UNIVERSITÉ DE NANTES
UFR de Sciences Pharmaceutiques et Biologiques

Année 2020 N° 07

THÈSE

Pour le

DIPLÔME D'ÉTAT

DE DOCTEUR EN PHARMACIE

Par

Antoine PERPOIL

Présentée et soutenue publiquement le 13 mars 2020

Dans quelle mesure l'intelligence artificielle pourrait-elle permettre d'optimiser la sécurité d'utilisation des biosimilaires ?

Président du jury :
Pr. S. BIRKLE, PU, UFR des Sciences Pharmaceutiques et Biologiques, CHU de Nantes

Membres du jury :
Mr. G. GRIMANDI, PU-PH, Doyen de la Faculté de Pharmacie de Nantes

Dr. F. BOCQUET, MCU-PH, UFR des Sciences pharmaceutiques et Biologiques, Paris Descartes

Mme. A. CHIFOLEAU, PH – Unité fonctionnelle de Pharmacovigilance, CHU de Nantes

Mr. JF. SIMONET : Directeur Compliance, Délégué à la protection des données personnelles Amgen France

Remerciements

Professeur Stéphane BIRKLE,

Qui me fait l'honneur de présider ce jury de thèse ainsi que pour ses enseignements qui m'ont poussé à choisir la thématique des médicaments biologiques similaires.

Docteur François BOCQUET,

Qui m'a encadré pour ce travail par de précieux conseils. Ces derniers m'auront permis d'orienter ma réflexion et d'identifier les éléments clés relatifs à mon sujet.

Docteur Anne CHIFOLEAU

Qui m'a guidé lors de mon externat à la Direction de la recherche. Cette première expérience m'aura permis d'acquérir de précieuses notions en pharmacovigilance.

Docteur Gael GRIMANDI,

Qui a dispensé des enseignements relatifs à l'encadrement des dispositifs médicaux. Ces derniers m'auront permis de développer certains mécanismes quant à la lecture des textes législatifs.

Jean-François SIMONET,

Qui m'aura accompagné tout au long de cette année. De nombreuses opportunités m'auront été proposées et je lui suis reconnaissant d'avoir placé sa confiance en moi.

Mes parents,

Pour leur soutien perpétuel durant toute ma scolarité, en particulier pour cette année riche en évènements.

Mattéo,

Pour son aide précieuse.

Table des matières

Introduction

Les médicaments biologiques ont révolutionné la prise en charge des patients, en particulier dans le domaine de l'oncologie. Les médicaments biologiques similaires, copies de médicaments biologiques de référence, contribuent au progrès thérapeutique. Il existe toutefois des questionnements quant à ces médicaments biologiques similaires également désignés par le terme de biosimilaires.

Les questionnements ont plusieurs origines à commencer par la structure de ces biomédicaments. Ces biosimilaires sont particuliers d'un point de vue structural[1] et ne sont ainsi pas comparables aux génériques, copie de médicament d'origine chimique. En effet, l'origine biologique engendre une variation structurale de tout médicament biologique[3] qui bien que maitrisée soulève des questions. Ces questions sont notamment celles de l'impact de la variabilité structurale sur la survenue de réactions immunitaires. Or cette réaction immunitaire est complexe à évaluer car elle présente plusieurs causes, dont la variabilité du produit[5,8], mais également la pathologie cible, la population[6] ou les pratiques.

Ce risque de réaction immunitaire est donc étudié pendant les phases de développement afin de mettre sur le marché un biomédicament sûr, de qualité et efficace[10]. Des études précliniques et des essais cliniques sont mis en place afin d'évaluer la biosimilarité d'un biosimilaire. En d'autres termes, des études sont conduites pour démontrer que la copie d'un biomédicament est hautement similaire au médicament biologique de référence[11].

Mais l'évaluation du biomédicament pendant les phases de développement présente des limites[12,13]. Les essais cliniques notamment vont étudier la survenue de ces réactions immunitaires mais pour une durée déterminée. Toute réaction qui surviendrait au long cours serait donc pas étudiée. A cette limite temporelle s'ajoute la possibilité pour un biomédicament similaire de bénéficier des informations obtenues pour le biomédicament de référence. Il s'agit ainsi d'une opportunité d'extrapoler les données de la référence en utilisant le biosimilaire pour une indication pour laquelle il n'aurait pas été étudié.

Aux termes de son évaluation, le biosimilaire obtient une autorisation de mise sur le marché (AMM) lui permettant d'être utilisé en dehors d'un essai clinique. La biosimilarité démontrée permet ainsi d'obtenir l'AMM et le biosimilaire peut dès lors être utilisé en soins

courants. De la même façon, les pratiques médicales pourraient soulever des questions supplémentaires quant au risque de réaction immunitaire.

Lors d'une utilisation en soins courants, un biomédicament peut être prescrit à un patient pour une pathologie qui requiert un traitement au long cours. Or ce traitement implique des renouvellements de prescription et une utilisation sur une longue période du même biomédicament. Cependant, il est possible pour le praticien de prescrire un biomédicament différent de celui initialement prescrit pour le patient. Cette pratique, appelée *switch* ou interchangeabilité[16], permet ainsi au praticien dans certains cas de figure de prescrire un biosimilaire à la place du biomédicament de référence. Il se pose toutefois une question de la tolérance du patient vis-à-vis du traitement. En effet, ces changements itératifs peuvent parfois – même si cela doit faire l'objet d'une analyse prudente - être à l'origine de réactions immunitaires et conduire à une rupture de tolérance

Si ce risque de rupture de tolérance est étudié par des études de *switch*, il est possible de s'interroger sur leur caractère suffisant. Comme les essais cliniques, ces études peuvent présenter des limites sur la durée de surveillance, le nombre de patient inclus ou encore les paramètres faisant l'objet d'une analyse. Cette limite serait d'autant plus importante que l'utilisation des biosimilaires est croissante.

Les biosimilaires permettent de réaliser des économies pour le système de Sécurité sociale[15]. A noter que des incitations à utiliser le biosimilaires ont été mises en place récemment destinées aux établissements de santé mais aussi aux médecins libéraux[17]. La question d'un droit de substitution pour augmenter la diffusion des biosimilaires et les économies générées a également fait l'objet de nombreuses hésitations de la part du législateur. En effet, le PLFSS de 2014 prévoyait un droit de substitution par le pharmacien du biomédicament de référence par son biosimilaire en initiation de traitement. Toutefois, ce droit de substitution a été abrogé par le projet de loi du PLFSS 2020 notamment pour des raisons de sécurité sanitaire[19]. La question de la substitution pharmaceutique ne manquera sans doute pas de revenir dans les débats tant les enjeux économiques et sociétaux en présence sont importants.

