suggèrent une forte activité d'intelligence artificielle pour la prévention en imagerie, en chirurgie assistée par robot.

2.2.1.4 Les limites de cette revue scientifique

Ce comptage de la production scientifique dans Pubmed présente plusieurs limites.

D'abord, celles induites par le fonds principalement adressé, Pubmed. Ce fonds ne dispose pas de toutes les revues susceptibles de traiter du sujet sur un abord scientifique.

Ensuite, les mots clés utilisés par les auteurs de chaque publication, à un moment donné, ne représentent pas forcément le sujet recherché. En effet, les mots clés choisis dans le

thésaurus de référence MESH, adoptés par Pubmed ne correspondent pas aux termes tels qu'on les exprime aujourd'hui. Par exemple, nous pouvons remarquer, d'après les graphiques représentant le nombre d'articles parus avec les termes *Big Data Healthcare* dans Pubmed que cette notion de Big Data a été utilisée de manière plus importante à partir de 2013 où 85 articles parlent de ces notions et en qu'actuellement, 315 articles ont été parus. Si on utilise les termes *Artificial Intelligence Healthcare*, nous pouvons constater que l'utilisation de ces termes s'est faite plus progressivement et tendent à stagner depuis 2010 puisque 262 et 379 articles ont respectivement été publiés en 2013 et 2019. Enfin, nous avons également pu constater que des termes comme *Data Reuse* ne sont que très peu utilisés puisque seulement 18 articles sont ressortis de cette requête.

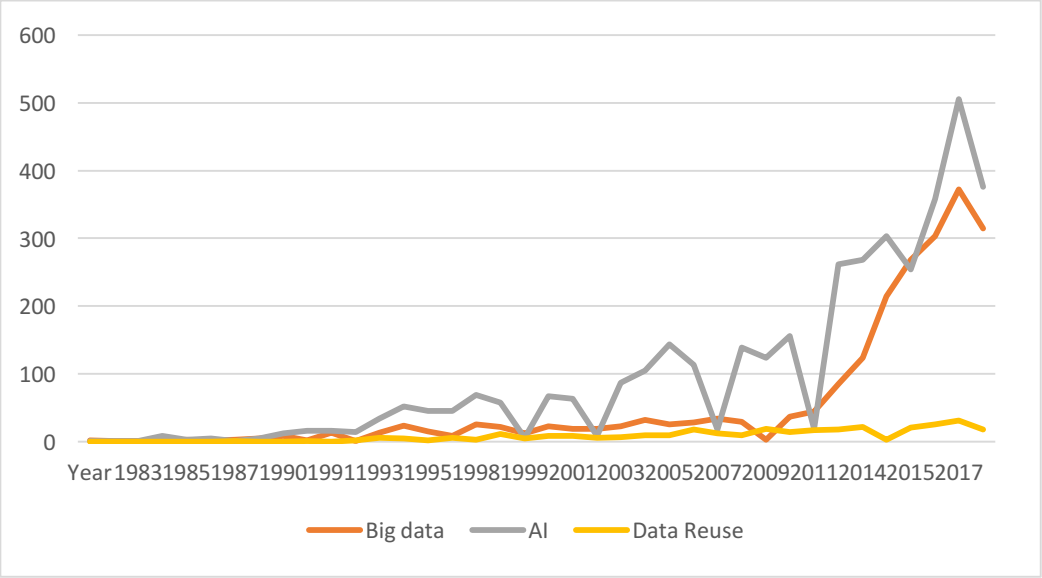

Figure 7 Utilisation des mots clés

Enfin, le volume d'articles repérés ne permet pas la lecture du texte intégral de tous ces articles. Les analyses de cette revue reprises dans ce mémoire s'appuient essentiellement sur les métadonnées de ces articles, essentiellement le titre, les mots clés et les résumés.

2.3 Les Big Data : quelles données candidates dans le secteur santé, pour la prévention ?

L'institutionnalisation de la prévention a pris un siècle en France, au travers de diverses politiques publiques. Les organismes et instituts au service de ces politiques de prévention peuvent-elles utiliser les données disponibles ? Quelles-sont-elles ?

Les Big Data ou données massives sont aujourd'hui de plus analysées et cela dans tous les domaines d'activités. Nous allons voir quelles données sont disponibles et quels peuvent-être les freins quant à leur utilisation dans le domaine de la santé. Nous poursuivrons ensuite avec une partie concernant la réglementation de ces données, en France et en Europe puis nous essaierons de comprendre quelle est la relation entre les français et la protection de leurs données et discuterons enfin de la nécessité de certifier l'intelligence artificielle.

2.3.1 Les data disponibles

Les big data dans les soins de santé peuvent être utilisées pour le diagnostic, le pronostic, la prévention et la prévision de maladies, ainsi que pour concevoir un programme de gestion de la santé et des maladies sur mesure basé sur le mode de vie de chaque individu. L'application des big data dans le secteur de la santé comprend 3 principaux sous-domaines : les projets de génome humain pour une médecine personnalisée, l'analyse des informations cliniques pour la prise de décision clinique et du système de facturation hospitalière pour réaliser des évaluations économiques de la santé et enfin, utiliser les données mobiles comme indicateurs de santé publique (9).

2.3.1.1 Quels types d'informations sont disponibles ?

Les données médicales peuvent être obtenues sous de nombreux aspects : données génétiques, données de coût, images radiographiques, résultats de tests de laboratoire, et à partir de médias sociaux ou de logiciels mobiles. En 2020, le volume de données devrait dépasser les 35 zettaoctets. Cet important volume de données posera le problème du stockage et de la récupération des données.

Toutes ces informations collectées peuvent-être catégorisées en deux types de données distincts : la donnée structurée et non structurée. La donnée structurée sera constituée d'informations (mots, signes, chiffres etc.) contrôlées par des référentiels et présentés dans des cases qui permettent leur interprétation et traitement par des machines. La donnée non structurée représente toutes les informations qui ne sont pas organisées en base de données comme les documents, les messages, les images ou vidéos etc. (18). Les méthodes utilisées pour analyser ces deux types d'informations seront différentes.

L'hétérogénéité de ces données est l'une des problématiques à adresser : le dossier médical informatisé complet d'un patient est composé de données de structures très différentes. Les compétences à mobiliser pour les agréger, puis les analyser, sont multiples et spécifiques : traitement de signal, probabilités/statistiques, biostatistiques (en particulier pour l'évaluation), informatique (différents champs : réseaux de neurones / machine learning / deep learning, optimisation multi-objectifs, etc.). La figure suivante positionne sur un quadrant les principales données disponibles à l'hôpital.

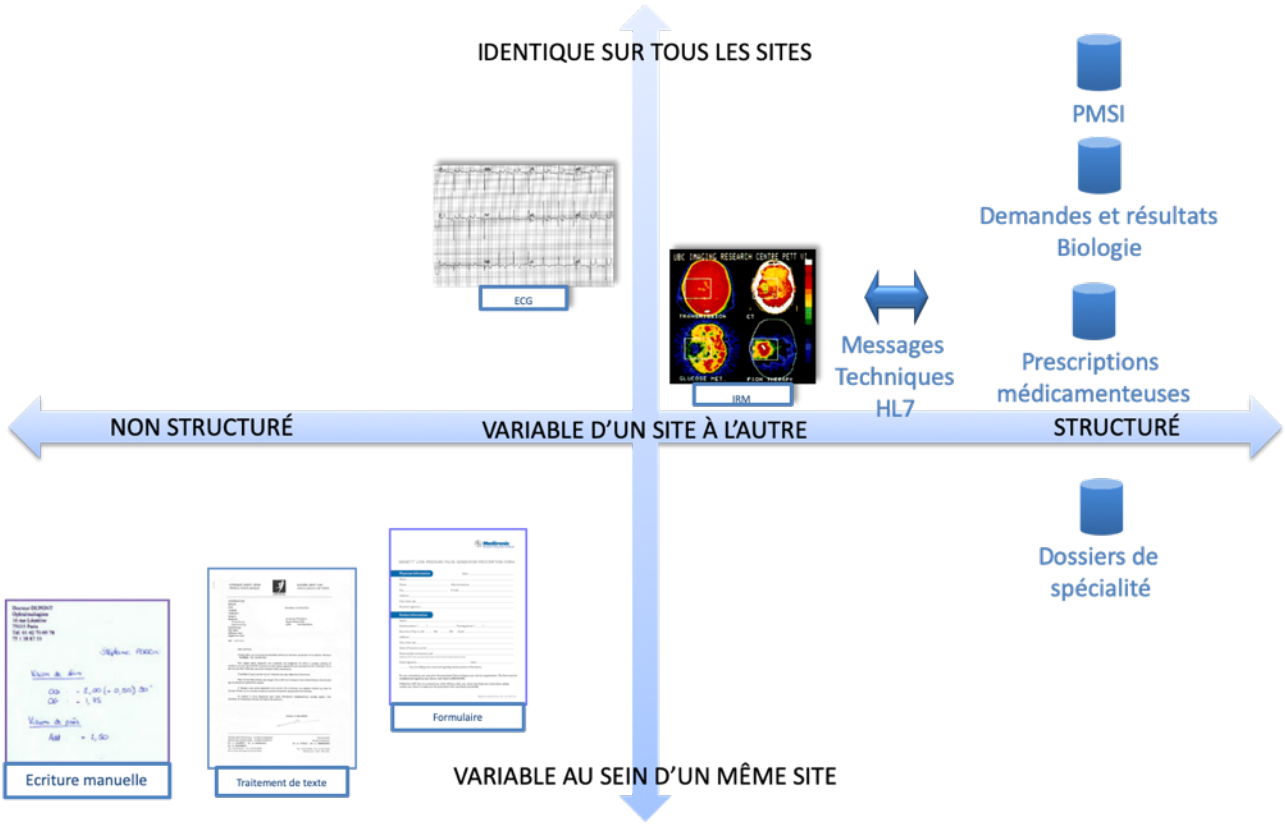

Figure 8 Représentation synthétique de la diversité des données disponibles à l'hôpital

Ces données ont-elles toutes la même véracité ? Peut-on leur accorder la même confiance ? Il est nécessaire de distinguer pour cela les données issues d'examens qui auront une forte véracité et une forte fiabilité mais dont la quantité d'informations par individu sera restreinte avec les données déduites, indirectes ou extrapolées, provenant par exemple des outils de quantified self comme les montres connectées. En effet, pour que les modèles d'intelligence

artificielle soient les plus représentatifs de la réalité, les données d'entrée doivent être en qualité et en quantité correcte. La sincérité et la rigueur des patients utilisant ces outils pour suivre leur état de santé sont donc une nécessité.

D'autres sources de données peuvent également être prises en compte en santé publique. En France, on dénombre environ 260 bases de données de santé publiques et plus de 500 bases de données médico-économiques, cohortes et études en cours (19). Ces bases offrent un accès à des données objectives et exhaustives, à l'échelle populationnelle. Parmi ces bases, les plus couramment utilisées sont celles de la SNIIRAM (Système National d'Informations Inter-régimes de l'Assurance Maladie) dans laquelle sont enregistrés tous les remboursements de l'Assurance Maladie et cela tout au long de la vie du cotisant mais également celles de l'ATIH (Agence Technique de l'Information sur l'Hospitalisation) qui permettent, grâce au PMSI (Programme de Médicalisation des Systèmes d'Informations) de collecter des informations médico-économiques sur les activités hospitalières d'un territoire (20).

Exemple de « nouvelle » base de données utilisée pour la prévention : les réseaux sociaux et moteurs de recherche peuvent aujourd'hui être utilisés pour prédire l'avancement d'une épidémie. En effet, l'avantage considérable de ces données provient de leur très rapide disponibilité, quasiment en temps réel. Elles permettraient d'estimer plus précocement la dynamique d'une épidémie par rapport au système de veille épidémiologique actuel.

Ainsi, Google Flu, filiale de Google a été créée en 2012 suite à l'épidémie de grippe H1N1. En effet, en 2009, grâce aux recherches effectuées par les internautes et à leur localisation, Google a été capable de prédire les risques et d'établir un modèle de propagation de cette épidémie (21). Toutefois, Google Flu a fermé en 2015 suite à des mauvaises prédictions. Aujourd'hui, Google continue de collecter toutes ces données mais a choisi de travailler en collaboration avec les scientifiques et les centres pour la prévention et le contrôle des maladies, qui constituent le système de veille sanitaire classique.

Les réseaux sociaux permettent également d'évaluer et de contrôler l'ampleur d'une épidémie et donc de déterminer les actions de prévention à entreprendre. Le principe consiste à associer les requêtes construites par les usagers à un territoire particulier.

Une étude parue en 2012 (22) relève que Twitter a été utilisé en 2011 pour créer un modèle épidémiologique lors de l'épidémie de Choléra qui avait touché Haïti. Les scientifiques ont

évalué la corrélation entre le nombre de reportages sur le choléra diffusés dans les médias, les publications sur Twitter et les cas relevés par le gouvernement lors des 100 premiers jours de cette épidémie. Les scientifiques ont estimé que les tendances du volume de sources informelles étaient en corrélation significative avec les données officielles et étaient disponibles jusqu'à deux semaines plus tôt. Ils en ont conclu que les données informelles peuvent être utilisées en complément des données officielles en cas d'épidémie afin d'obtenir des estimations fiables de la dynamique de la maladie.

2.3.1.2 Comment les utiliser ?

Compte tenu du volume de données important et de la variété de qualité des données, il est nécessaire de prendre en compte tous les aspects du jeu de données, notamment la façon dont les données sont collectées, extraites, intégrées, interprétées ainsi que la sélection de méthodes statistiques, lors du traitement et de l'analyse.

L'analyse des données est plus simple sur de la donnée structurée. Cependant, beaucoup des données existantes sont non structurées et la plupart des données récemment collectées ne le sont pas. Les données issues des dossiers médicaux électroniques peuvent être structurées si cela a été prévu. Les informations génétiques, quant à elles, sont intrinsèquement structurées. Par contre, les parties de dossiers médicaux rédigées à la main ne sont pas des données structurées (transmissions ciblées des infirmières en texte, libre, notes cliniques, commentaires, et surtout les courriers). De même que la plupart des données sur les comportements de santé personnels générées à partir des médias sociaux et des appareils mobiles ne sont pas structurées.

Il existe un moyen de faciliter l'analyse des données non structurées, comme les courriers : il est nécessaire pour cela bénéficier de la coopération du personnel médical et soignant, en incitant à une rédaction structurée de courriers.

2.3.2 Quels sont les freins actuels quant à l'utilisation des big data de santé, pour la prévention ?

2.3.2.1 Une infrastructure technique particulière est nécessaire

Dans les systèmes de santé, les données doivent être traitées en temps réel. Le système doit amasser les données en tout temps et fournir des résultats de recherche à tout moment. Pour utiliser et analyser les données massives à l'échelle d'un établissement de santé et sans modifier l'organisation actuelle, très décentralisée, des données de santé hospitalières, les hôpitaux ou autres centres devront se munir de capacités « serveurs » très importantes. Cela peut représenter un coût non négligeable pour les établissements de soins qui disposent par ailleurs d'un système d'information complexe, et d'un budget limité.

2.3.2.2 Des compétences particulières, et rares, sont requises

Comme évoqué précédemment, outre la complexité technique d'amasser des données, l'analyse des données massives nécessite des ressources humaines bien particulières. En effet, avant de vouloir construire des modèles à partir des données, il est nécessaire de choisir les données les plus probantes, les rendre exploitable par les outils informatiques, les agréger, définir des classifier permettant de comparer les données entres elles etc. Pour réaliser ces outils basés sur les données massives, plusieurs corps de métiers interviendront : les data-analysts, data-scientists, data-manger, data protection officers etc. Aujourd'hui encore, ces ressources humaines sont rares.

2.3.2.3 Autres freins : la faible incitation des professionnels à l'utilisation des outils ; la faible utilisabilité des systèmes d'information en santé

Il existe également de nombreux défis pratiques quant à l'utilisation des données personnelles que nous ne pouvons ignorer.

Tout d'abord, le personnel médical et soignant ne semble pas motivé pour collecter des d'avantage de données électroniques, sans incitations. L'une des incitations pourrait être la promesse de bénéficier de ces données collectées, en priorité, pour des usages médicaux et soignants, mais aussi pour l'amélioration des pratiques professionnelles.

Autre défi à relever : celui de l'ergonomie, de l'utilisabilité. Les outils de saisie présents à l'hôpital sont réputés comme étant peu ergonomiques. Leur principale vocation historique relève de la gestion de l'hôtellerie (gestion des lits et chambres), des logistiques liées au patient (brancardage, repas, prescriptions, linge, etc.), et du médico-administratif

(facturation, PMSI/T2A). Des outils ergonomiques devraient permettre au professionnel concerné de saisir une seule fois, au plus près dans le temps et dans l'espace de la production de l'information, l'information utile. L'ergonomie devrait pouvoir réduire, comme pour les outils informatiques grand public, la courbe d'apprentissage.

Deuxièmement, les soins de santé sont une industrie énorme avec des relations complexes, structurées et enchevêtrées. Les hôpitaux, les centres de santé de ville, les compagnies d'assurance, l'Agence nationale de sécurité du médicament et des produits de santé, le centre de surveillance des dispositifs médicaux, les ensembles de données de surveillance de la santé publique sont comme des îlots de données isolées qui ne peuvent pas être facilement reliés (23).

Troisièmement, l'absence de système normalisé de collecte d'informations accroît la difficulté de partage et de collecte des données. De plus, la pénurie de ressources humaines en technologies de l'information de la santé entrave le déroulement rapide du processus d'informatisation(24).

2.3.2.4 La réglementation française

Bien que les individus soient de plus en plus enclins à partager leurs données personnelles, la collecte de ce type d'informations soulève de nombreux problèmes éthiques et réglementaires. Il n'est pas aisé de contrôler les utilisations commerciales des données de santé publique. Compte tenu de la complexité du domaine de la Santé, la protection des individus et des populations doit mettre l'accent sur l'utilisation appropriée, l'évaluation et la minimisation des risques. L'anonymisation des données doit être robuste, surveillée et appliquée. De plus, la sécurité des données se doit d'être particulièrement élevée. En effet, il est fondamental de préserver la confidentialité et la sécurité des données tout en en tirant parti (25). Pour cela, la France, en traduisant les directives européennes en droit français, en les anticipant, ou en les durcissant, a mis en place un cadre strict.

Les données de santé sont des données à caractère personnel particulières car considérées comme sensibles. Elles font, à ce titre, l'objet d'une protection particulière par les textes (règlement européen sur la protection des données personnelles, loi Informatique et

Libertés, code de santé publique etc.) afin de garantir le respect de la vie privée des personnes (26).

La CNIL (Commission Nationale de l'Informatique et des Libertés) est le régulateur des données personnelles. Elle a plusieurs missions dont celles d'informer les entreprises et les particuliers sur leurs droits et devoirs, d'accompagner la mise en conformité des entreprises, notamment face aux exigences de la RGPD et de contrôler et sanctionner les organismes s'ils ne respectent pas les prérogatives établies.

Le RGPD est le règlement européen sur la protection des données. Il est entré en application le 25 mai 2018 et est fondé sur la responsabilisation des acteurs de la santé et décrit le rôle donné au consentement des personnes.

Toute opération portant sur des données de santé est définie comme un traitement de données de santé. Le RGPD l'interdit à condition d'obtenir le consentement exprès des personnes concernées ou que le traitement fasse partie des exceptions prévues par la loi.
Le traitement d'une donnée de santé est légal dès que la personne concernée a donné son consentement exprès, libre et univoque. Elle devra en outre être bien informée des objectifs du traitement. Le RGPD étant un règlement européen, les Etats membres peuvent toutefois mettre en place des règles plus strictes, comme cela est le cas en France dans le cadre de la loi Informatique et Libertés modifiée en juin 2018.

Le traitement des données de santé peut être légal malgré l'absence du consentement préalable de la personne concernée lorsque le traitement est réalisé pour atteindre certaines finalités ou en cas d'obligation légales :

- Prévention de la santé publique ;
- Préservation des intérêts vitaux de la personne concernée ;
- Appréciation médicale : médecine préventive, soins, diagnostics ;
- Médecine du travail ou encore pour déterminer le pourcentage légal d'emploi de personnes avec handicap ;
- Gestion des systèmes et services de santé ou de la protection sociale.(27)

Pour les entreprises réalisant du traitement de données à visée préventive, le RGDP exige quelques prérogatives supplémentaires comme désigner une personne déléguée à la protection des données dont le rôle majeur sera de veiller au respect du RGPD, de tenir un

registre des traitements comprenant notamment les finalités du traitement, la durée de conservation des données etc. et une analyse des risques et des impacts. En effet, tous les traitements des données de santé peuvent avoir des impacts non négligeables sur la vie de la personne concernée. Cette analyse doit contenir certains éléments comme les mesures mises en place pour protéger les données (anonymisation, chiffrage).

2.3.3 Les Français sont-ils opposés à l'utilisation de leurs données personnelles

Comme nous venons de le constater, la législation européenne et la législation française sont assez strictes vis-à-vis des données personnelles et de leur traitement. Il semble intéressant à ce stade de connaître l'avis des français concernant la protection et l'utilisation de leurs données de santé à des fins de santé publique.

Selon un sondage Odoxa réalisé en 2017 (28) **72% des français estiment que leurs données de santé sont utiles pour faire avancer la recherche et la santé pour tous**. Aussi, **83% accepteraient de partager leurs données de santé si cela répondait à certains objectifs dont l'amélioration des diagnostics et des traitements médicaux ainsi que pour accélérer la recherche médicale**. Toutefois, 65% d'entre nous n'accordent pas une confiance aveugle et attendent que certaines conditions soient mises en place par les institutions françaises concernant la protection de leurs données de santé. Seuls 6% de la population transmettraient leurs données de santé sans aucune condition. Nous pouvons ainsi constater que la législation mise en place par l'Europe et l'Etat français correspondent aux besoins de la population en termes de protection des données de santé.

De plus, 60% des Français considèrent l'exploitation des Big Data comme positive et pensent qu'il vaut mieux être soigné dans un centre qui a accès à des bases de données de cas patients du monde entier plutôt qu'à une équipe médicale dont l'expertise se limite à sa région (29).

Un sondage BVA de 2018 montre que 21% de la population française utilise le smartphone comme assistant de suivi de santé. 77% jugent les applis santé utiles et 15% les utilisent régulièrement. À noter que 61% des Français sont rassurés si c'est un professionnel de santé qui leur conseille l'appli ou le site à utiliser (29).

51% des Français estiment que les objets connectés peuvent être utiles pour mieux prendre en main leur santé. Aussi, 14% en possèdent ou en utilisent, et 70% d'entre eux l'utilisent fréquemment tandis que 27% partagent déjà leurs données avec leur professionnel de santé. Un sondage Mazars nous apprend également que 69% des français se déclarent favorable au dossier médical centralisé et 53% sont prêts à transmettre leurs données collectées via des objets connectés à des professionnels de santé. Ainsi, un français sur deux est favorable à l'utilisation de logiciels d'intelligence artificielle capables d'interpréter les données médicales mais aussi celle des programmes d'intelligence prédictive, qui peuvent formaliser des recommandations (30).

2.3.4 Comment gagner la confiance des professionnels de santé dans les outils basés Data et intelligence artificielle, pour la prévention ?

Dans le domaine de la santé, les médicaments et dispositifs médicaux nécessitent une certification pour pouvoir être utilisés ou vendus. Les logiciels contenant de l'intelligence artificielle pour traiter les données de santé peuvent être utilisés pour faire de la prévention voire du diagnostic thérapeutique. Dans ce cas, la certification est-elle obligatoire ? Comment et qui attestent de la qualité et de la sécurité que ces logiciels ?

Dans un premier temps, nous nous rappellerons en quoi consiste la mise sur le marché d'un médicament et d'un dispositif médical puis nous discuterons du cas des logiciels utilisant des mécanismes dits « d'intelligence artificielle ».

2.3.4.1 Mise sur le marché des médicaments et des dispositifs médicaux.

La mise sur le marché d'un médicament est extrêmement contrôlée et règlementée. Le médicament doit posséder une autorisation de mise sur le marché (AMM) délivrée par une autorité compétente. Cette AMM garantit que le médicament possède un profil de qualité, de sécurité et d'efficacité satisfaisant. Elle est conditionnée par la démonstration du rapport bénéfice/risque dans une indication thérapeutique donnée et pour une population donnée. Pour certifier les médicaments, il faut d'abord qu'ils soient testés sur les animaux,

puis sur homme afin d'identifier les propriétés pharmacocinétiques et pharmacodynamiques (31).

Pour les dispositifs médicaux, la règlementation est différente. Le marquage CE (Conformité Européenne) est obligatoire et nécessite d'avoir au préalable démontré la conformité du dispositif médical. Il est délivré par l'ANSM (Agence Nationale de Sécurité du Médicament et des Produits de santé). Il dépendra de la classe du dispositif médical et consistera à examiner le dossier de conception du dispositif médical fourni par le fabricant ou par le biais d'essais du produit en lui-même. Cela permet d'assurer la conformité du dispositif médical aux exigences essentielles en accord avec des normes techniques et intègre également l'évaluation des données cliniques (32).

2.3.4.2 Les algorithmes et l'intelligence artificielle, besoin d'être certifiés ?

Aujourd'hui, que l'on développe un algorithme ou un logiciel utilisant de l'intelligence artificielle, en France ou aux Etats-Unis, aucune loi ne demande réellement qu'ils soient certifiés, malgré plusieurs avancées et annonces dans ce domaine. Toutefois, les professionnels du secteur s'attendent à ce que ce type de certification voit le jour, dans un futur plus ou moins proche.

2.3.4.2.1 Dans les autres pays : l'avis d'ARTERYS, un acteur de l'analyse d'image basée intelligence artificielle, présente dans plusieurs pays

Afin d'obtenir un regard connaissant à la fois le contexte français et d'autres contextes, une interview a été faite dans le cadre de ce mémoire : l'interview de Maya Khalifé, Product Manager chez Arterys (33), société spécialisée dans la lecture automatisée d'images : « *Pour l'instant, que ce soit aux US ou en France, rien d'officiel n'est sorti, par contre, il semblerait que les méthodes d'analyse de la donnée médicale aient besoin d'être certifié. Aux US, il devrait s'agit d'une certification par la FDA et en France du marquage CE.* »

En confirmation de cette appréciation, le règlement européen 2017/745 relatif aux dispositifs médicaux indique que, si l'intelligence artificielle est utilisée à des fins de diagnostic, de prévention, de contrôle de maladies, de blessures, de prédiction de maladie etc. alors elle devra être considérée comme un dispositif médical et dans ce cas devrait bénéficier de la même obligation de certification. La plupart des logiciels d'intelligence artificielle relèveront de la classe Iia (34).

La certification des logiciels ne contenant pas d'intelligence artificielle se fait en fonction de trois classes, A, B et C. Si un logiciel connaît un dysfonctionnement (un « bug »), et qu'il est sans danger pour le patient, alors il est classé A ; C si ce dysfonctionnement a des conséquences graves. Pour faire certifier les logiciels de classe A, on doit définir le résultat obtenu alors que pour ceux de classe C, il faut en plus expliquer comment chaque composant du logiciel fonctionne. Pour les logiciels comportant de l'intelligence artificielle, cela ne sera pas toujours possible à démontrer puisque, par nature, les modèles issus de l'apprentissage ne sont pas explicables, leurs motivations des décisions proposées sont illisibles pour l'humain et sont considérés comme des boîtes noires 22/09/2019 22:34:00.

Pour que les logiciels d'intelligence artificielle soient validés, il semblerait qu'il faille, dans un premier temps, démontrer la qualité des données utilisées pour l'apprentissage en décrivant la quantité de ces données, leur représentativité, les biais liés à la collecte, à l'analyse, au choix des données et que le seuil d'erreur soit minime. Dans un second temps, il faudra tester l'algorithme avec des données de même type que celles décrites dans les documents de certification et que le taux d'erreur soit toujours tolérable. Cela semble difficile puisque les données, donc les modèles « appris » à partir des données, sont très souvent dépendants de leur site d'origine.

De plus, il faut également prendre en compte que toutes nouvelles mises à jour devront être certifiées. Cela va également à l'encontre des demandes du secteur de la santé puisque l'on travaille avec des données de routine, en temps réel et que les ajustements provenant du « terrain » sont réguliers. Autrement dit, à chaque fois qu'un fabriquant souhaitera faire progresser l'intelligence artificielle, il devra refaire les spécifications, vérifications et validations nécessaires pour certifier une nouvelle version de son logiciel.

Aujourd'hui et face à ce constat, il semblerait que deux alternatives se dessinent. La première consisterait à certifier le fabriquant du logiciel et non le logiciel lui-même, c'est-à-dire certifier sa chaîne de production de modèles (son « atelier »), plutôt que ce que produit l'atelier. La seconde alternative qui émerge permettrait d'apprendre simultanément sur plusieurs sites, tout en respectant les contraintes de sécurité. Cet apprentissage « en une passe » utilise les techniques dites d'apprentissage décentralisé ou *Decentralized Machine Learning*.

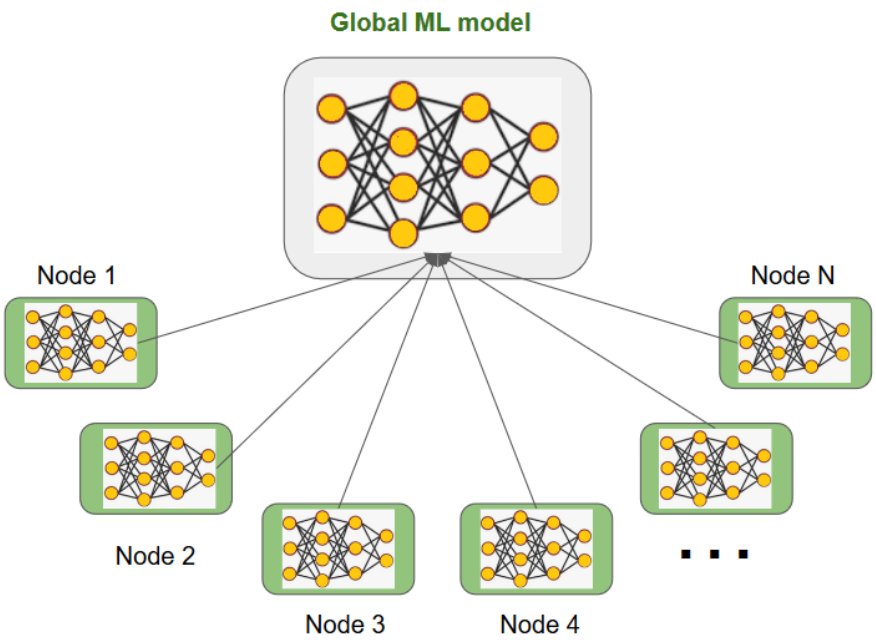

Figure 9 Apprentissage décentralisé (25)

Ces techniques d'intelligence artificielle permettraient d'apprendre sur plusieurs jeux de données distincts (mais de même structure), sans à aucun moment devoir réunir ces jeux de données en un seul endroit. Le modèle issu de cet apprentissage décentralisé pourrait être certifié, et pourtant concerner plusieurs sites.