Il apparait ainsi que plusieurs éléments spécifiques aux biosimilaires peuvent jouer sur la survenue d'une réaction immunitaire. Des informations supplémentaires seraient nécessaires

pour répondre à ces questionnements et ainsi compléter les données des études cliniques et des études de *switch*. Dans la mesure où ces informations seraient obtenues après la mise sur le marché, l'analyse porterait sur des données de vie réelle.

Les données de vie réelle sont toutes les données liées à la prise en charge d'un patient en dehors des études cliniques. Une analyse de ces données pourrait ainsi apporter les informations liées à ces questionnements. Il y aura dès lors une phase indispensable de collecte des données issues de la prise en charge des patients[20]. Cette collecte de données impliquera une capacité d'analyse de données massives[21].

Ces données massives impliquent d'obtenir des données en très grand volume, de nature très variée, avec un flux important[22]. Ces données sont intéressantes car riches en information, permettant un suivi d'une population hétérogène, et intégrant de nombreuses composantes liées aux biosimilaires. Il s'agit donc d'une première pierre à l'édifice, puisqu'elles seraient à l'origine de la capacité de l'analyse à répondre aux questionnements de la communauté de santé.

Ces données sont, quoi qu'il en soit, stockées dans des bases[23] auxquelles l'initiateur de l'analyse devrait avoir accès. Il existe en l'occurrence différentes natures de bases, qui ne présentent pas les mêmes caractéristiques. En effet, ces bases peuvent varier dans leur format, la nature des données qu'elles contiennent et même dans les conditions d'accès[24]. Nous pouvons citer notamment les registres, les dossiers médicaux et les bases médico-administratives.

Pour fournir de l'information, ces bases doivent être exploitées dans le cadre d'une étude. Il peut être envisagé de conduire des analyses conjointes ou isolées de ces bases. Quoi qu'il en soit, une méthode appropriée capable de traiter ces données massives sera requise. Or des méthodes classiques pourraient ne pas être à même de réaliser cette tâche. C'est la raison pour laquelle est discuté dans la présente thèse l'intérêt de l'utilisation d'une méthode d'intelligence artificielle[28] pour répondre à notre problématique.

Cette méthode présente l'intérêt de réaliser une tâche, en fonction d'instructions qu'elle aura développées elle-même[29]. Au regard de la problématique des biosimilaires, la méthode de l'apprentissage machine paraît intéressante. Ce dernier réalise une tâche de classification, en associant des données de prise en charge à un résultat clinique. En faisant cela, la méthode

est ensuite capable d'identifier ce qui lui a permis de classer[31]. Ainsi dans un contexte où le résultat clinique est connu, l'apprentissage machine peut-être envisagé pour trouver un lien entre les données. Ce lien permettrait ainsi de mettre en évidence des facteurs de risques qui concourent à la survenue d'une réaction immunitaire.

Cette analyse par l'apprentissage machine s'intègre dans une étude de données de vie réelle qui vise à compléter les essais cliniques. La conception de ces études devra se faire dans le respect de la réglementation, notamment la loi Jardé[25]. Il faudra toutefois considérer précocement que la conception de l'étude déterminera ce cadre réglementaire applicable à l'étude.

En effet, une étude qui analyse des données une fois que la prise en charge est terminée, dite rétrospective, n'aura pas le même cadre qu'une étude qui collecte de nouvelles données, dite prospective[25]. L'initiateur de l'étude devra donc déterminer s'il mène une étude rétrospective ou prospective selon les questions auxquelles il souhaite répondre. Une étude rétrospective serait ainsi plus adaptée pour comprendre les causes alors qu'une étude prospective permettrait d'anticiper des événements. Le traitement de données personnelles, et de données de santé impose également le respect d'un cadre réglementaire particulier. L'initiateur de l'étude sera ainsi amené à démontrer qu'il n'exerce un traitement que sur des données anonymisées. Cette anonymisation implique plusieurs étapes de préparation des bases de données.

La conception d'une étude de données de vie réelle intègre le choix d'une méthode d'analyse. Dès lors, le fonctionnement d'un modèle d'intelligence artificielle sera présenté. Il s'agit d'une méthode qui réalise des tâches définies par des instructions intégrées dans un algorithme. L'apprentissage machine permet par principe un développement de ses propres instructions selon les données qu'il traite. Il y a donc une première phase nécessaire, dite d'apprentissage, où le modèle est soumis à des données pour créer les instructions[28].

Durant l'apprentissage, le modèle propose un résultat selon les données et le compare au résultat réel observé. Par cette comparaison, le modèle corrige ses propres instructions afin d'aboutir à des résultats corrects. Une fois qu'il est éduqué, le modèle d'apprentissage machine est utilisé sur les bases de données sélectionnées pour l'étude. Il peut ainsi reproduire son traitement. Dès lors, nous pourrions souhaiter utiliser un modèle

d'apprentissage machine, tel qu'un réseau de neurones, sur les données de vie réelle des biosimilaires. Le réseau de neurones est un ensemble d'algorithme effectuant des opérations mathématiques afin de transformer des données utilisées. Dans ce document, l'utilisation du réseau de neurones a pour but d'apporter des réponses aux questionnements sur les biosimilaires. L'utilisation d'une méthode d'intelligence artificielle peut ainsi être valorisée de plusieurs façon. En effet, un modèle d'intelligence artificielle peut permettre d'identifier des tendances dans les données et de proposer un résultat qui revient à classer les patients dans des catégories.

Dans un premier temps, il s'agit d'explorer et de comprendre les éléments qui jouent un rôle important dans la réaction immunitaire liée à l'utilisation d'un biomédicament de référence et de son biosimilaire. Pour cela, il est possible de répondre à une première question telle que l'impact de l'interchangeabilité sur le risque de réaction immunitaire. Cette question implique de déterminer les types de résultats que le modèle peut fournir. En effet, le modèle pourra classer les patients selon qu'ils ont un risque ou non d'avoir une réaction immunitaire. Cette classification pourrait toutefois être plus complexe. Une fois qu'elle est faite, le modèle apprendra à classer correctement les patients selon les données de prises en charge dans une catégorie.