En d'autres termes, si dans le logiciel est initialement prévu l'apprentissage sur des données issues d'établissements différents alors cet algorithme sera certifié.

Il reste à vérifier que l'écart d'efficacité entre le modèle centralisé résultant de ces méthodes décentralisées, et ce qu'aurait été la précision d'un modèle sur chacun des sites, est acceptable pour les usages ciblés.

Des projets de recherche sont menés pour évaluer ces approches, notamment le projet PAMELA (35).

Cette seconde alternative ne semble à l'heure actuelle accessible qu'aux entités disposant de gros volumes de données issus d'un nombre significatif de sites. C'est le cas de quelques laboratoires de recherche institutionnels, qui au travers d'entrepôts multi-sites apprendre sur de gros volumes de données extraits de plusieurs centres hospitaliers (comme les projets menés par HUGO (36)).

Pour la plupart des entreprises privées, chaque nouvelle vente nécessitera une nouvelle certification puisque les données d'entrées sont, le plus souvent, territoires dépendants,

donc *probablement* différentes des données précédentes. De plus, cela risque de compliquer les ventes de ce type de technologies puisque les coûts seront probablement impactés et que les délais de mise en œuvre du projet seront augmentés.

Quelques entreprises privées semblent pouvoir s'affranchir de ces contraintes : d'une part, les acteurs mondiaux comme Google, Amazon, Facebook, IBM, Microsoft (les « GAFAM ») ont la capacité à investir suffisante, ainsi que de réels moyens d'accès aux données de santé, au travers de leurs offres, comme en France celles s'appuyant sur leur agrément « hébergeur de données de santé ». Leurs activités leur permettent de disposer de gros volumes de données en un point physique unique.

D'autre part, des acteurs comme OPTUM ont constitué des jeux de données en traitant directement avec des états, ou de très gros groupes d'hôpitaux ou d'assurances santé privées. OPTUM vend des jeux de données dits « de vie réelle », censément déidentifiés, tout en restant représentatifs.

2.4 En conclusion de cette partie

Dans cette première partie de l'exposé, nous avons cherché à définir les deux notions clés du sujet, et à les replacer dans une dynamique historique : la prévention et les big data (liées à l'intelligence artificielle). Ces notions sont polymorphes, difficiles à définir, et en évolution. Nous avons suggéré cependant que les techniques récentes d'intelligence artificielle sur des données massives pourraient, sous certaines conditions, être utilisées dans le cadre de la prévention. Cette utilisation se heurterait à quelques limites, rencontre plusieurs freins, qui ont également été abordés.

Mais que l'utilisation des big data via l'intelligence artificielle pour la prévention soit possible, ne signifie pas forcément qu'elle soit toujours faisable, notamment d'un point de vue économique. Dans la seconde partie de ce mémoire, nous nous intéresserons à l'évolution de l'économie de la santé d'abord en France, puis en prenant quelques exemples d'autres pays, pour se rendre compte que la personnalisation de la prévention grâce aux big data est une tendance mondiale, mais qu'elle peut se heurter, en fonction des spécificités de chaque pays, à des résistances. La viabilité économique d'un système de prévention

individuelle, basé data, étant l'une d'entre elles : au-delà d'expérimentations subventionnées ponctuellement, quel modèle économique pour la prévention personnalisée basée Data ?

III. Quels impacts pourraient avoir ces outils de prévention sur l'économie globale de la santé ?

Avant de s'intéresser à la personnalisation de la prévention mondiale, il est important de comprendre le fonctionnement de la prévention en France, sur un plan économique. Ainsi, nous analyserons dans une première partie le développement de la prévention moderne puis son mode de financement et terminerons avec la place du numérique au sein du système de santé.

3.1 La prévention en France

3.1.1 Mise en place de la prévention en France

Comme nous l'avons vu dans une partie précédente, la prévention au sens « actions publiques de prévention » est ancienne. Toutefois, le modèle récent qui s'est imposé repose sur le modèle d'Ehrlich & Becker : *Analyse de la prévention* datant de 1972. Il a permis de mettre en relation deux types de prévention, autoprotection et auto-assurance avec l'achat d'assurance. Les auteurs expliquent que l'autoprotection vise à réduire la probabilité du sinistre, c'est-à-dire l'apparition d'une maladie. Elle renvoie à la prévention primaire. L'auto-assurance et la souscription d'un contrat d'assurance peuvent s'apparenter à une prévention de type secondaire dont l'objectif est de limiter l'extension du sinistre en cas de survenue de la maladie (37).

Ce modèle mixte « prise en charge globale financée par la collectivité et complément de responsabilisation individuelle qui se traduit par une assurance facultative souscrite » est toujours d'actualité.

Plus récemment, la loi du 4 mars 2002 place la prévention au premier plan dans la politique de santé. Cette loi introduit une nouvelle notion dans le financement de la prévention : elle est énoncée comme un risque qui peut être couvert par une assurance. Depuis cette loi, c'est maintenant à l'assurance maladie de prendre en charge le financement des actions de prévention réalisées dans le cadre des nouveaux programmes de prévention, au même titre que les actes médicaux (38).

3.1.2 Financement de la prévention

En 2017, en France, la prévention s'est élevée à 6,1 milliards d'euros, en hausse de 4,2% par rapport à 2016. On peut découper ce budget en deux parties, la prévention individuelle qui représente 3,8 milliards d'euros et la prévention collective qui atteint les 2,3 milliards d'euros. Ces enveloppes ont augmenté respectivement de 2,9% et 6,3% par rapport à l'année précédente. Autrement dit, la prévention institutionnelle représente 91 euros par habitant, soit 2,3% des dépenses courantes de santé.

La prévention primaire représente à elle seule plus de la moitié du budget total dont la plus grande partie revient à la médecine du travail (55%). La prévention secondaire concerne 11,4% de l'enveloppe. La plus grande part de la prévention collective revient à la prévention à visée environnementale avec 31,3% de l'enveloppe totale institutionnelle. Elle concerne l'hygiène du milieu, la lutte contre la pollution, la prévention des accidents de travail etc. La prévention collective à visée comportementale remporte 6,7% de la somme consacrée à la prévention. Elle a pour but d'éduquer la population à différentes thématiques comme par exemple le développement du programme national nutrition santé.

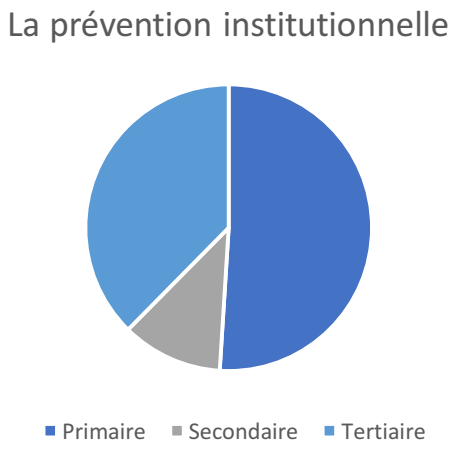

Figure 10 Répartition du financement de la prévention institutionnelle

Le financement de ces types de prévention provient d'une part de l'état et des collectivités locales à hauteur de 38%, de la sécurité sociale à raison de 15% et le reste, soit 27%, du secteur privé.

Si l'Etat et les collectivités locales, la sécurité sociale et le secteur privé cotisent tous pour la prévention institutionnelle, la répartition est différente en fonction de la nature de la prévention. En effet, le secteur privé financera la prévention primaire à hauteur de 51%, notamment par les crédits accordés par les entreprises à la médecine du travail. La sécurité sociale finance chaque type de prévention mais surtout la prévention secondaire alors que l'Etat et les collectivités locales participeront essentiellement au financement de la prévention collective en étant les financeurs à hauteur de plus de 82%. (39) (40)

3.1.3 La prévention et le numérique, au cœur du système de santé

Depuis quelques années, nous avons pu constater que le numérique est de plus en plus présent dans les programmes de prévention populationnelle.

La loi de modernisation du système de santé ou loi du 26 janvier 2016 a, dans un premier temps, placé la prévention au cœur du système de santé dont la priorité a été portée sur les addictions chez les jeunes (tabac, alcool etc.), sur la lutte contre l'obésité et sur un meilleur dépistage des infections sexuellement transmissibles. Pour cela, les médias ont été

fortement utilisés : on pouvait voir des publicités contenant ces messages, des affiches dans les rues ou à la radio. Elles ont été diffusées à l'échelle populationnelle, sans ciblage précis.

Figure 11 Exemples de campagnes de prévention réalisées à l'échelle nationale via les médias

Pour compléter la loi de modernisation du système de santé, dans l'avis du 28 juin 2017 *« refonder les politiques de prévention et de promotion de la santé »*, le Haut Conseil a de nouveau insisté sur l'importance de replacer la prévention et la promotion de la santé au cœur de la stratégie. Les orientations de conception et de mise en œuvre de ces deux grands principes ont été formulées à l'automne 2017 lors de la définition de la Stratégie Nationale de santé pour la période 2018-2022.

Dans un second temps, en avril 2019, Madame Agnès Buzyn, ministre des Solidarités et de la Santé a présenté la nouvelle feuille de route concernant la politique de santé et plus particulièrement sur la politique du numérique en santé, définie dans le cadre de *Ma Santé 2022*, stratégie de transformation du système de santé. Cette feuille de route est décrite en cinq axes :

- Renforcer la gouvernance du numérique en santé
- Intensifier la sécurité et l'interopérabilité des systèmes d'information en santé
- Accélérer le déploiement des services numériques socles

- Déployer au niveau national des plateformes numériques de santé
- Soutenir l'innovation et favoriser l'engagement des acteurs

Dans ce rapport, Madame Buzyn décrit des projets étant actuellement en déploiement en France. Parmi ceux-ci, l'Espace Numérique de Santé est une plateforme publique, personnelle et personnalisable, disponible dès la naissance, sur laquelle les citoyens peuvent se connecter et accéder directement à leurs données et ainsi participer à la construction de leurs parcours de soins, en lien avec les acteurs de la santé et du médico-social. Cela a pour vocation de favoriser la prévention, la coordination, la qualité et la continuité des soins.(41)

Le Health Data Hub est la future plateforme des données de santé. Cette plate-forme permettra d'analyser les données de santé, notamment pour la recherche clinique. En écrivant ce rapport, les auteurs ont constaté que ces données étaient sous utilisées, non pas à cause de raisons techniques mais plutôt culturelle ou organisationnelle. Pour pallier cela, il est convenu que les données de santé financées par la solidarité nationale appartiennent au patrimoine commun. Les données seront ainsi mises en commun, dans le respect de l'éthique et des droits fondamentaux, et leur accès sera simple, la procédure sera unifiée et sécurisée. Grâce à cet entrepôt de données de santé national, le gouvernement espère évoluer vers une approche épidémique et populationnelle. Le Health Data Hub pourra également « fournir des échantillons représentatifs d'une population, avec un niveau d'informations qui sera assez élevé sur le plan quantitatif et qualitatif », explique Frédéric Valluet, expert Digital et Innovation lors de la Paris Healthcare Week 2019.

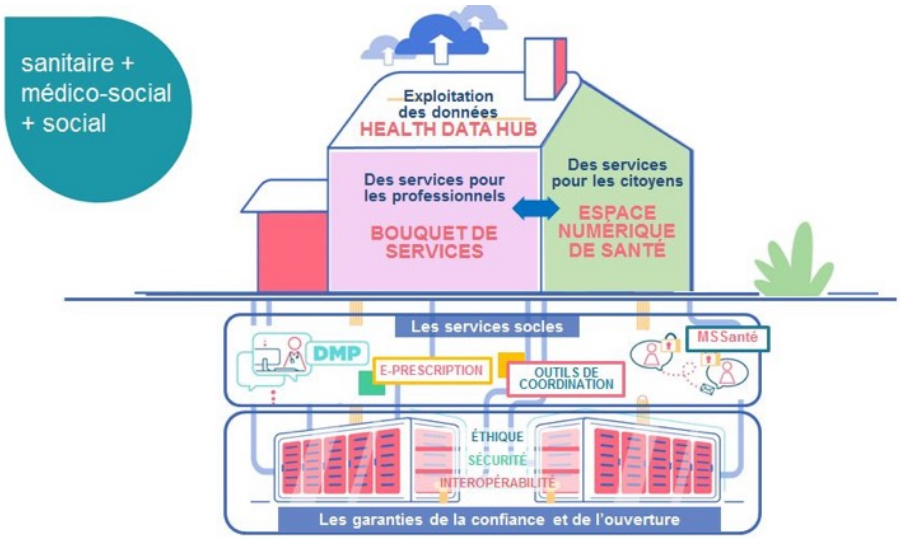

Figure 12 Présentation de la feuille de route "Accélérer le virage numérique"

Ces nouvelles plateformes pourront également permettre « que les gens qui ont des facteurs de risques ne développent les pathologies, mais aussi éviter que les gens qui ont des pathologies développent des complications » selon Antoine Malone, responsable du pôle Prospective Europe International de la Fédération Hospitalière de France (FHF), également interviewé à l'occasion de la Paris Healthcare Week.

Lors de ce salon, la FHF a également présenté une méthode d'analyse de la donnée dans le cadre de la prévention populationnelle. Ces algorithmes permettent d'identifier des territoires et de cibler des populations de patients souffrant de diabète ou d'insuffisance cardiaque afin de les prendre en charge dans une logique de responsabilité populationnelle. La responsabilité populationnelle est un concept de la FHF qui vise à regrouper l'ensemble des acteurs de la santé autour du patient afin de lui proposer une prise en charge globale. Ces travaux sont entrepris depuis 2017 et ont déjà permis de certifier deux algorithmes.

Le premier algorithme a permis d'identifier 132 territoires de santé qui correspondent aux consommations de soins réelles des populations. Les données qui ont permis de valider ce premier modèle sont issues de l'analyse des flux de population vers les établissements de soins, d'abord public puis privés et cela pour toutes activités confondues (Médecine-Chirurgie-Obstétrique, psychiatrie, Hospitalisation à Domicile et Soins de Suite et Réadaptation). Ces données ont été couplées avec les données issues du programme de médicalisation des systèmes d'information (PMSI) et de données populationnelles provenant de l'Insee.

Le second algorithme permet de trouver, au sein des territoires de santé identifiés par le premier algorithme, des patients atteints ou à risque d'insuffisance cardiaque ou de diabète. Cet algorithme résulte de l'analyse des données PMSI sur 5 ans et des données du Sniiram (Système national d'information inter-régimes de l'assurance maladie). Les résultats des patients retrouvés par ce modèle algorithmique est comparable aux affections de longue durée enregistrées par l'assurance maladie. Antoine Malone expliquait d'ailleurs que cet algorithme qui fonctionnait aujourd'hui pour les patients atteints ou à risque de diabète et d'insuffisance cardiaque devrait être également applicable à d'autres maladies chroniques comme la bronchopneumopathie chronique obstructive dont des travaux sont déjà prometteurs.

Les algorithmes ont nécessité une validation par une société savante française, « pour être sûr de travailler avec le meilleur savoir médical, validé dans un contexte français » explique Monsieur Malone.

Ces deux algorithmes ont pour vocation d'être « accessibles à tous » et d'être généralisés décrivait Antoine Malone, dans le but d'élaborer des stratégies de territoire en coordonnant les actions entre les praticiens de soins et en ajustant les ressources humaines et matérielles nécessaires. Ces stratégies seront différentes en fonction du niveau de gravité du patient. Ainsi, quatre niveaux ont pu être identifié, trois concernant les patients atteints par la pathologie : « complexe », « instable » et « stable » et un dernier niveau concernant les patients dits « à risque ». Cela permettra aux professionnels de santé de planifier des actions adaptées, allant du contrôle des symptômes à des actions de prévention ciblées et de « faire créer, par des professionnels de santé, des programmes de prise en charge de population extrêmement robustes » a-t-il ajouté. Dans ce sens, le Dr Sandra Gomez, directrice de projets médico-économiques et impliquée dans ces travaux, insistait sur l'importance de décloisonner les informations de santé : « une fois qu'on décloisonne, on redonne du sens et on peut agir sur la façon d'organiser les soins sur un territoire par rapport à de réels besoins » (42).

Dans cette partie, nous avons pu constater que, bien que la prévention soit un sujet important en France, le ministère de la Santé, via notamment à *Ma santé* 2022, s'intéresse plus globalement à l'utilisation du numérique pour faire évoluer les systèmes de santé.

Cependant, grâce aux interviews d'Antoine Malone, nous pouvons tout de même voir que certaines personnes commencent à s'interroger sur la réutilisation des données dans le cadre spécifique de la prévention.

Intéressons-nous maintenant à l'utilisation des big data dans des systèmes différents du système français. Nous avons choisi de nous intéresser à trois systèmes complétement différents : le modèle danois, le modèle chinois et le modèle adopté par l'organisation « Kaiser Permanente », présente aux Etats-Unis. Ce dernier est très souvent pris en exemple pour imaginer les systèmes de santé de demain.

3.2 La prévention dans d'autres systèmes que le système français

Afin de comprendre les différents aspects de la politique de santé française, il est important de la comparer à d'autres contextes. Réaliser un panorama des systèmes de prévention dans tous les pays du monde n'est pas l'objet de ce mémoire. Un focus a été fait sur trois systèmes.

Nous discuterons d'abord du modèle danois dans lequel la centralisation des données est assez ancienne, mais dont les difficultés d'utilisation sont assez marquées.

Ensuite, nous aborderons le modèle chinois pour lequel les données sont également centralisées et utilisées, avec une approche différente concernant le respect de la confidentialité, et la limitation des utilisations des données collectées.

Enfin, nous nous intéresserons à l'organisation « Kaiser Permanente », qui prend en charge des assurés dans un programme de santé incluant prévention et prise en charge, au travers d'un suivi régulier servi depuis plusieurs décennies par un système généralisé de collecte et d'analyse de données.

3.2.1 Le modèle danois

Le Danemark est un pays d'Europe du Nord de 5,7 millions d'habitants dont les revenus sont en moyenne plus élevés que les pays similaires, et bénéficiant d'une couverture

universelle des soins financée par l'impôt. Les infrastructures informatiques sont très élaborées, les services de santé sont numérisés et il existe des registres nationaux qui collectent, de la naissance à la mort, des données structurées sur tous les citoyens. (43)

Au Danemark, le modèle de la médecine 4P est au centre du système de santé. La médecine a pour but de rendre la médecine prédictive, préventive, personnalisée et participative. (44)

La personnalisation de la médecine et à fortiori de la prévention nécessite de lourds investissements dans la collecte, le stockage et l'utilisation des données de santé (45). Suite à ce constat, l'Union Européenne a décidé d'investir 1 milliard d'euros dans le cadre du septième programme cadre pour la médecine personnalisée. D'autres financements ont également été débloqués pour harmoniser les bio banques et les infrastructures de données dans le but de faciliter la mise en commun des données. Plus récemment, en janvier 2019, dans le cadre du programme Horizon 2020, l'Union Européenne a investi 21 millions d'euros supplémentaires pour constituer un dossier de santé interopérable.

Dans le système danois, la mise en commun facilite la recherche et l'expérimentation clinique mais aussi la stratification de groupe de population et la prévention personnelle ciblant les individus à risque de médecine générale. Les big data ont complétement modifié les attentes des professionnels de santé, mais aussi des patients. Ces derniers voient les Data jouer un rôle important dans leur santé.

L'intensification de la collecte de données implique de nouveaux rôles et responsabilités pour les patients et professionnels de santé.

Le registre national de patient a été instauré en 1967. Il contient les données du patient telles que les traitements, les résultats de biologie, les résultats d'examens médicaux, les causes du décès, que les soins soient fait à l'hôpital ou en cabinet de ville. Il regroupe également toutes les données issues des interactions avec les autorités publiques, comme des informations concernant le logement et le lieu d'habitation, les revenus, etc. Quasiment toutes les données sont aujourd'hui numérisées et sont reliées entre elles. Ce registre national d'identification est utilisé dans le domaine de la santé, mais pas seulement. En effet, elles sont également consultables et complétées par les services commerciaux comme les entreprises de transport ou de télécommunication et le citoyen est relié à son lieu d'habitation, à sa famille.

A titre de comparaison, en France, associer dans un fichier un code postal et une date de séjour dans un établissement donné est considéré comme identifiant. En contournement, des scores extrapolés, basés sur l'adresse, renseignent sur l'état de précarité, comme le score EPICES (46).

Au Danemark, les données issues du registre national du patient peuvent être utilisées à des fins administratives et de recherche, sans avoir eu besoin du consentement éclairé de l'individu à qui appartenait ces données. Cette utilisation des données personnelles est considérée comme peu controversée par les danois. (47)

Le Danemark, suite au constat de l'OCDE en 2013 expliquant que « la mine d'or n'est que peu exploitée » a décidé de se faire aider par des entreprises du secteur privé pour conseiller le gouvernement sur la manière de valoriser les données en développant des usages métiers concrets. Suite à ce constat, en 2016, le ministère de la santé et les régions danoises ont élaboré une politique générale visant à optimiser l'utilisation de des données relatives à la santé.

Concrètement, en termes de prévention, les sociétés privées venant en aide au gouvernement ont conseillé d'effectuer des bilans de santé ciblés afin d'aider les patients les plus complexes et les plus vulnérables. De plus, les médecins généralistes ont fait part de leur besoin de nouveaux outils informatiques pour les aider à détecter les citoyens présentant un risque élevé de maladie chronique (48). Ce rapport a permis d'établir un nouveau rôle pour les généralistes. En effet, avant cela ils n'avaient pas accès aux bases de données, à part pour les patients ayant eux-mêmes demandé une assistance médicale. A la suite de ce rapport, leur rôle préventif a été rétabli puisqu'ils doivent maintenant chercher les patients présentant le plus de risques afin de leur faire comprendre qu'ils sont sujets à des maladies chroniques.

3.2.2 Le modèle chinois

3.2.2.1 Les réformes de santé en Chine

En Chine, le gouvernement central est en charge de la législation, de la politique et de l'administration pour le secteur de la Santé. Depuis les années 1950, le système de santé chinois a beaucoup évolué.

La mise en place du système de santé publique chinois, dans les années 1950, a permis une diminution importante des maladies infectieuses et des affections chez la femme enceinte, le nourrisson et l'enfant ainsi qu'une diminution du taux de mortalité et une augmentation de l'espérance de vie. Cette phase initiale était concentrée sur la prévention et chaque citoyen pouvait profiter d'une offre de soins rudimentaire.

Dans les années 1970, la Chine a vu son modèle économique se transformer et évoluer vers une économie de marché. Le secteur de la santé a suivi ce même modèle. Le coût financier de la sécurité sociale et des soins de santé étant devenus trop importants, les entreprises d'Etat n'ont plus été en mesure d'assumer ce budget. Ainsi, du fait de ce bouleversement du système de santé, de nombreuses personnes habitant des zones rurales ou étant inactives en zones urbaines ont cessé de recevoir des soins de santé de la part de l'État. Il incombait maintenant aux patients de financer leurs propres dépenses de santé.

En 1998, un régime d'assurance de santé de base a été créé. Ce régime d'assurance était différent selon les citoyens. Cela a engendré de fortes inégalités entre les zones côtières et les provinces excentrées, entre les zones urbaines et rurales et entre les couches socio-professionnelles. Ce système comportait de grosses lacunes puisque 50% de la population urbaine et plus de 80% de la population rurale n'étaient pas couverts par l'assurance maladie ou n'avaient pas accès aux soins de santé (49).

Les réformes suivantes visaient à réduire ces inégalités au sein de la population. Toutefois, encore aujourd'hui, ces inégalités territoriales demeurent et l'accès aux soins est très déséquilibré et dépend du système d'enregistrement des familles. En effet, si la famille est enregistrée comme habitant en campagne, il est très difficile et très couteux de modifier son affectation, même si la famille déménage en zone urbaine. Ainsi, on distingue au moins trois types de systèmes de soins de santé différents :

- UEBMI : régime d'assurance médicale de base des salariés résidant en milieu urbain. Tous les employeurs urbains sont tenus de faire bénéficier leur personnel de ce régime. Entreprise et salariés cotisent pour cette assurance à hauteur, respectivement, de 6% de la masse salariale mensuelle pour le patronat et 2% de la rémunération salariale mensuelle (en France : 7% ou 13% selon le salaire).

- URBMI : régime d'assurance médicale de base en faveur des personnes résidant en milieu urbain. Il a été créé pour les personnes sans activités (étudiants, handicapés etc.). L'affiliation à ce régime repose sur le volontariat et est financé par des cotisations des assurés et des subventions publiques.
- NCMS : nouveau système de santé coopératif en milieu rural, géré par le ministère de la Santé. L'affiliation à ce régime est, tout comme l'RBMI, basé sur la volonté du citoyen. Ce régime bénéficie d'importantes subventions publiques mais est également financé par les cotisations des agriculteurs. Le NCMS est souvent qualifié de coopératif puisque les fonds qui alimentent ce budget sont mis en commun.

La réforme de 2006 avait pour vocation de permettre l'instauration d'un système de santé de base et à garantir l'accès des citoyens à des services de soins efficaces et abordables. Cette réforme avait également pour objectif de mettre en place des interventions de santé publique, notamment en matière de prévention des maladies, des épidémies et de développer les campagnes sur l'hygiène ainsi que les services de santé maternelle et pédiatrique.

3.2.2.2 Les big data, une affaire d'Etat

En Chine, les objectifs énoncés pour chaque secteur de l'économie, y compris la santé, sont à très long terme. Par exemple, le programme national « Healthy China 2030 » a été publié en 2016 et planifie des actions pour les quinze prochaines années. Dans ce programme, le président Xi Jinping a mis la santé au centre de l'ensemble des politiques du pays. Le pays a décidé de se concentrer davantage sur la prévention, notamment grâce à un meilleur style de vie et à la pratique régulière du sport, que sur le traitement des maladies.

Dans un autre secteur que celui de la santé, la Chine ambitionne à devenir le leader mondial dans l'utilisation de l'Intelligence artificielle et l'a montré au reste du monde lors de l'exposition mondiale sur l'intelligence artificielle qui s'est tenue à Shangai le 19 septembre 2018. Aujourd'hui, le pays utilise cette technologie dans divers politiques. Récemment, dans le cadre d'une politique visant à sécuriser la population, le gouvernement est en train de se doter d'un système national de notation de la population sur lequel la police s'appuie, en plus des 170 millions de caméras de vidéo-surveillances déjà présentes (50).

L’activité scientifique en intelligence artificielle, tous secteurs confondus, en Chine, est très importante. Quantitativement, elle est maintenant supérieure à celle des Etats-Unis et de l’Europe (51).

Le domaine de la santé n’échappe pas à cette ambition nationale. En effet, les politiques récentes encouragent l’utilisation d’internet dans le secteur médical et des soins de santé. En 2017, le marché de la santé digitale s’est rapidement élargi depuis 2014 et a généré plus de 10 milliards de yuans (soit 247 milliards d’euros) en 2017 (52).

En Chine, les conditions d’exploitation des données personnelles sont beaucoup moins strictes qu’en France et en Europe, et la pénurie de médecins, seulement 1,5 pour 1000 habitants contre 3,3 en Europe permettent à de nombreuses sociétés privées d’exploiter ce nouveau marché. C’est déjà le cas pour des start-up facilitant le diagnostic médical (53). Par exemple, *PereDoc* a développé un algorithme permettant de détecter les cancers du poumon, à partir d’images collectées dans plus de 180 hôpitaux. Cela est fait de manière systématique et permet de compenser le manque de médecins généralistes, de spécialistes et le temps nécessaire d’obtention de rendez-vous.

Ainsi, le Centre chinois de Contrôle et de Prévention des Maladies a, en 2012, collaboré avec Baidu pour mener une étude dans l'analyse de données volumineuses et utilisera la technologie de stockage de données et le moteur de recherche de Baidu pour construire le premier système de prévision de la grippe en Chine. Concrètement, il s’agissait de proposer, en temps réel, la détection et la prédiction épidémiologique du virus. Comme nous l’avons vu précédemment avec Google Flu, ils ont également utilisé les recherches internet du moteur de recherche Baidu pour modéliser et surveiller l’épidémie (54). Ce Centre National de Contrôle et de Prévention a plusieurs rôles dont celui d’exploiter le système de surveillance national et de mener des enquêtes spéciales pour comprendre, d’une part les maladies transmissibles et d’autre part les maladies chroniques et pouvoir orienter les programmes de prévention. De plus, ce centre est également amené à mener des activités d’évaluation des résultats et des impacts des programmes nationaux de prévention et de traitement (55). La détection précoce et le contrôle des maladies transmissibles bénéficieront à des milliards de personnes.

En supplément, la construction d'une plate-forme Cloud dans le secteur sanitaire en Chine met l'accent sur la prévention des maladies chroniques non transmissibles et la surveillance de l'état de santé de la population. D’ailleurs, de nombreux appareils et applications

installées dans les smartphones peuvent désormais collecter des informations personnelles sur la santé, telles que le pouls, la tension artérielle et le taux de sucre dans le sang, permettre la surveillance en temps réel de l'état physique essentiel et émettre des recommandations relatives à une alimentation saine, aux exercices et aux médicaments. Ces notions d'appareils permettant le quantified self seront décrits plus précisément dans une autre partie de ce mémoire.

3.2.3 Un modèle américain : le Kaiser Permanente

3.2.3.1 Le Kaiser Permanente : la naissance du prépaiement

Le Kaiser Permanente est né dans les années 1930 grâce au Docteur Stanley Garfield. Ce jeune chirurgien, travaillant auprès des hommes impliqués dans la construction de l'aqueduc de la rivière Colorado, décida d'ouvrir un hôpital, le Contractors General Hospital, dans le désert de Mojave afin de soigner les travailleurs malades et blessés. Toutefois, il constata rapidement qu'il ne pourrait pas maintenir l'hôpital ouvert, faute de ressources financières suffisantes. En effet, il avait beaucoup de mal à se faire rembourser les frais engagés pour soigner ses patients assurés auprès de leurs assurances santé, et il constata également que de nombreux travailleurs ne disposaient pas d'assurance. Un jour, Harold Hatch, un ancien industriel devenu assureur proposa son nouveau projet au Dr Garfield : les compagnies d'assurances devraient verser un montant journalier, par travailleur pris en charge par l'hôpital et cela à l'avance. Garfield proposait désormais aux travailleurs deux options : pour cinq cents par jour, ils bénéficiaient de cette nouvelle forme de couverture maladie et pour cinq cents supplémentaires, cette couverture était étendue aux problèmes médicaux non liés au travail. Ce fut un succès puisque des milliers de travailleurs de l'aqueduc vinrent s'inscrire à l'hôpital.