Une fois que cela est fait, nous souhaitons connaitre le raisonnement qui a conduit à cette classification. En effet, un réseau de neurones opère un traitement qui est d'associer les données les unes aux autres pour obtenir un résultat. Cette association se traduit dans les faits par de interactions qui pourraient permettre d'identifier des facteurs de risques significatifs. Par exemple, il serait possible de déterminer si le *switch* au bout de 6 mois de traitement est plus risqué qu'un *switch* un mois après l'initiation.

Mais il s'agit d'une première utilisation de l'intelligence artificielle qui est dépendante de données soumises au modèle. Des données supplémentaires pourraient être soumises selon les cas de figure. Ainsi, l'analyse pourrait être répétées 3 ans après la première analyse mais également suite à la mise sur le marché d'un nouveau biosimilaire. Il s'agit ici d'une première déclinaison de l'IA selon les spécificités de biosimilaires. Le modèle d'apprentissage machine d'une première analyse pourrait être mis à profit par les analyses suivantes. Il y aurait dès lors une valorisation du modèle développé, et une exploitation conjointe de résultats obtenus lors de chaque analyse.

Bien que le droit de substitution du pharmacien dans le cas des biosimilaires soit aujourd'hui abrogé, de nouvelles analyses plus complexes peuvent être envisagées sur la mise en pratique de cette substitution. D'autres analyses quant à elles pourraient porter sur de nouvelles données spécifiques à un problème. Il s'agit d'une opportunité pour intégrer des données propres à une problématiques telles que les données génétiques pour évaluer la prédisposition de réponse à un traitement selon le terrain génétique.

Il serait également possible d'exploiter le second versant de l'intelligence artificielle. Cette deuxième possibilité correspondrait à l'exploitation de la capacité d'un réseau de neurones à classer un résultat. Au lieu de se placer dans un cas de figure où la réponse est connue, le modèle pourrait être utilisé pour proposer un résultat lorsque la réponse est inconnue. Le praticien pourrait ainsi valoriser l'anticipation de l'issue d'une prise en charge au regard des donnée collectées. Le réseau de neurones profite ainsi d'un raisonnement déjà construit lors d'une analyse précédente, pour classer les patients dans les catégories définies auparavant selon le niveau de risque de réaction immunitaire.

Ces différentes utilisations de l'intelligence artificielle pourraient être prometteuses mais devront également respecter des cadres réglementaires adaptés. La puissance de calcul de la méthode devrait notamment faire l'objet d'une démonstration de conformité. En effet, l'initiateur de l'étude devrait démontrer qu'aucun traitement contraire à la loi n'est effectué par une telle méthode. De plus, il pourrait être nécessaire de démontrer que le traitement ne fragilise pas le caractère anonyme des données.

Les méthodes d'intelligence artificielle présentent également une certaine opacité dans leur traitement. Le traitement effectué par un réseau de neurones peut rendre l'interprétation difficile soulevant dès lors la question de boite noire. Or ce manque de transparence pourrait s'accompagner d'une réticence de la part de la communauté de santé à exploiter les résultats apportés par de telles méthodes. Cela limiterait ainsi l'impact des informations apportées aux questions des biosimilaires.

Trois parties seront donc développées dans ce document. Une première partie dédiée à la présentation des questionnements relatifs aux biosimilaires et aux risques de réactions immunitaires. Une seconde partie permettra de développer l'exploitation des données de vie réelles générées par l'utilisation des biosimilaires en soins courants. Une troisième partie

permettra d'exposer une réflexion quant à l'utilisation de l'intelligence artificielle, en l'occurrence l'apprentissage machine, pour apporter des éléments de réponse aux questionnements relatifs aux biosimilaires.

1. L'utilisation des biosimilaires soulève des questions sur la sécurité au long cours

a) L'impact de la variabilité des médicaments biologiques sur la survenue de réactions immunitaires

i. Définition du médicament biologique et du médicament biologique similaire

1. Les médicaments biologiques, définition et particularité structurale

Il existe une diversité de molécules utilisées en thérapeutiques qui présentent des structures variées. Ces molécules peuvent être d'origine chimique ou biologique. Les molécules chimiques sont obtenues par des réactions de synthèse produisant des molécules de petites tailles. Il est alors question de principe actif, qui sera intégré *in fine* dans une forme galénique pour obtenir un médicament. Contrairement aux molécules d'origine chimique, les molécules d'origine biologiques sont obtenues en utilisant des microorganismes. L'utilisation d'organisme vivant conduit ainsi à l'obtention de grosses molécules, notamment des protéines thérapeutiques. Ces dernières sont constituées d'une structure protéique d'acides aminés, dite séquence primaire, et de chaines de sucres attachées à la structure protéique. La production d'une telle protéine implique des procédés d'obtention plus complexes, avec par exemple des remaniements génétiques de la source biologique. Il est possible de citer par exemple les technologies d'ADN recombinant pour produire des protéines complexes[1].

La protéine thérapeutique obtenue est utilisée en tant que médicament tel que défini à l'article L.5111-1 du Code de la santé publique. La protéine est ainsi présentée comme possédant des propriétés curatives ou préventives. Mais les particularités d'obtention de cette protéine thérapeutique justifient une définition sur mesure qui conduit à la définition du médicament biologique par la directive 2001/83/CE qui définit le médicament biologique[2]. La directive 2003/63/CE apporte une distinction entre les médicaments biologiques et les médicaments chimiques. Ainsi, la définition du médicament biologique est présentée comme la résultante d'une succession de procédés techniques, ainsi que les contrôles réalisés pour garantir l'activité, l'efficacité et l'innocuité du médicament. Cette définition conduit ainsi à un cahier des charges plus complexes et donc des exigences règlementaires plus contraignantes[2] pour les médicaments biologiques.