Lorsque le projet d'aqueduc fut terminé, un autre industriel, Monsieur Henry Kaiser, demanda au Dr Garfield de prendre en charge 6500 travailleurs et leurs familles qui œuvraient pour ce qui était alors l'un des plus grands chantiers de construction, celui du barrage de Grand Coulee. Stanley Garfield n'hésita pas très longtemps et décida d'y restaurer un ancien hôpital en un hôpital très moderne et d'y recruter des médecins pour travailler dans un « cabinet de groupe prépayé ». Encore une fois, il rencontra un grand succès. En 1941, le chantier du barrage pris fin et Kaiser demanda à Garfield de trouver

une solution pour les 30 000 travailleurs des chantiers navals de Richmond mobilisés en prévision de l'intervention massive des Etats-Unis dans la seconde guerre mondiale, qui se préparait. Garfield créa, en association avec Kaiser, un centre de soins regroupant plus de 75 médecins et prenant en charge ces travailleurs. C'est ainsi que le KP, pour Kaiser Permanente, est né. A la fin de la guerre, le 21 juillet 1945, le KP fut ouvert au public. En 10 ans, plus de 300 000 personnes ont rejoint ce programme, dans le nord de la Californie.

Aujourd'hui, KP est toujours basé sur la coopération étroite entre deux organisations : le régime de santé, d'une part, et les hôpitaux de la Fondation Kaiser à but non lucratif et les groupes médicaux Permanente d'autre part.

3.2.3.2 Le Kaiser Permanente actuel

Le Kaiser Permanente est maintenant un système de soins américain à but non lucratif, créé en 1945. Il est aujourd'hui utilisé dans neuf états, assure 12.3 millions de citoyens américains et emploie plus de 300 000 personnes, infirmiers et médecins confondus.

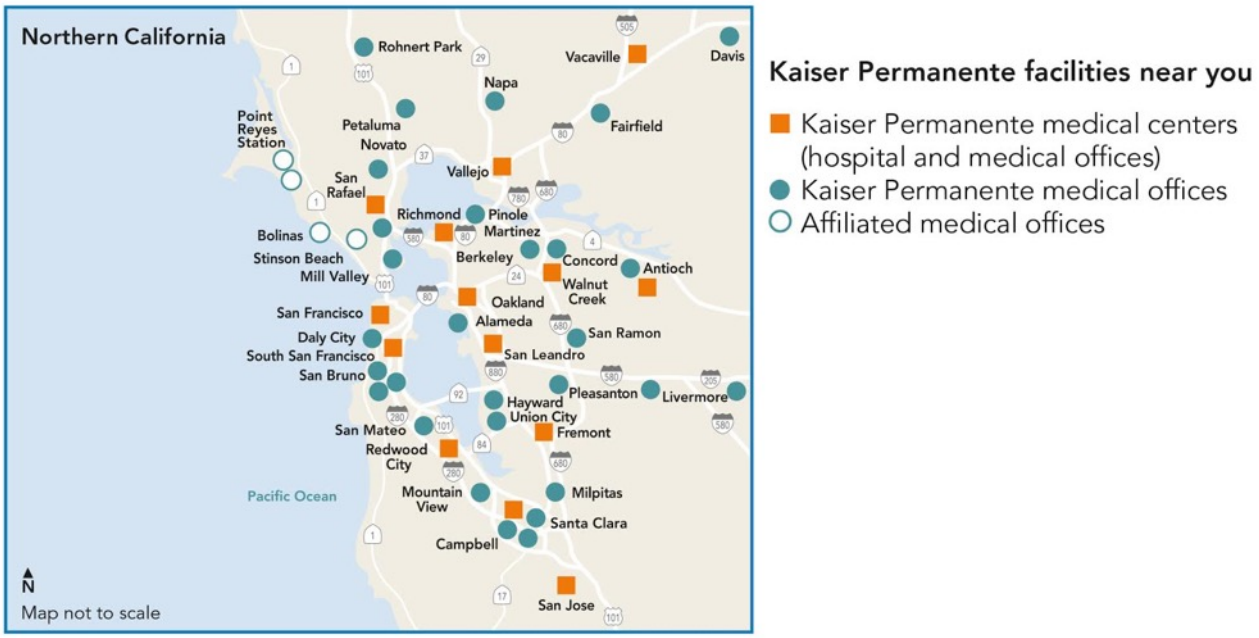

Figure 13 Centres de soins Kaiser Permanente en Californie du Nord (56)

Le Kaiser Permanente est considéré comme l'un des systèmes de santé les plus performants.

Ce modèle repose sur une gestion des patients et de leurs aidants en fonction de leur profil de risque. Ces profils sont classés selon quatre niveaux :

- Le niveau 0 concerne un patient en bonne santé,
- Le niveau 1 correspond à un patient ayant une maladie chronique à faible risque,
- Le niveau 2 s'adresse à un patient ayant une maladie chronique à haut risque,
- Le niveau 3 s'applique à un patient atteint d'une maladie chronique complexe et à un très haut niveau de risque.

Selon le niveau auquel appartient le patient, différentes mesures lui seront proposées, allant d'actions de prévention, de mesures favorisant le bien-être, à une logique de « case management » nécessitant une prise en charge complète et transverse et l'intervention d'un gestionnaire des cas complexes.

Le cœur du modèle du Kaiser Permanente repose sur les outils de gestion, basés sur les technologies de l'information, des données des patients, outils que KP a développés. Il s'agit d'un système de dossiers de santé électronique, intitulé HealthConnect qui héberge plusieurs types d'usages, comme ceux disponibles au travers d'un portail auquel le patient peut se connecter, puis consulter les comptes rendus de ses dernières visites, les résultats des tests de biologie, les prescriptions. Ce portail est aussi une plate-forme d'échanges entre les patients et les médecins, ou entre les médecins entre eux, leur permettant de réaliser des consultations par téléphone ou en visioconférence. Cette plateforme permet également d'organiser les parcours patients, d'identifier l'efficacité et la rentabilité des traitements administrés, d'avoir un retour des patients systématique leur permettant d'améliorer continuellement leur système.

Ces données sont collectées lors d'un séjour à l'hôpital mais également lors de visites chez les généralistes, à la pharmacie ou lors de soins infirmiers à domicile.

Aujourd'hui, le Kaiser Permanente dispose de l'une des bases de données les plus riches du monde.

Les données lui ont permis de développer des modèles prédictifs reconnaissant notamment les signes avant-coureurs de certaines pathologies et ainsi de prendre en charge le patient

en amont de la maladie et de lui administrer la prévention des risques qui lui convient le mieux. Des algorithmes permettent par exemple d'alerter les médecins lorsque le patient risquerait de se détériorer plus rapidement qu'un autre. Cet algorithme appelé « Advance Alert Monitor » est actuellement utilisé dans 21 centres de soins. « Le système AAM synthétise et analyse les statistiques de l'état civil, les résultats de laboratoire et d'autres variables afin de générer des scores de risque de détérioration horaires pour les patients hospitalisés adultes dans les unités de médecine-chirurgie et de soins de transition », déclare Dick Daniels, vice-président exécutif et DSI de Kaiser Permanente. « Les équipes d'hôpitaux distants évaluent les scores de risque toutes les heures et informent les équipes d'intervention rapide de l'hôpital en cas de détection d'une éventuelle détérioration. L'équipe d'intervention rapide procède à une évaluation du patient à son chevet et calibre le déroulement du traitement avec le médecin hospitalier. » Le modèle Advance Alert Monitor s'appuie sur trois modèles analytiques prédictifs et permet d'analyser plus de 70 éléments du dossier de santé électronique d'un patient afin de générer un score de risque composite.

Ainsi, toutes les actions de prévention menées par le Kaiser Permanente commencent à porter leurs fruits : le ciblage et l'éducation thérapeutique du patient consommant du tabac par exemple de permis de diminuer considérablement le nombre de fumeurs, bien plus qu'ailleurs aux Etats-Unis ou en Californie.

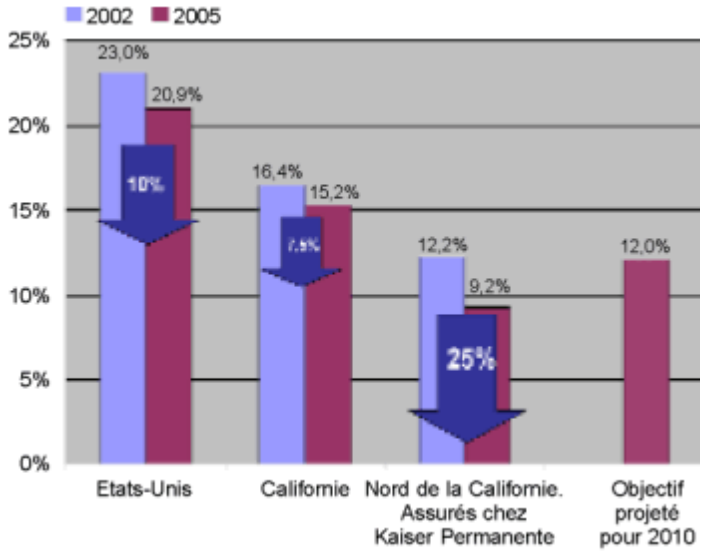

Figure 14 Un exemple de médecine préventive (57), Kaiser Permanente

3.2.3.3 « Collaborative Cardiac Care Service », un exemple de prévention tertiaire chez des patients atteints de coronaropathie.

Le “Collaborative Cardiac Care Service” a été mis en place en 1996 par le Kaiser Permanente du Colorado et a pour vocation de prendre en charge des patients atteints de coronaropathie.

Il est composé d’une équipe d’infirmiers et de pharmaciens mais travaille également en collaboration avec les patients, les médecins, cardiologues et autres professionnels de santé, pour proposer des stratégies de réduction des risques pour les patients.

Les missions du “Collaborative Cardiac Care Service” sont variées puisqu’elles interviennent dans la modification du mode de vie, aident à l’ajustement médicamenteux (conciliation médicamenteuse entre les prescriptions à l’entrée de l’hôpital, celles prescrites pendant le séjour, et celles préconisées à la sortie), à l’éducation du patient, à la surveillance en laboratoire et à la gestion des effets indésirables.

Le “Collaborative Cardiac Care Service” utilise un logiciel d'enregistrement électronique de dossiers médicaux et de suivi des patients pour documenter toutes les interactions avec les patients, suivre les rendez-vous des patients et collecter des données pour évaluer les résultats à court et à long terme.

En 2008, le “Collaborative Cardiac Care Service” suivait plus de 12 000 patients atteints de coronaropathie. Le “Collaborative Cardiac Care Service” a démontré une amélioration des résultats de substitution, notamment : une augmentation du dépistage du cholestérol de plus de 30%, avec plus de 97% des patients se faisant régulièrement vérifier leur taux de cholestérol, depuis leur sortie de l’étude.

Aussi, le pourcentage des patients ayant un taux de cholestérol LDL sous la limite des 100 mg/dL a augmenté de 47,2% (seulement 25,5% des patients avaient moins de 100mg/dL de c-LDL dans le sang et 72,7% étaient en dessous de ce taux à la sortie de ce programme). Le “Collaborative Cardiac Care Service” a également montré une réduction de 76% en moyenne et chez les patients inclus dans les 90 jours après leur coronaropathie, une diminution de 89% (58).

Les résultats suggèrent que la mise en œuvre complète, agressive et continue des stratégies de prévention utilisant des systèmes permettant un suivi étroit et une surveillance des patients, est associée à une réduction de la mortalité.

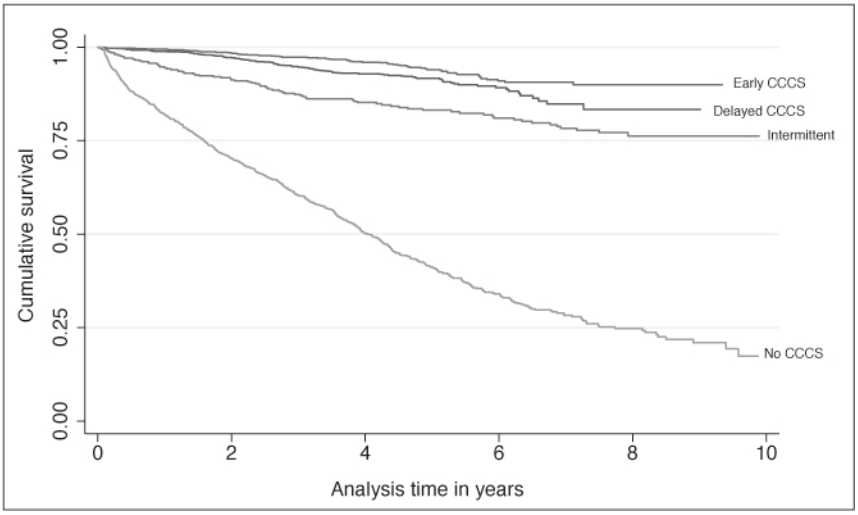

Figure 15 Courbe de Kaplan Meyer montrant l'efficacité du programme "Collaborative Cardiac Care Service"(58)

Concernant le financement de cette étude, les données avaient été récupérées d'une étude précédente, ce qui a permis d'économiser plusieurs milliers de dollars en frais d'hospitalisation. De plus, les auteurs de l'étude expliquent qu'il leur a été très difficile de calculer précisément les coûts totaux mais ils estiment que la réduction de la morbidité et de la mortalité engendrée par ce programme a permis d'économiser 3 millions de dollars, somme correspondant aux 266 cas par an évités (frais d'hospitalisation, salaire du personnel soignant, de la thérapeutique associée …).

IV. La prévention personnalisée constitue-t-elle un nouveau marché ?

Afin de déterminer si les logiciels de prévention personnalisée représentent un vrai marché, économiquement viable, nous devons :

- D'une part voir s'il y a une demande (comprendre quelles sont les demandes des patients, de l'Etat, des collectivités locales, des médecins, etc.).
- D'autre part, nous devons déterminer quelles offres peuvent être proposées, et en quoi elles répondent aux besoins, dans un équilibre économique (ou d'équilibre de

valeurs, au-delà de l'aspect financier). Certaines pistes seront illustrées au travers de deux exemples concrets, en production.

4.1 La demande

La prévention personnalisée grâce aux Big Data est un enjeu pour plusieurs catégories d'individus et d'organisation. En effet, nous pouvons distinguer la demande de la population, des médecins et de l'Etat et des collectivités locales puisque chacun aura un rôle différent dans la prévention.

4.1.1 La demande des patients : rester en bonne santé, mieux se connaître, éviter des incidents, suivre finement sa maladie chronique, fournir simplement à son médecin des informations, s'auto-éduquer

Pour la population, une personnalisation de la prévention serait gage d'une meilleure santé, d'une amélioration du bien-être. Pour cela, depuis quelques années, des outils dits de « *quantified self* » permettent de suivre, quasiment en temps réel, ses propres informations de santé. Ce genre d'outil va, au cours du temps, enregistrer différentes données. Ces données peuvent être le poids, le rythme cardiaque, la température corporelle ou le nombre de pas effectués dans une journée. Aujourd'hui, ils peuvent être permettre d'augmenter ses performances mais également d'apprendre à connaitre son corps. Par exemple, *Propeller Health for asthma* est un capteur que l'on fixe sur les inhalateurs qui permettent d'enregistrer la fréquence d'utilisation et qui sonne lorsque le patient n'a pas pris sa dose de médicament. L'application apprend sur les crises d'asthme du patient et sur les habitudes d'utilisation du médicament, et aide le patient à gérer les symptômes et à identifier les déclencheurs dans le but de prévenir l'apparition des crises d'asthme. Comme nous l'avons vu dans une partie précédente, les français utilisent beaucoup ces objets mais ils souhaiteraient d'avantage que les données collectées puissent être utiles aux médecins et aider au diagnostic, en tant que données de routine.

4.1.2 Pour les professionnels de santé : gagner en rendement des actions de prévention

Pour un professionnel de santé, généraliser à toute une population une action de prévention, sans personnaliser (à l'aide de données ou non), est la plupart du temps totalement inefficace. En effet, le temps médical mobiliser pour poser le diagnostic ou le confirmer (lorsque des non-médecins, ou des systèmes informatisés émettent la première hypothèse) sur toute une population mobiliserait des ressources médicales importantes, au détriment de cas avérés.

Les professionnels de santé attendent de systèmes de prévention personnalisée basée sur la data d'être mobilisés à bon escient, par des patients dont la probabilité de subir un incident aigu, ou de développer une nouvelle maladie, est réellement significative.

4.1.3 Pour l'état : améliorer la « coût-efficience » des mesures de prévention

Pour l'Etat, en ciblant la prévention, les résultats espérés pourraient être meilleurs et pourraient ainsi permettre de diminuer les coûts des soins et des traitements d'une maladie.

Aujourd'hui, la demande concerne essentiellement trois champs d'activité : la cancérologie, la cardiologie et la neurologie (59). Que ces trois domaines médicaux arrivent en tête des recherches associant intelligence artificielle et diagnostic n'est pas totalement inattendu. Ce sont les trois principales causes de décès dans le monde. Par conséquent, les diagnostics précoces sont essentiels pour prévenir la détérioration de l'état de santé du patient.

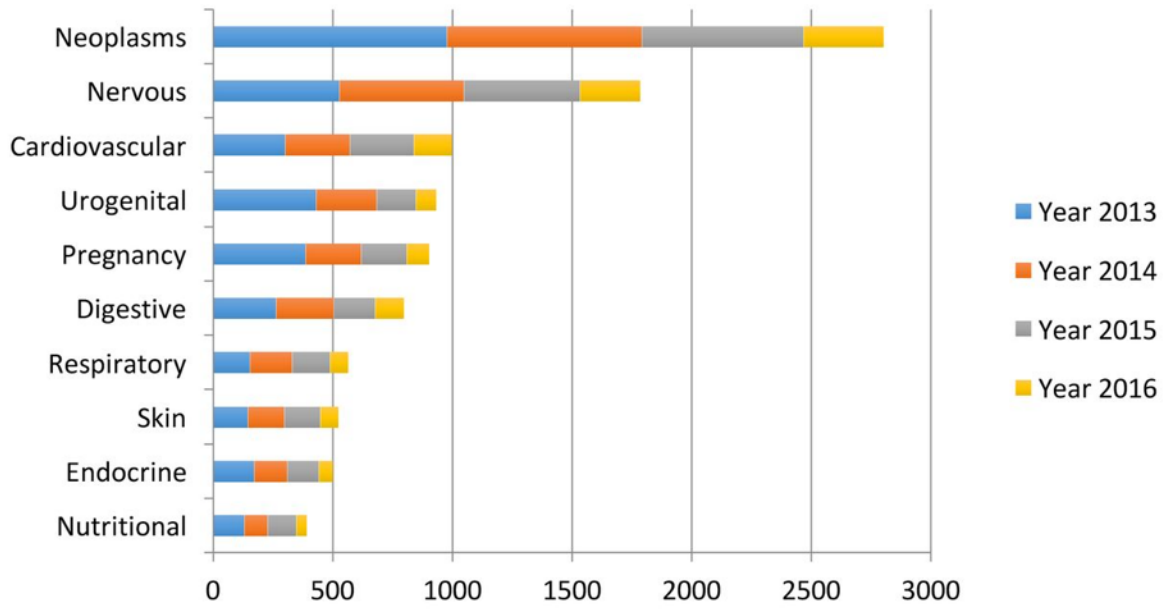

Figure 16 Les dix principales maladies considérées dans la littérature sur l'intelligence artificielle

En segmentant la population sans apriori, l'intelligence artificielle appliquée à la prévention vise à réduire les inégalités entre individus. Ces inégalités peuvent être financières mais également en termes d'accès aux soins. Prendre conscience que toutes les zones ne sont pas pourvues de la même façon en médecins et hôpitaux permettra de concentrer la prévention sur ces zones. Cela permettrait à l'Etat de lancer des programmes de prévention sur un temps plus long, plus spécifiques et dont les modes de communications seront plus ciblés. Cela nous a été confirmé par le Docteur Grosbois qui expliquait « Les Big Data devront permettre de mieux prendre en charge les patients les plus précaires, les perdus de vue » (60).

Pour l'Etat, le but ultime de la personnalisation de la prévention serait de réduire significativement le nombre de patients atteints de maladies chroniques par du profilage de patients à risques, et de maladies transmissibles par un meilleur ciblage des populations concernée pour mettre en œuvre des campagnes de dépistages efficaces.

4.1.4 Illustration par une expérience régionale : le passeport vasculaire© Centre Hospitalier de Valenciennes

Une expérience réelle a été étudiée pour illustrer ce sujet : celle de la construction d'un système de prévention personnalisée, généralisée, pour tous les patients admis dans un

hôpital. Ce projet est mené au Centre Hospitalier de Valenciennes par le Docteur Lernout, qui a été interviewé dans le cadre du mémoire.

Ce projet s'appelle le « passeport vasculaire ».

Le passeport vasculaire utilise à l'hôpital de Valenciennes des techniques d'Intelligence Artificielle pour classer tout patient présent, pour d'autres raisons qu'une pathologie vasculaire, dans un niveau de risque de développer une pathologie vasculaire.

Ainsi, un patient admis aux urgences pour un bras cassé, pourra se voir proposer une consultation de dépistage, si les informations présentes dans son dossier médical informatisé le classent dans une catégorie de risque vasculaire. L'outil embarquant une intelligence artificielle a été conçu par le service du Dr Lernout, la société ALICANTE et la Direction Informatique de l'hôpital de Valenciennes. Cet outil est en cours d'évaluation.

Lors de l'interview menée dans le cadre du mémoire, le Docteur Lernout a insisté sur ce qui semble une évidence, mais n'est pas toujours appliqué : il vaut mieux recevoir un patient en consultation de prévention que lorsque le sinistre est arrivé. Ce « sinistre » peut arriver la nuit ou les jours fériés lorsqu'il y a moins d'assistance médicale, et que le coût d'une opération est toujours bien plus élevé que celui d'une consultation préventive.

Au travers cette description des demandes émanantes des parties prenantes de la prévention, nous pouvons constater que les professionnels de santé et l'Etat souhaite mieux cibler la prévention pour qu'elle soit plus efficace et plus rentable. Le patient à quant à lui le besoin de mieux comprendre, gérer et apprendre de ses données de santé.

4.2 L'offre

Aujourd'hui il existe de nombreuses sociétés qui utilisent les données de santé pour faire de la prévention personnalisée, chaque société ayant son secteur de spécialité, chaque société se positionnant sur un maillon précis de la chaîne de valeur. Ces différentes possibilités sont listées dans une première partie présentant les différentes offres, et leur

business model associé. Ensuite, nous illustrerons par plusieurs cas concrets d'offre de prévention basée Data :

- Arterys, société d'analyse d'images sur le « cloud » spécialisée notamment dans les maladies pulmonaires ;
- Le Centre Hospitalier de Valenciennes et ALICANTE, autour du passeport vasculaire.

4.2.1 Les offres disponibles et les business models associés

Aujourd'hui, plusieurs offres, associées à plusieurs business models, sont présents sur le marché :

- **La vente de données** dont le tarif dépendra du volume, de la qualité, de la nature, du contenu, de la fraicheur, de la représentativité etc. des données.

  OPTUM par exemple est une société spécialisée dans la vente de données, d'informations et de logiciels dans le domaine de la santé et du management.

- **La vente de compétences** à des acteurs non technologiques du secteur santé. C'est par exemple ce qu'une SSII (société de services et d'ingénierie en informatique) propose : elle va agir en tant que sous-traitant pour créer un outil informatique, parce que le client ne dispose pas des compétences nécessaires en interne. Les acteurs sont effectivement les SSII, mais aussi certains cabinets de conseil, ou des plates-formes associées à des laboratoires universitaires. Ces derniers sont les plus crédibles quant au respect des méthodes d'évaluation de la précision de modèles prédictifs.

- **La fourniture d'une plate-forme technologique** permettant de créer des modèles basés intelligence artificielle. Parmi ces acteurs, nous pouvons distinguer plusieurs catégories :

  - Acteurs cloud permettant un paiement à l'usage. Parmi ceux-là, nous retrouvons les plus grosses entreprises comme Google avec Google Cloud Platform, Amazon avec AWS et ses fonctions intelligence artificielle intégrées,

Microsoft avec AZURE et ses fonctions intelligence artificielle intégrées. Dans le secteur santé, nous pouvons également citer un projet du MIPIH (61) qui consiste à offrir un accès à des services et de la puissance de calcul en environnement hébergeur de données de santé. Ces acteurs adoptent en plus du modèle de paiement à l'usage, celui du business model de l'écosystème, de la marketplace : des éditeurs tiers doivent pouvoir héberger leur application spécifique métier dans l'environnement, s'en rendant d'ailleurs dépendant. L'accès au marché est dans ce cas accéléré, mais le lien de dépendance peut avoir une forte influence sur le destin de l'offre voire de la société hébergée. Concernant les sociétés implantées aux Etats-Unis, le Cloud Act (62) qui permet la saisie de données en tous points du globe en cas de décision de justice aux Etats-Unis, freine l'adoption par de nombreux acteurs de ces plates-formes.

- Acteurs plus classiques de l'édition comme la société SAS (63) spécialisée dans l'analytique, la business intelligence et le management des données.

- **La fourniture de progiciels, de solutions clés en main** : chez les acteurs actuels, il existe plusieurs modèles de tarification : en souscription, en paiement à l'usage, ou en mode « licence + maintenance ». Le client n'achète pas forcément de l'intelligence artificielle (même si le caractère valorisant qu'a actuellement l'adoption par une entreprise d'une solution basée intelligence artificielle peut intervenir dans la décision d'achat), mais le retour sur investissement d'une solution basée intelligence artificielle. De plus, la solution basée intelligence artificielle doit :

    - Soit faire quelque chose grâce à l'intelligence artificielle, qui n'était pas possible avec les techniques précédentes (exemple : analyse sémantique de courriers médicaux francophones)

    - Soit améliorer ce qui était fait auparavant. C'est dans cette catégorie que se trouvent ARTERYS, ALICANTE.

- **Les partenariats** : dans cette catégorie, il existe également un ensemble de schémas variables, mêlant relations partenariales et relations commerciales.

  Le modèle le plus répandu, amené par l'apparition de startups d'un nouveau genre, s'appuie soit sur la proposition de projets de « preuves de concept » gratuits pour le

site client, soit sur la valorisation de ces « preuves de concept » auprès d'investisseurs dans le cadre d'une levée de fonds.

Toutefois, il semblerait que ce modèle trouve actuellement ses limites, notamment parce que les sites bénéficiant des « preuves du concept » ont un niveau d'implication en général faible, face à l'absence d'enjeux financiers.

On trouve aussi des co-constructions d'offres entre acteurs de la recherche, du monde hospitalier, et de l'édition ou de l'intégration de logiciels et, selon la réussite de l'évaluation du produit co-construit, ou testé dans le cadre d'une « preuve du concept », les partenaires se partagent les différents droits de diffusion, de propriété intellectuelle, de l'actif issu du partenariat. Ces modèles sont souvent financés ou cofinancés par des subventions ponctuelles : leur pérennisation peut poser problème.

4.2.2 Illustrations

Après avoir vu les différents business models actuellement disponibles et utilisés dans l'édition de logiciels basés sur l'intelligence artificielle, nous allons nous intéresser à Arterys et au Passeport Vasculaire.

4.2.2.1 Arterys

Arterys est un éditeur de logiciels informatiques qui développe des outils pour accompagner les médecins dans leurs diagnostics. Ils proposent plusieurs services : permettre de visualiser le flux sanguin dans le corps humain, mais également des services en oncologie et en neurologie. Les données sont enregistrées dans le cloud (qui s'appuie actuellement sur le cloud d'Amazon, AWS) en vue d'être analysées, traitées et comparées au moyen de l'intelligence artificielle afin de fournir au médecin un maximum d'informations nécessaires pour l'aider dans son diagnostic. L'accès à la plateforme Arterys se fait exclusivement via un navigateur web et peut aussi permettre de visualiser les images des patients.

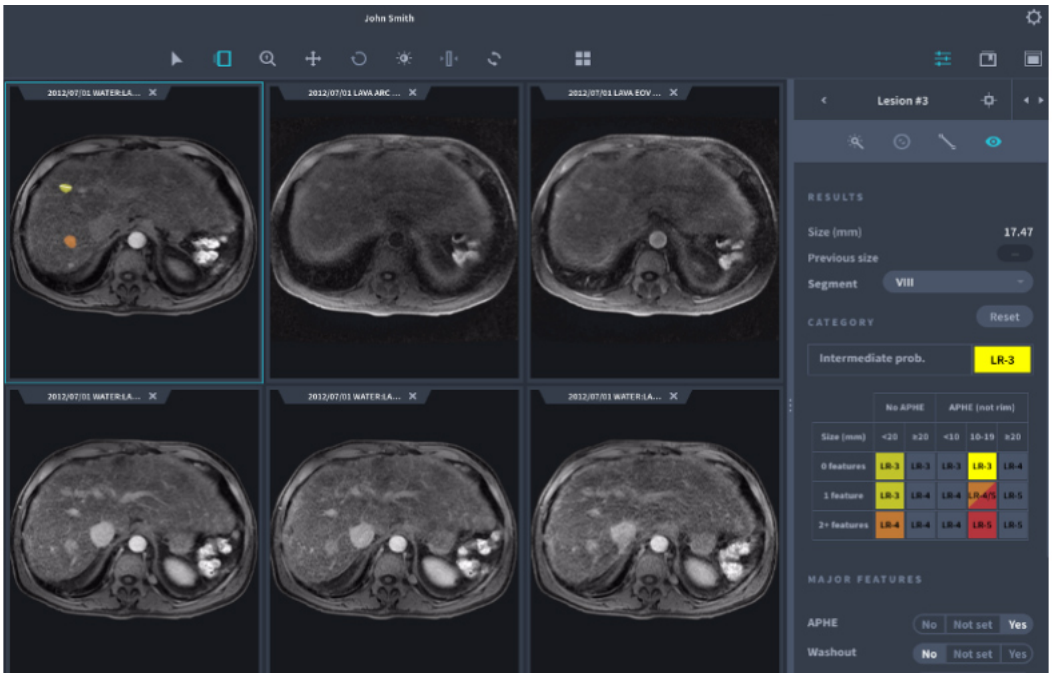

Figure 17 Copie d'écran de l'outil de visualisation et d'affichage de prédictions ©ARTERYS

Maya Khalifé (64), product manager chez Arterys explique : « Chez Arterys, nous avons initialement développé une lecture d'images automatisée. Aujourd'hui, un nouveau projet est en cours de finalisation. Il s'appuiera sur les technologies que nous avons développées pour la lecture d'image automatisée, mais sera complété par une aide au diagnostic et au traitement. Le but est d'agréger des données de types différents : résultats imagerie, données textuelles du dossier patient, données biologiques etc. pour prévenir au mieux l'avenir du patient et ajuster le traitement. Il s'agit de donner plus d'éléments pour guider la décision médicale. »

Ce nouveau projet traitera les images pulmonaires et sera entrainé pour reconnaitre les nodules pulmonaires. Ainsi, en disposant de données patients suivis dans le temps, il sera capable de faire remonter les images où se trouve un nodule pulmonaire et, en fonction de ses données, de proposer des examens supplémentaires ou un traitement correspondant au patient et à ses habitudes. Si un patient a déjà eu des expériences de chimiothérapie et qu'elle n'avait pas été concluante alors un autre mode de traitement sera proposé.