Le premier laboratoire ayant développé le médicament biologique aura déposé un brevet pour la molécule thérapeutique d'intérêt. Le dépôt d'un brevet permet dès lors d'une protection de l'exploitation d'un produit une fois sur le marché. Seul le détenteur du brevet pourra commercialiser le médicament biologique sans imitation de la molécule thérapeutique par un concurrent. Cette période de monopole d'exploitation permet ainsi de compenser les coûts de développement engagés par le laboratoire. En effet, des coûts très importants peuvent être nécessaires pour le développement d'un biomédicament afin d'obtenir une autorisation de mise sur le marché (AMM). Il faut citer notamment les phases de modification des microorganismes, de mise au point des procédés de production ou encore d'études cliniques. Toutes ces phases de développement complexe conduisent à un prix élevé pour ces médicaments biologiques[3]. Ainsi, le brevet permet au laboratoire de compenser les coûts élevés de développement du médicament biologique. Mais une fois le brevet expiré, d'autres laboratoires concurrents pourront copier ce médicament biologique.

2. La copie des médicaments biologiques : les médicaments biologiques similaires

Un médicament biologique similaire est une copie du premier médicament biologique mis sur le marché par un laboratoire. Le premier biomédicament est qualifié de référence, tandis que la copie est qualifiée de biologique similaire, aussi appelé biosimilaire. Une fois le brevet du biomédicament de référence arrivé à expiration, il est possible de réaliser une copie de la molécule thérapeutique pour développer un médicament biologique similaire. La réalisation d'une telle copie est réglementée, le biosimilaire devra répondre à une définition précise. Cette réglementation vise à garantir que le biosimilaire soit hautement similaire à sa référence, sans différence clinique significative et équivalent en efficacité et en sécurité[4].

Les textes rédigés par l'EMA précisent qu'il est à priori possible de copier toute molécule biologique[2]. Il faut pour cela connaitre la molécule biologique de référence, sa structure, son procédé de fabrication et son contrôle [2]. Mais à la différence du biomédicament de référence, le biosimilaire bénéficiera des phases de développement moins coûteuses. Les coûts de développements sont par conséquents réduits, de même que le coût du médicament biologique. L'intérêt du biosimilaire est donc de mettre à disposition des praticiens des molécules biologiques moins coûteuses et donc d'accroitre l'accès au traitement pour les patients[3].

Il est question de copie de médicament biologique alors qu'en fait, le biosimilaire présente des différences vis-à-vis de la référence. En effet, des différences de présentations du médicament existent puisque le biosimilaire peut présenter moins d'indications que la référence, une administration différente, des particularités de préparation, de stockage ou encore d'information sur la sécurité[5]. Mais une différence qui requière d'autant plus d'attention, est celle concernant la structure même de la molécule thérapeutique d'origine biologique contenu dans le médicament biologique similaire.

3. *Le médicament biologique similaire : une copie très proche sans être identique à la référence*

Pour comprendre cette différence structurale entre le biosimilaire et sa référence, il faut reprendre le cas du médicament d'origine chimique. Le premier médicament d'origine chimique mis sur le marché prend le nom de médicament *princeps*. Tout comme le médicament biologique, le principe actif du *princeps* peut être copié pour aboutir à un médicament générique. Comme les médicaments chimiques ont des procédés très prédictibles et stables, il est possible de copier à l'identique le principe actif[1] du *princeps*. En d'autres termes, le principe actif copié, obtenu par les réactions de synthèse, aura exactement la même structure que le principe actif du *princeps*.

Toutefois, les biosimilaires ne sont pas des génériques[1]. Contrairement aux génériques, il est impossible de copier à l'identique la molécule thérapeutique d'origine biologique. En effet, les molécules biologiques sont produites par des procédés plus complexes avec lesquels il n'est pas possible de reproduire à l'identique une molécule[3]. Biosimilaires et génériques n'ont ainsi pas la même définition.

Des définitions s'appliquent donc en distinguant médicaments biologiques similaires et génériques. Sur le plan juridique, les médicaments biologiques similaires sont ainsi définis par défaut par rapport aux médicaments génériques[2], avec un régime plus strict pour les biosimilaires [2]. La directive 2004/27/CE apporte une définition du générique et du médicament biologique similaire [2]. Il faut remarquer que la définition d'un produit doit être associée aux exigences réglementaires. La directive 2001/83/CE donne les exigences réglementaires pour l'obtention du dossier d'AMM des génériques. Ces derniers ont des exigences allégées vis-à-vis des *princeps* avec notamment des études précliniques et cliniques

qui ne sont pas exigées pour prouver la bioéquivalence [2]. Les biosimilaires quant à eux ne bénéficient pas des exigences réglementaires allégées car ne rentrent pas dans la même définition. L'encadrement spécifique des biosimilaires peut être abordé *via* les particularités structurales des biomédicaments produits.

ii. La variabilité des médicaments biologiques : la micro-hétérogénéité

1. La production d'un biomédicament par des procédés biologiques

Il n'existe pas de copie parfaite d'un médicament biologique puisque les procédés sont complexes. Ces derniers intègrent des technologies complexes qui sont à l'origine de cette impossibilité de reproduction à l'identique. Dans le cas de la production d'anticorps monoclonaux, il est possible d'utiliser des technologies telles que les animaux transgéniques, les virus ou encore les lymphocytes immortalisés[6]. Ces technologies sont utilisées pour obtenir la molécule d'intérêt. Pour cela, des manipulations sont effectuées sur des microorganismes afin d'en modifier le matériel génétique dans le but d'obtenir une protéine d'une structure bien définie. En effet, il est attendu de cette structure un effet biologique thérapeutique.

Ces différentes technologies sont intégrées dans des procédés de production pour produire industriellement la molécule biologique. Deux phases de production peuvent être distinguées, une première phase d'obtention de la molécule (*Upstream*) et une seconde phase de purification (*Downstream*). Cette dernière est nécessaire car le procédé conduira à la production d'autres molécules non souhaitées telles que les impuretés. Une fois les procédés réalisés, il est nécessaire de vérifier la qualité du produit, soit sa conformité aux attentes. La qualité du biomédicament peut varier selon le bon déroulé des procédés et toute variation sur un paramètre critique du procédé pourra avoir un impact sur la qualité du produit[7]. Toutefois, sans impacter la qualité du produit fini, les procédés sont à l'origine de variation des médicaments biologiques. En effet, il est dit qu'un biosimilaires est une molécule donnée pour un procédé de fabrication donné[2].