4.2.2.2 Le passeport vasculaire

Le passeport vasculaire est un projet qui est en train d'être développé entre le Centre Hospitalier de Valenciennes et Alicante, éditeur de logiciels informatiques spécialisé dans l'intelligence artificielle dans le domaine de la santé. Ce projet a vu le jour en constatant que dans la région de Valenciennes, la mortalité était bien supérieure à la mortalité nationale. Il a été estimé qu'a minima 600 morts par an pouvaient être évitées.

« Il s'agit, explique le Docteur Lernout, d'une part d'une liste de travail dans laquelle, grâce à des méthodes algorithmiques, des patients considérés comme présentant des risques cardiovasculaires plus important que la normale se verraient prescrire des consultations spécifiques. Le but étant d'encourager les personnes ayant un risque significatif à venir consulter et à les suivre dans le temps. Pour cela, nous disposerons également d'un calendrier évolutif pour éviter les perdus de vue » (65).

C'est un projet transverse puisque « (…) les patients qui seront convoqués ne seront pas forcément venus consulter dans le service cardio-vasculaire, mais également parce que plusieurs corps de métiers nous aiderons : des infirmiers transverses seront affectés au projet, des secrétaires médicales devront s'assurer du suivi des patients et des consultations externes pourront être prescrites (par exemple chez un tabacologue, nutritionniste etc.). Le but étant de responsabiliser le patient sur son état de santé. » (65)

Dans le cadre du passeport vasculaire, les données utilisées sont des données de santé qui seront, dans un premier temps, collectées suite à la venue du patient au sein de l'hôpital et qui pourront être complétées lors de la première consultation, par l'intermédiaire d'un questionnaire.

Concernant le business model du passeport vasculaire : l'étude préliminaire au projet a montré que la prévention personnalisée générait davantage de gains pour le patient, et de valeurs pour l'hôpital (qu'il s'agisse de recettes liées à l'activité en cas d'opérations lourdes, mais aussi d'image de marque auprès du patient), que de coûts liés à l'investissement initial et au fonctionnement du système.

Avec ce projet, c'est une nouvelle organisation de la prévention qui est en train de voir le jour au Centre Hospitalier Valenciennes. Le parcours de prévention sera concrètement personnalisé grâce aux Big Data.

Conclusion

Prévention santé, basée Data et intelligence artificielle : une activité en croissance

D'après les éléments que nous avons décrits dans les parties précédentes, nous pouvons retenir l'hypothèse suivante : l'utilisation de la Data et des techniques d'intelligence artificielle pour la prévention est un secteur en forte croissance.

Les faits étayant cette hypothèse sont, notamment : l'augmentation de l'activité de recherche scientifique dans cette spécialité ; le positionnement d'un nombre croissant d'acteurs, du monde du logiciel mais aussi des dispositifs médicaux dont les objets connectés, sur ce nouveau marché ; l'importante activité législative et règlementaire, en France et en Europe, sur ces sujets, notamment autour de la certification des dispositifs médicaux logiciels ; et, le retour terrain des trois acteurs représentatifs interviewés en support de ce travail.

Prévention santé, basée Data et intelligence artificielle : de nombreux défis à relever

Défi 1 : Les systèmes de prévention, qu'ils soient basés Data ou non, sont difficiles à dupliquer d'un contexte à un autre

Le premier défi qui semble devoir être relevé est celui, pour un pays et ses agences, de trouver son chemin, en fonction de paramètres divers. D'une manière générale, un système de prévention n'est pas qu'affaire de technique et de financement : il est fortement lié à son contexte culturel et juridique.

D'après ce que nous avons pu voir, de nombreux pays occidentaux tentent cette nouvelle approche en termes de prévention. Il semblerait tout de même que certains pays ou certaines organisations aient pris de l'avance par rapport à d'autres.

En effet, le Danemark dispose depuis 1967 d'un numéro d'identification citoyen ce qui permet à l'Etat de collecter toutes les données relatives à cette personne, que ce soient des données personnelles, des données de santé ou des données environnementales par exemple.

En France, le Dossier Médical Partagé vise à rendre disponible pour le patient et les médecins, toutes les données de santé de ce patient. Cette énième évolution du vieux projet (2007) de Dossier Médical Partagé a été annoncée il y a quelques mois, mais éprouve déjà des difficultés dans sa mise en œuvre puisqu'il revient à chaque citoyen de l'ouvrir et de l'utiliser, ou non.

La différence dans la dynamique d'adoption pourrait s'expliquer en partie, selon certains observateurs, par le fait que le Danemark est un pays majoritairement et historiquement protestant et que, dans les pays à forte culture protestante, le rapport des citoyens à la publicité des données privées est très différent des autres pays, notamment la France. En effet, dans le protestantisme, la communauté prévaut sur l'individualité, et si les données personnelles peuvent servir à la communauté, alors la population ne s'y opposera pas. Cela est complétement entré dans les mœurs puisque, comme nous l'avons vu précédemment, les danois sont confiants quant à l'utilisation de leurs données personnelles pour le bien commun (45).

En plus de ces différences sur la culture du partage de données, s'ajoutent parfois les fortes différences de législation, découlant de fortes différences entre les systèmes politiques. Entre la Chine par exemple, qui instaure une obligation de fourniture des données à un Etat qui en monopolise les applications, en santé et dans d'autres secteurs, et la France qui protège de plus en plus la propriété qu'a un citoyen de ses données, il existe un fossé immense.

Défi 2 : Le défi de la centralisation des données pour produire des modèles efficaces dans un grand nombre de contextes, malgré l'hétérogénéité des données

La nature des données, leurs équilibres internes, les dialectes de codage, etc., forment la *signature unique* d'un jeu de données. Or un modèle issu d'un apprentissage sur un jeu de données x ne sera pas (aussi) performant sur un jeu de données y. A l'exception des méthodes récentes de Decentralized Learning ou Federated Learning, aucun modèle prédictif issu des techniques d'intelligence artificielle connexionniste ne semble pouvoir être réutilisé à l'identique d'un hôpital à un autre, d'un territoire à un autre.

Défi 3 : Gagner la confiance et l'adhésion des utilisateurs finaux de ces systèmes, grâce à la certification et à l'ergonomie

Aucun système de prévention personnalisée basée Data ne s'imposera s'il n'inspire pas la confiance des professionnels de santé

Une certification doit s'imposer pour garantir la méthodologie de production des modèles basés intelligence artificielle.

Le processus de certification doit permettre de gagner la confiance des professionnels de santé dans l'usage des techniques basées sur les Data, utilisant l'intelligence artificielle.

Ce processus doit reprendre et adapter les mécanismes utilisés par l'industrie pharmaceutique et l'industrie du médicament.

Les mêmes organismes, essentiellement publics, de recherche, doivent être mobilisés en amont pour assurer un cadre méthodologique strict, propre à garantir in fine aux professionnels de santé le fonctionnement de dispositifs médicaux logiciels aux propriétés vérifiées et affichées.

La prévention en continu peut s'appuyer sur des objets connectés. Mais la qualité de la mesure, le caractère acceptable des marges d'erreur dans ces prises de mesure doivent être connus et acceptés des médecins. Les algorithmes embarqués, ou accédés via Internet depuis ces objets, doivent être certifiés avec la même rigueur que les médicaments et les dispositifs médicaux.

Entre la préconisation d'un algorithme, ou l'alerte qu'il génère, et la modification dans les comportements du patient, existe une chaîne d'acteurs et de systèmes très différente d'un système à l'autre. Qu'il s'agisse du système au sens système de santé (différent d'un pays à l'autre, d'un Etat à l'autre, d'une assurance à l'autre), mais aussi du « système culturel ».

L'utilisation de techniques d'intelligence artificielle explicable doit être incitée. La recherche publique doit soutenir le développement de techniques d'intelligence artificielle explicable.

Un accent particulier doit être mis sur l'utilisabilité : le meilleur modèle basé intelligence artificielle, s'il n'est pas intégré dans un logiciel ergonomique intégré aux process du quotidien, ne sera pas utilisé, et ne délivrera pas ses bénéfices.

Les principales attentes du terrain : la Data et l'intelligence artificielle pour une prévention personnalisée au quotidien

La prévention massive, notamment par l'intermédiaires des médias est souvent considérée comme peu efficace et comme extrêmement couteuse.

Les médecins interviewés dans le cadre de cette étude ont confirmé ces observations : le Docteur Lernout, chirurgien vasculaire au Centre Hospitalier de Valenciennes explique que « (…) la prévention a d'abord été très générale. Le résultat est qu'elle a couté très cher et qu'elle était peu productive. Suite à cela, des mesures extrêmes avaient même été prises puisque les députés avaient par exemple voté le non-remboursement des mammographies pour dépister le cancer du sein » (62).

Le Docteur Grosbois, médecin travaillant chez FormAction Santé, spécialisé dans la réhabilitation et l'éducation du patient atteint de pathologies chroniques, également interviewé dans le cadre de ce mémoire, confirmait ce point en disant que la prévention primaire continuera d'être diffusée via les canaux médiatiques mais que la prévention secondaire et tertiaire se doit d'être ciblée et personnalisée : « Dans le futur, les bases de données vont permettre de sélectionner des strates de la population pour cibler au mieux les populations concernées. La prévention primaire passera toujours par les médias parce qu'elle est globale » (57). Dans ce passage de l'interview, le Docteur Grosbois aborde le fait que le traitement des Big Data permet de cibler des parties de la population non pas au niveau régional ou départemental mais de se limiter à des territoires de santé. Ces territoires de santé seront identifiés via l'analyse des Big Data et peuvent se trouver au niveau d'un quartier, d'un bassin de population etc. Toutes les personnes rencontrées ont effectivement affirmé que c'est grâce à cette stratification de la population que la prévention sera efficiente. La stratification de la population permet de cibler une typologie d'individus dont les comportements seront semblables et qui devraient être répondre positivement à une certaine forme de communication. Si la prévention est rarement rentable selon les métriques classiques (budgétaires, de court terme), elle demeure néanmoins attractive et offre de bons retours sur investissement sous réserve qu'elle soit correctement évaluée et ciblée. Par conséquent, une prévention ciblée pourrait être plus efficiente que certains traitements curatifs (35).

Une évolution probable de l'organisation de la prévention

Dans le futur, la médecine personnalisée dite médecine 4P prendra probablement une place de plus en plus importante au sein des systèmes de santé, au même titre que la prévention personnalisée, qui en est une composante. Les technologies étant également en constante évolution, on pourrait bientôt disposer de bases centralisées, sécurisées et alimentées en temps réel, de serveurs de calculs ultra-rapides permettant de calculer des modèles algorithmes beaucoup plus rapidement et pourquoi pas aussi, un jour, en temps réel.

Il reste à espérer qu'un filtre éthique s'imposera toujours en amont, distinguant, suivant la position de Kant, ce qui est possible techniquement de ce qui est moralement souhaitable. Les initiatives, en France, comme ETHIK-intelligence artificielle, le cadre proposé par le Comité Consultatif National d'Ethique, les positions du Conseil d'Etat, et d'autres travaux à visée normative, vont en ce sens.

Bibliographie

1. G. Roddenberry. Star trek. 1966.

2. P. Verhoeven. RoboCop. 1987.

3. A. Niccol. Bienvenue à Gattaca. 1997.

4. M. Bay. The Island. 2005.

5. M. Williams. Scanadu Scout : a real-life « tricoder » brings a touch of « Star Trek » to medicine. PCWorld. 2015;

6. Qu'est-ce que la prévention santé ? [Internet]. Previssima. [cité 14 août 2019]. Disponible sur: https://www.previssima.fr/question-pratique/quest-ce-que-la-prevention-sante.html

7. Ehrlich I, Becker GS. Market Insurance, Self-Insurance, and Self-Protection. J Polit Econ. 1 juill 1972;80(4):623-48.

8. Gordon RS. An operational classification of disease prevention. Public Health Rep. 1983;98(2):107-9.

9. Cai X, Nie F, Huang H. Multi-View K-Means Clustering on Big Data. In: Twenty-Third International Joint Conference on Artificial Intelligence [Internet]. 2013 [cité 5 août 2019]. Disponible sur: https://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6979

10. JL. Mathieu. Prévention : quels enjeux de santé publique ! L'Harmattan. 2018. (Analyse de politiques publiques).

11. D. Pastre. intelligence artificielle. 2000;

12. Qu'est ce que l'intelligence artificielle. Perspect Sanit Soc. nov 2017;255:41.

13. Les grandes étapes de l'histoire de la prévention : Millenaire 3, Société [Internet]. [cité 22 août 2019]. Disponible sur: https://www.millenaire3.com/Interview/2005/les-grandes-etapes-de-l-histoire-de-la-prevention

14. pubmeddev. Home - PubMed - NCBI [Internet]. [cité 21 août 2019]. Disponible sur: https://www.ncbi.nlm.nih.gov/pubmed/

15. CISMeF. Constructeur de Requêtes Bibliographiques Médicales [Internet]. [cité 21 août

2019]. Disponible sur: https://crbm.chu-rouen.fr/querybuilder/

16. PRISMA [Internet]. [cité 21 août 2019]. Disponible sur: http://www.prisma-statement.org/

17. VOSviewer - Visualizing scientific landscapes [Internet]. VOSviewer. [cité 22 août 2019]. Disponible sur: https://www.vosviewer.com//

18. CHABIN PM-A. Données structurées et données non structurées [Internet]. Arcateg, méthode d'archivage par catégorie. 2018 [cité 13 août 2019]. Disponible sur: https://www.arcateg.fr/2018/09/06/donnees-structurees-et-donnees-non-structurees/

19. Cartographie des bases de données publiques en santé - data.gouv.fr [Internet]. [cité 21 août 2019]. Disponible sur: /fr/datasets/cartographie-des-bases-de-donnees-publiques-en-sante/

20. Big data en santé [Internet]. Inserm - La science pour la santé. [cité 14 août 2019]. Disponible sur: https://www.inserm.fr/information-en-sante/dossiers-information/big-data-en-sante

21. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 19 févr 2009;457(7232):1012-4.

22. Chunara R, Andrews JR, Brownstein JS. Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak. Am J Trop Med Hyg. 1 janv 2012;86(1):39-45.

23. Quantin C, Coatrieux G, Fassa M, Breton V, Vlieger P de, Bourquard K, et al. Gestion décentralisée des documents médicaux des patients. Doc Numer. 2009;Vol. 12(3):23-35.