Une notion essentielle rattachée à l'utilisation de procédés biologiques est celle de variabilité de la molécule biologique d'intérêt. Cette variabilité ne désigne pas ici les potentielles impuretés mais la structure même de la molécule d'intérêt. Cette dernière a été définie comme étant une structure protéique et des chaines de sucres rattachées à celle-ci. Or il n'est pas possible de contrôler le greffage de ces chaines de sucres sur la protéine (figure 1). Il y a

dès lors des variations de chaines de sucres, qualifiées de modifications post traductionnelles et désignées par le terme de micro-hétérogénéité. Ces dernières sont réalisées par la cellule de manière imprévisible durant le transport intracellulaire de la protéine afin de la protéger [1]. Cela conduit à des variations de structure bien que la chaine d'acides aminés soit identique [1] à celle attendue.

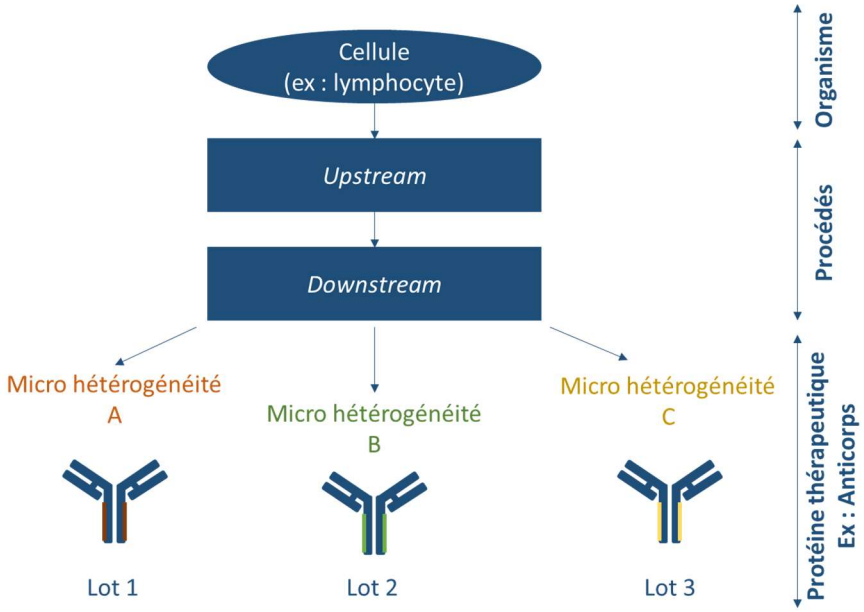

Figure 1: la micro-hétérogénéité d'une protéine

2. *La production par le vivant : micro-hétérogénéité et maitrise des dérives*

L'utilisation des procédés biologiques impliquent ainsi des variations de la molécule d'intérêt. Mais ces variations de chaines de sucres sont définies, contrôlées et tolérées dans une certaine mesure. En effet, ces variations, qualifiées de micro-hétérogénéité, sont propres à l'utilisation de microorganismes[2]. Il existe une tolérance vis-à-vis de ces variations et l'EMA accepte la micro-hétérogénéité en tant qu'elle correspond à l'augmentation de phosphorylation de structure de mannose[1] (un type de chaine de sucres). Cette micro-hétérogénéité est donc tolérée pour les médicaments biologiques de référence et les biosimilaires. Cependant, ce ne sont toutefois pas les seules variations qui peuvent survenir par l'utilisation de tels procédés.

La micro-hétérogénéité est tolérée parce qu'elle n'a pas d'impact à priori significatif sur l'effet clinique du médicament biologique. Mais il existe d'autres modifications qui pourraient être

à l'origine d'un éloignement du produit de ces caractéristiques initiales[8]. Il s'agirait en l'occurrence de modifications portant atteinte à la structure protéique, modifiant ainsi la chaine d'acide aminés. Ces modifications peuvent être causées par des altérations au cours du temps du matériel génétique d'un microorganisme utilisé pour produire une protéine d'intérêt. Dès lors le microorganisme pourrait conduire à l'obtention d'une structure de la molécule biologique différente de celle initialement prévue. Ce phénomène de dérive conduirait alors à l'obtention d'un produit dont les caractéristiques sont différentes du produit attendu[1]. Dans le cas du biosimilaire, il pourrait y avoir une perte de la similarité en l'absence de correction de la dérive [1], ce qui pourrait également causer des effets différents chez l'homme. Ainsi, ces dérives à l'origine d'un changement significatif de structure ne sont pas tolérées.

3. Le questionnement de l'impact de la micro-hétérogénéité

Il y a donc une tolérance des variations mineures du produit qui existent à chaque production du médicament biologique. Ainsi, chaque lot d'un médicament biologique de référence sera unique au regard des modifications imprévisibles des chaines de sucres. Le premier lot de l'année présentera des chaines de sucres différentes du second lot. Ainsi, il existera des différences parmi les lots de production d'un même médicament biologique de référence. Les variations de ces lots ne peuvent toutefois pas amener à considérer qu'il s'agit d'un biosimilaire [8].

De la même façon, chaque lot de biosimilaire présentera des caractéristiques uniques vis-à-vis de ces chaines de sucres. D'autant plus que le laboratoire produisant le biosimilaire aura établit un procédé de production distinct de celui du laboratoire produisant la référence [1]. Pour la référence comme pour le biosimilaire, les fabricants visent à maintenir chaque lot de produit dans les limites acceptables prédéfinies[8]. Pour cela sont utilisés une source biologique et un procédé stable[9]. Par la suite des contrôles seront effectués pour contrôler que le produit se trouve dans les limites acceptables. Ceci dans le but de démontrer l'obtention d'un médicament biologique sûr, de qualité et efficace.

Il existe une certaine tolérance concernant les variations acceptables du produit. Le biomédicament peut présenter des différences au niveau de la qualité tant qu'elles sont acceptables et justifiées[2]. A l'inverse, ne sont pas acceptables les déviations qui causent

l'émergence de différences cliniques significatives[1]. Ainsi, il peut y avoir un questionnement de la micro-hétérogénéité vis-à-vis de la sécurité d'utilisation des biosimilaires. Il peut subsister un doute sur l'impact indirect de ces variations de chaines de sucres à chaque lot sur l'émergence de différences significatives. En effet, la littérature met en évidence que les professionnels de santé s'interrogent quant au rôle que pourraient jouer ces variations structurales sur l'émergence de différences significatives[1]. En l'occurrence, il serait question d'une survenue au long cours de différences cliniques par l'administration d'un produit dont les lots sont uniques.

iii. Le risque de réaction immunitaire

1. L'effet clinique : la réaction immunitaire et ses composantes

Des évènements indésirables, tels qu'une réaction immunitaire, peuvent être déclenchés par un biomédicament. Il serait plus exact de mentionner le risque de réactions immunitaires [1] car celles-ci sont incertaines. Dans les faits, il s'agit d'une réaction de l'organisme à un produit telle qu'une réaction fulminante[6] qui diffère selon les individus. Cette réaction est délétère car elle peut à la fois entrainer des dommages sur l'organisme et détruire le médicament biologique, le rendant inefficace. Cette inefficacité peut ainsi traduire un échec thérapeutique expliqué par la production d'anticorps par l'organisme, dirigés contre le médicament. Ces anticorps sont appelés anticorps anti médicament [6]. Questionner le rôle des variations d'un biomédicament dans la survenue d'une telle réaction implique de comprendre les mécanismes de cette réaction.