24. Des ressources humaines trop peu nombreuses [Internet]. IT for Business. 2017 [cité 13 août 2019]. Disponible sur: https://www.itforbusiness.fr/des-ressources-humaines-trop-peu-nombreuses-15464

25. Lu C, Wu Z, Liu M, Chen W, Guo J. A patient privacy protection scheme for medical information system. J Med Syst. déc 2013;37(6):9982.

26. Santé | CNIL [Internet]. [cité 14 août 2019]. Disponible sur: https://www.cnil.fr/fr/sante

27. Leclerc X. Le RGPD est-il appliqué au secteur de la santé ? | DPMS [Internet]. Mise en conformité RGPD. 2018 [cité 14 août 2019]. Disponible sur: https://www.dpms.eu/rgpd/rgpd-donnees-sante/

28. Données de santé : les Français sont prêts à les partager à quelques conditions [Internet].

Odoxa. [cité 14 août 2019]. Disponible sur: http://www.odoxa.fr/sondage/convaincus-potentiel-de-leurs-donnees-de-sante-francais-prets-a-partager-a-quelques-conditions/

29. Sondage BVA : les Français sont prêts pour la digitalisation de la santé – La Veille des acteurs de la Santé – L'info santé à la source ! [Internet]. [cité 15 août 2019]. Disponible sur: https://www.veille-acteurs-sante.fr/2018/01/18/sondage-bva-les-francais-sont-prets-pour-la-digitalisation-de-la-sante/

30. Mazars. Santé : 1 Français sur 2 favorable à l'utilisation de l'Intelligence Artificielle pour interpréter les données médicales - Mazars - France [Internet]. [cité 15 août 2019]. Disponible sur: https://www.mazars.fr/Accueil/News/Communiques-de-presse/Enquete-Les-Francais-la-prevention-augmentee

31. Comment se décide une autorisation de mise sur le marché (AMM) ? [Internet]. [cité 16 août 2019]. Disponible sur: https://www.leem.org/comment-se-decide-une-autorisation-de-mise-sur-le-marche-amm

32. La mise sur le marché d'un dispositif médical - ANSM : Agence nationale de sécurité du médicament et des produits de santé [Internet]. [cité 18 août 2019]. Disponible sur: https://www.ansm.sante.fr/Dossiers/Dispositifs-medicaux/La-mise-sur-le-marche-d-un-dispositif-medical/(offset)/1

33. Medical Imaging Cloud AI [Internet]. Arterys. [cité 21 août 2019]. Disponible sur: https://www.arterys.com/

34. Règlement (UE) 2017/745 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) n° 178/2002 et le règlement (CE) n° 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE (Texte présentant de l'intérêt pour l'EEE.) [Internet]. OJ L, 32017R0745 mai 5, 2017. Disponible sur: http://data.europa.eu/eli/reg/2017/745/oj/fra

35. PAMELA – Personalized and decentrAlized MachinE Learning under constrAints [Internet]. [cité 19 août 2019]. Disponible sur: https://project.inria.fr/pamela/

36. GCS Hugo [Internet]. [cité 21 août 2019]. Disponible sur: http://www.chu-hugo.fr/fr/

37. Barnay T. Prévention et vieillissement en bonne santé : un regard économique. Gerontol Soc. 18 août 2012;35 / HS n° 1(5):185-203.

38. S. Loubière JMT. L'économie de la prévention. 2005;

39. avis_prevention_-_hcaam_-_25_octobre_2018.pdf [Internet]. [cité 8 août 2019]. Disponible sur: https://www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/avis_prevention_-_hcaam_-_25_octobre_2018.pdf

40. cns18.pdf [Internet]. [cité 29 août 2019]. Disponible sur: https://drees.solidarites-sante.gouv.fr/IMG/pdf/cns18.pdf

41. 190425_Dossier_virage_numerique_masante2022.pdf. In [cité 13 août 2019]. Disponible sur: https://esante.gouv.fr/sites/default/files/media_entity/documents/190425_Dossier_virage_numerique_masante2022.pdf

42. S. Sargentini. Big-Data-AH8.pdf [Internet]. [cité 13 août 2019]. Disponible sur: http://www.avenir-hospitalier.fr/images/AH8/Big-Data-AH8.pdf

43. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sørensen HT. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. BMC Med Res Methodol. 28 mai 2011;11:83.

44. Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. Pers Med. 2013;10(6):565-76.

45. Leonelli S, Spichtinger D, Prainsack B. Sticks and carrots: encouraging open science at its source. Geo. 2015;2(1):12-6.

46. Abrantes P, Sabatier S, Guenot C. Centre technique d'appui et de formation des Centres d'examens de santé. 2005;18.

47. Sheikh Z, Hoeyer K. "That is why I have trust": unpacking what 'trust' means to participants in international genetic research in Pakistan and Denmark. Med Health Care Philos. 2018;21(2):169-79.

48. Deloitte_Transformationsplan.pdf [Internet]. [cité 12 août 2019]. Disponible sur: https://www.sum.dk/Sundhedsprofessionelle/Sundhedsdataprogrammet/~/media/Filer%20-%20dokumenter/Sundhedsdataprogrammet/Deloitte_Transformationsplan.ashx

49. Darimont B, Liu D. Le système de santé de la République populaire de Chine: entre privatisation et système public de santé. Rev Int Sécurité Soc. 2013;66(1):105-25.

50. Zhang S. Big Data for Healthcare in China: a Review of the State-of-art. 10(1):9.

51. China produces more AI papers than EU or US, but their impact is lower [Internet]. Science|Business. [cité 22 août 2019]. Disponible sur: https://sciencebusiness.net/news-byte/china-produces-more-ai-papers-eu-or-us-their-impact-lower

52. 137-AifangMA_intelligence artificielleenChine_2018-10-29_w.pdf [Internet]. [cité 4 août 2019]. Disponible sur: http://www.fondapol.org/wp-content/uploads/2018/11/137-AifangMA_intelligence artificielleenChine_2018-10-29_w.pdf

53. He H, Pan L, Pa L, Cui Z, Ren X, Wang D, et al. Data Resource Profile: The China National Health Survey (CNHS). Int J Epidemiol. 01 2018;47(6):1734-1735f.

54. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring Influenza Epidemics in China with Search Query from Baidu. PLOS ONE. mai 2013;8(5):e64323.

55. Chinese Center for Disease Control and Prevention [Internet]. [cité 20 août 2019]. Disponible sur: http://www.chinacdc.cn/en/aboutus/orc_9362/201809/t20180926_194258.html

56. Kaiser Permanente® | UC Berkeley | University of California [Internet]. [cité 21 août 2019]. Disponible sur: https://my.kp.org/universityofcalifornia/plan-details/uc-berkeley/

57. Netgen. Un grand réseau de soins médicaux intégrés : l'exemple de Kaiser Permanente en Californie. Témoignage d'un praticien [Internet]. Revue Médicale Suisse. [cité 20 août 2019]. Disponible sur: https://www.revmed.ch/RMS/2007/RMS-99/32063

58. Sandhoff BG, Kuca S, Rasmussen J, Merenich JA. Collaborative Cardiac Care Service: A Multidisciplinary Approach to Caring for Patients with Coronary Artery Disease. Perm J. 2008;12(3):4-11.

59. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 1 déc 2017;2(4):230-43.

60. Grosbois JM. Big Data et prévention personnalisée : l'exemple d'un cabinet de réhabilitation et d'éducation patient. (Disponible sur demande).

61. Accueil [Internet]. [cité 21 août 2019]. Disponible sur: https://www.mipih.fr/

62. US CLOUD Act and International Privacy [Internet]. Data Protection Report. 2019 [cité 21 août 2019]. Disponible sur: https://www.dataprotectionreport.com/2019/08/u-s-cloud-act-and-international-privacy/

63. Analytique, Business Intelligence et Data Management [Internet]. [cité 21 août 2019].

Disponible sur: https://www.sas.com/fr_fr/home.html

64. Khalife. Big Data et prévention personnalisée : l'exemple d'Arterys (disponible sur demande).

65. Lernout. Big Data et prévention personnalisée : l'exemple du passeport vasculaire (disponible sur demande).

Table des matières

Remerciements **2**

Sommaire **3**

Table des illustrations **5**

Introduction **6**

La prévention grâce aux Big Data, une fiction… ancienne 6

La fiction qui inspire la réalité 6

La démocratisation du quantified self : la prévention basée Data pour tous ? 7

Présentation du sujet 7

Cadrage du sujet, limites de l'étude 7

Organisation du présent mémoire 8

I. Définitions préliminaires **9**

1.1 La prévention, un concept difficile à définir 9

1.2 Les Big Data 11

1.3 L'intelligence artificielle 11

II. Quelles utilisations de la Big Data et de l'intelligence artificielle dans la prévention ? **12**

2.1 La prévention dans l'histoire récente de la médecine et dans l'organisation de la santé, en France 12

2.1.1 L'apparition du « Mieux vaut prévenir que guérir » dans les politiques publiques 12

2.1.2 L'avant-guerre 12

2.1.3 L'entre-deux guerres 14

2.1.4 Le développement de la santé publique dès la fin de la seconde guerre mondiale 14

2.2 Bibliographie synthétique : la prévention grâce aux Data et à l'intelligence artificielle, dans l'activité scientifique après les années 1950 15

2.2.1 Méthodologie : construction des requêtes 15

2.2.1.1 Outils utilisés pour construire la revue des publications scientifiques 15

2.2.1.2 Les mots clés recherchés et construction de la requête 16

2.2.1.3 Discussion 17

2.2.1.4 Les limites de cette revue scientifique 18

2.3 Les Big Data : quelles données candidates dans le secteur santé, pour la prévention ? 19

2.3.1 Les data disponibles.....20

2.3.1.1 Quels types d'informations sont disponibles ? 20

2.3.1.2 Comment les utiliser ?..... 23

2.3.2 Quels sont les freins actuels quant à l'utilisation des big data de santé, pour la prévention ? 23

2.3.2.1 Une infrastructure technique particulière est nécessaire 24

2.3.2.2 Des compétences particulières, et rares, sont requises 24

2.3.2.3 Autres freins : la faible incitation des professionnels à l'utilisation des outils ; la faible utilisabilité des systèmes d'information en santé 24

2.3.2.4 La réglementation française..... 25

2.3.3 Les Français sont-ils opposés à l'utilisation de leurs données personnelles 27

2.3.4 Comment gagner la confiance des professionnels de santé dans les outils basés Data et intelligence artificielle, pour la prévention ? 28

2.3.4.1 Mise sur le marché des médicaments et des dispositifs médicaux. 28

2.3.4.2 Les algorithmes et l'intelligence artificielle, besoin d'être certifiés ? 29

2.3.4.2.1 Dans les autres pays : l'avis d'ARTERYS, un acteur de l'analyse d'image basée intelligence artificielle, présente dans plusieurs pays.....29

2.4 En conclusion de cette partie 32

III. Quels impacts pourraient avoir ces outils de prévention sur l'économie globale de la santé ?.....33

3.1 La prévention en France 33

3.1.1 Mise en place de la prévention en France 33

3.1.2 Financement de la prévention.....34

3.1.3 La prévention et le numérique, au cœur du système de santé 35

3.2 La prévention dans d'autres systèmes que le système français.....40

3.2.1 Le modèle danois 40

3.2.2 Le modèle chinois 42

3.2.2.1 Les réformes de santé en Chine 42

3.2.2.2 Les big data, une affaire d'Etat 44

3.2.3 Un modèle américain : le Kaiser Permanente 46

3.2.3.1 Le Kaiser Permanente : la naissance du prépaiement 46

3.2.3.2 Le Kaiser Permanente actuel 47

3.2.3.3 « Collaborative Cardiac Care Service », un exemple de prévention tertiaire chez des patients atteints de coronaropathie. 50

IV. La prévention personnalisée constitue-t-elle un nouveau marché ? 51

4.1 La demande 52

4.1.1 La demande des patients : rester en bonne santé, mieux se connaître, éviter des incidents, suivre finement sa maladie chronique, fournir simplement à son médecin des informations, s'auto-éduquer 52

4.1.2 Pour les professionnels de santé : gagner en rendement des actions de prévention 53

4.1.3 Pour l'état : améliorer la « coût-efficience » des mesures de prévention 53

4.1.4 Illustration par une expérience régionale : le passeport vasculaire© Centre Hospitalier de Valenciennes 54

4.2 L'offre 55

4.2.1 Les offres disponibles et les business models associés 56

4.2.2 Illustrations 58

4.2.2.1 Arterys 58

4.2.2.2 Le passeport vasculaire 60

Conclusion 61

Prévention santé, basée Data et intelligence artificielle : une activité en croissance61

Prévention santé, basée Data et intelligence artificielle : de nombreux défis à relever 61

Défi 1 : Les systèmes de prévention, qu'ils soient basés Data ou non, sont difficiles à dupliquer d'un contexte à un autre 61

Défi 2 : Le défi de la centralisation des données pour produire des modèles efficaces dans un grand nombre de contextes, malgré l'hétérogénéité des données *62*

Défi 3 : Gagner la confiance et l'adhésion des utilisateurs finaux de ces systèmes, grâce à la certification et à l'ergonomie *63*

Les principales attentes du terrain : la Data et l'intelligence artificielle pour une prévention personnalisée au quotidien 64

Une évolution probable de l'organisation de la prévention 65

Bibliographie..... 66

Table des matières 72

Valentine DEMOERSMAN

LA PERSONNALISATION DE LA PREVENTION GRACE AUX BIG DATA

La prévention est souvent jugée comme étant inefficace et extrêmement coûteuse dans le système de santé français. Toutefois elle est primordiale pour sensibiliser la population sur des thématiques de santé publique. L'essor des nouvelles technologies a permis le développement des analyses des Big Data, notamment dans le domaine de la santé. Algorithmes, Intelligence Artificielle et autres méthodes statistiques ont pu mettre en évidence que la prévention peut être personnalisable grâce aux données de santé. Afin de comprendre les différents aspects de la politique de santé française, il est important de la comparer à d'autres systèmes de santé : le modèle danois, chinois et le Kaiser Permanente américain et d'identifier des modèles d'évolution de notre système de santé actuel. Enfin, nous nous sommes demandés si les logiciels de traitement des Big Data pour personnaliser la prévention pouvaient représenter une nouvelle opportunité commerciale. Nous avons étudié les besoins des patients, des professionnels de santé et de l'Etat ainsi que ce que les entreprises privées étaient capable de proposer en retour et leurs business models. Nous avons pour cela pris l'exemple de deux outils, Arterys et Passeport Vasculaire.
Mots clés : Big Data ; Prévention ; Intelligence artificielle ; Business Model

---

Prevention is often judged to be ineffective and extremely costly in the French health system. However, it is essential to raise public awareness on public health issues. The rise of new technologies has enabled the development of big data analytics, particularly in the health sector. Algorithms, Artificial Intelligence and other statistical methods have shown that prevention can be personalized thanks to health data. In order to understand the different aspects of French health policy, it is important to compare it to other health systems: the Danish, Chinese and American Kaiser Permanente models and to identify evolutionary models of our current health care system. Finally, we wondered whether big data processing software to personalize prevention could represent a new business opportunity. We studied the needs of patients, healthcare professionals and the state as well as what private companies were able to offer in return and their business models. For this study we took the example of two tools, Arterys and Vascular Passport.
Key-words : Big Data, Prevention, Artificial Intelligence, Business Model