Une réaction immunitaire peut avoir plusieurs causes (figure 2). Il est donc possible de traiter la notion de facteurs de risque d'immunogénicité. Ces derniers correspondent aux éléments qui ont un rôle dans la survenue d'une réaction immunitaire. La littérature regroupe les facteurs de risques en catégories : le médicament, la pathologie cible et la population[6]. Il sera donc discuté des différents éléments liés aux biomédicaments, avec notamment les variations de structures et les impuretés. Mais il serait également question des éléments liés à la pathologie et aux comorbidités, des éléments propres aux patients ainsi qu'une quatrième catégorie de facteurs de risque : les pratiques[9].

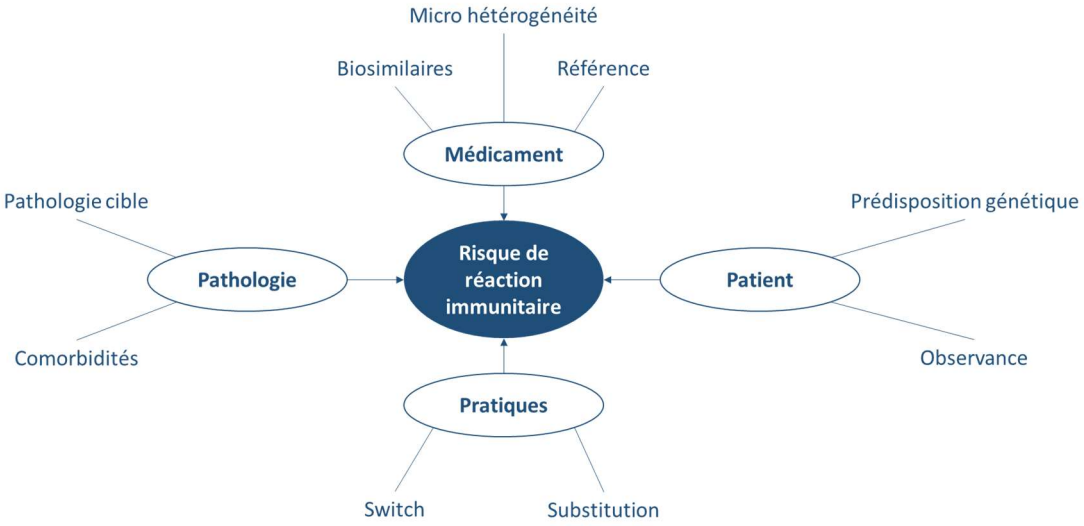

Figure 2: Les composantes de la réaction immunitaire

2. *L'impact de la variabilité du produit sur les réactions immunitaires*

La réaction immunitaire est complexe et il n'est pas aisé de faire le lien entre le produit et la survenue de cette réaction. Le lien peut en l'occurrence être fait en partant de la structure même du biomédicament. Tout d'abord, une réaction immunitaire se produit lorsque le système immunitaire reconnait la structure d'une molécule comme étant étrangère à l'organisme. Lors de l'administration d'un biomédicament, la molécule thérapeutique et les impuretés peuvent être reconnues par l'organisme. Dans le cas de la protéine thérapeutique, l'enchainement particulier d'acides aminés peut être reconnu comme étant un antigène et ainsi déclencher une réaction. Cette reconnaissance se traduit par la libération de cytokines[6] qui engendre une cascade de réactions potentiellement délétères pour l'organisme. Il y a donc une nécessité d'étudier cette capacité de la protéine thérapeutique à déclencher une réaction.

Une réduction du risque d'immunogénicité des protéines a été mise au point. En effet, une protéine thérapeutique est d'autant plus immunogène que son origine est étrangère à l'homme. Cela signifie qu'une protéine d'origine animale serait plus à risque d'engendrer une réaction immunitaire qu'une protéine d'origine humaine. C'est la raison pour laquelle il y a eu plusieurs générations de protéines thérapeutiques telles que les anticorps monoclonaux. Ces derniers étaient initialement issus de souris, et présentaient ainsi des structures murines [6]. Les anticorps ont ainsi fait l'objet de développement itératifs afin de remplacer progressivement les structures d'origine animale par une structure humaine [6]. Ces améliorations ont permis de rendre la protéine thérapeutique moins immunogène. En effet,

la littérature qui indique que la structure d'une molécule thérapeutique n'est pas à l'origine de réaction immunitaire majeure[6].

Cependant les structures étrangères à la molécule thérapeutique, produites malgré le fabricant durant le procédé, peuvent-elles aussi être immunogènes. Ces substances, aussi qualifiées de substances extrinsèques, peuvent de la même façon être reconnue par l'organisme et devenir ainsi potentiellement dangereuses. Il s'agit par exemple des impuretés, qui ne seraient pas éliminées, telles que les endotoxines bactériennes, l'ADN microbien, les protéines dénaturées ou encore les agrégations protéiques. La littérature met en évidence que ces éléments extrinsèques peuvent être à l'origine de réaction immunitaire [6]. L'immunogénicité de ces substances fait donc l'objet d'évaluation pendant les phases de développement du biomédicament.

3. La difficulté d'évaluation de la réaction immunitaire

Or la réaction immunitaire est difficile à évaluer, notamment à cause d'un délai de survenue variable. Il est possible qu'une seule injection suffise pour faire apparaitre les anticorps conduisant à une réaction de type vaccinale [6]. Cette dernière est observée notamment avec l'utilisation en thérapeutique de protéine d'origine animale, qui sera alors qualifiée d'immunogène lorsqu'elle provoque une réaction. Mais les progrès scientifiques ont permis de produire des molécules dont la structures imite de plus en plus les molécules d'origine humaine. Cette amélioration a ainsi conduit à un retardement de la potentielle survenue des réactions. Il peut donc y avoir plusieurs mois de traitement chronique avant que le patient ne produise des anticorps dirigés contre le médicament biologique [6]. Cet allongement des délais de survenue rend ces réactions immunitaires indétectables durant les phases de développement [1] (figure 3).

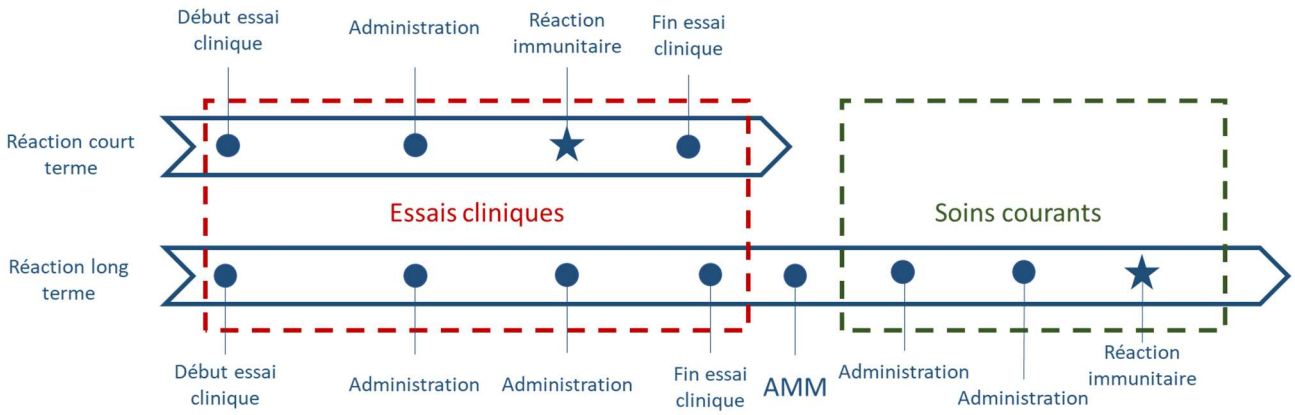

Figure 3: Cadre temporel d'une réaction

Les phases de développement peuvent donc mettre évidence une absence d'immunogénicité d'une protéine thérapeutique, or il n'est pas possible d'affirmer que l'administration répété d'anticorps humain ne génère pas de réaction immunitaire[6]. En effet, il existe de nombreuses composantes pouvant intervenir dans une telle réaction. La variabilité de structure d'un biosimilaire, bien qu'à priori non immunogène, ne devrait donc pas être considérée seule. L'ajout d'autres facteurs à cette variabilité pourraient amener à une augmentation du risque de survenue de réactions immunitaires.

4. L'incertitude quant à la réaction immunitaire

Une de ces composantes est en l'occurrence l'administration du biomédicament sur une longue durée. Il faut dans un premier temps rappeler qu'une prise en charge au long cours peut jouer un rôle dans la survenue d'une réaction immunitaire [6]. Le patient étant exposé à répétition à un produit, il est possible que les premières expositions soient sans effet mais qu'à termes l'organisme finisse par réagir. Tout dépendra de la nature de la prise en charge selon la fréquence d'administration et le dosage [6]. De plus, un patient traité au long cours par un médicament biologique pourrait également être traité par d'autres médicaments. Ces derniers sont destinés à traiter des comorbidités du patient[6]. Or ces traitements supplémentaires pourraient également avoir un impact sur la survenue de réaction *via* un changement du profil de sécurité[5].

A cette composante d'exposition long cours se combine l'administration d'un produit dont les lots varient. Il était déjà question une première fois de l'impact de la variabilité du biomédicament sur la sécurité du produit. Cette question se pose lors d'une première injection. Mais la question des administrations répétées peut également être posée, laissant

ainsi une incertitude au long cours. En effet, ce changement répété de lots différents pourrait présenter des risques d'immunogénicité[1]. De surcroit, il pourrait être discuté d'une majoration du risque selon les pratiques médicales. Il serait ainsi pertinent d'intégrer dans la réflexion l'impact de pratiques médicales visant à changer un biomédicament de référence par un biosimilaire, exposant ainsi le patient à un nouveau niveau de variation.

La problématique est donc la suivante : comment identifier les composantes qui sont à l'origine de la réaction immunitaire ? Il est vrai qu'une difficulté réside dans la multitude de facteurs. Nous avons présenté un médicament biologique, pour lequel il existe des variations à chaque lot. Ces variations ne sont pas toujours immunogènes mais une longue durée de traitement associée à d'autres médicaments pourrait accroitre le risque. D'autant plus que ce dernier serait majoré selon les pratiques médicales qui introduiraient un niveau supplémentaire de changement dans l'administration de biomédicament. Toutefois ce risque de réaction immunitaire n'est pas inconnu puisqu'il est dans un premier temps évalué durant les phases réglementaires de développement.

b) Une évaluation du risque de réaction immunitaire durant les phases de développement mais des incertitudes qui persistent

i. L'évaluation du biosimilaire : la preuve de biosimilarité

1. L'évaluation du biosimilaire : l'obtention de l'AMM

Tout médicament doit obtenir une autorisation de mise sur le marché (AMM). Il est pour cela nécessaire de démontrer la sécurité, la qualité et l'efficacité de ce médicament [1]. Cette démonstration est basée sur le recueil d'informations définies par des exigences réglementaires (figure 4). Dans l'Union Européenne, l'EMA (Agence Européenne du médicament) est chargée de la revue des informations. Ainsi, toute AMM d'un médicament biologique, y compris les biosimilaires, est délivrée par l'EMA [1].

Dans les faits, le biosimilaire sera évalué en comparaison du médicament biologique de référence. L'objectif est d'explorer la puissance, la sécurité et le profil immunologique des biosimilaires[1]. Sont également évalués les procédés de fabrication dans leur capacité à respecter les spécifications initiales[10]. Il s'agit notamment de l'occasion de démontrer que les dérives sont maitrisées et que la micro-hétérogénéité est maintenue dans un intervalle acceptable. S'en suivent des études précliniques, mais également cliniques. Ces dernières visent à évaluer les différences cliniques significatives entre biosimilaire et la référence. A noter que le biosimilaire bénéficie d'un consensus sur les concepts et les méthodes scientifiques d'évaluations règlementaires.

Figure 4: Evaluation réglementaire des biosimilaires

2. La détermination de la biosimilarité

Ces méthodes scientifiques sont destinées à évaluer la biosimilarité du médicament biologique copiant la référence. Ce concept de biosimilarité correspond à la démonstration que le biosimilaire présente des propriétés similaires à la référence. Les propriétés sont en l'occurrence celle de la structure, du profil de sécurité et d'efficacité ou encore de l'activité biologique[4,11]. La démonstration de la biosimilarité se traduit par des études analytiques de

pharmacocinétique et pharmacodynamie [4], mais également d'études précliniques et cliniques.

Le concept de biosimilarité comme son nom l'indique ne peut désigner qu'une similarité ou grande similitude du biosimilaire vis-à-vis de sa référence. En effet, un laboratoire copie un biomédicament de référence en mettant au point les procédés pour obtenir le biosimilaire. Cette mise au point dépend de connaissances d'ingénierie inversée. Malgré ce travail, le biosimilaire aura un profil de sécurité et d'efficacité très proche de la référence sans jamais y être identique [1]. C'est la raison pour laquelle, contrairement au générique, le biosimilaire doit faire l'objet d'exercices de comparabilité. Ces derniers permettent ainsi de couvrir les particularités des biosimilaires et leur variabilité naturelle.

3. Les exercices de comparabilité en amont des essais cliniques

La première étape de ces exercices de comparabilité est l'étude de la structure de la molécule. Il a été mentionné la variabilité structurale de la molécule liée à l'utilisation de procédés biologiques. Cette variabilité entraine des différences de structures entre biosimilaires et références. Il y a donc une étude de cette structure par des analyses en profondeur de la structure du biosimilaire. Des analyses de séquençage de la protéine in vitro [1] s'intéressent à la structure de la protéine et aux modifications post traductionnelles. Il existe également la spectroscopie de résonnance magnétique ou encore la chromatographie afin d'explorer la structure de la molécule biologique [1]. Mais ces analyses structurales auront des limites et devront donc être poursuivies par des tests précliniques.

Ces études précliniques permettent d'explorer l'effet de la molécule sur un organisme vivant. Ces études vont à la fois porter sur des cellules vivantes dites modèles cellulaires et sur des animaux. L'objectif est d'explorer des possibles variations qui n'auraient pas été mises en évidence avec les analyses structurales. Il y aura donc une comparaison de la pharmacocinétique et de la pharmacodynamie du biosimilaire et de la référence[10]. La pharmacocinétique concerne l'effet de l'organisme sur la molécule alors que la pharmacodynamie s'intéresse à l'effet de la molécule sur l'organisme. Il peut par exemple s'agir d'un effet qui doit être obtenu par administration de la molécule, tel que la cytotoxicité cellulaire dépendante de l'anticorps (ADCC). Les études précliniques apporteront également

des informations quant à la toxicologie du biosimilaire à travers des études inter espèces de cross réactivité [10].

ii. Les essais cliniques des médicaments biologiques et leurs limites

1. La soumission du biosimilaire aux essais cliniques

Les études précliniques d'un médicament doivent être complétées par des essais cliniques. Ces derniers visent à étudier la causalité entre l'utilisation d'un médicament et un effet dans des conditions idéales[12]. En effet, des patients seront inclus dans l'étude selon qu'ils répondent à des critères d'inclusions afin de placer l'évaluation dans des conditions idéales. Dans le cas des génériques, dont le principe actif est identique à celui du *princeps*, seules les études de bioéquivalence sont requises[2]. Il n'est alors pas nécessaire de conduire des essais cliniques. Il faut cependant des études cliniques complètes pour les biosimilaires [2]. Ces études cliniques des biosimilaires ont pour objectif de démontrer l'absence de différences cliniques significatives entre le biosimilaire et la référence[10].

2. La limite des essais cliniques dans le recueil de l'information

Les essais cliniques sont menés dans des conditions idéales afin d'obtenir une AMM. Or les patients qui recevront le traitement, une fois l'AMM obtenue, ne seront pas sélectionnés avec des critères de sélection aussi rigoureux. Il y aura bien entendu une utilisation dans le respect de l'AMM mais les patients seront plus hétérogènes que ceux inclus durant les études cliniques. Il pourrait ainsi y avoir davantage de comorbidités au sein de la population réellement traitée que chez les patients inclus[13]. De plus, les essais cliniques toucheront vraisemblablement une population avec une faible variabilité génétique[12]. Or, cela pourrait traduire un manque de représentativité des variations d'un gène qui influencerait la réponse à un biomédicament. En somme, les essais cliniques ne rendraient pas entièrement compte de l'effet du médicament biologique sur une population réellement traitée [12]. Il y aurait ainsi un problème de validité externe, car les études pourraient présenter des informations difficilement transposables à une utilisation en soins courants.

Plus spécifiquement, les essais cliniques de biosimilaires présentent une limite vis-à-vis de l'évaluation du risque de réactions immunitaires. Il faut rappeler que la réaction immunitaire peut avoir un délai de survenue important. Or les essais cliniques ayant une durée limitée, ils pourraient ainsi ne pas mettre en évidence l'immunogénicité au long cours du médicament

biologique. Dans les faits, les essais cliniques pourraient ne pas détecter les événements rares ou retardés[13]. Or cela pourrait accroitre un besoin de suivi des patients sur le long terme lors d'une utilisation en soins courants. Cette notion de suivi en soins courants nous amènera à discuter de l'exploitation de données générées par une prise en charge après l'obtention de l'AMM.

Tout d'abord, il faut mentionner que les essais cliniques sont longs et coûteux[7]. Il y aura donc des limites techniques et financières à utiliser les essais cliniques pour générer de l'information sur le produit. Les données générées par l'utilisation d'un produit lors de son exploitation commerciale pourraient compléter ce manque d'information des essais cliniques[12]. Il s'agit en l'occurrence des données de vie réelle qui permettraient d'identifier les éléments prédictifs parmi des réponses hétérogènes à un traitement[14]. En d'autres termes, il s'agirait d'explorer les données de soins courants afin d'acquérir de l'information quant à l'utilisation des biosimilaires. Cela permettrait ainsi d'identifier les potentielles causes de réactions immunitaires parmi de nombreuses possibilités. *In fine*, il pourrait y avoir vocation à déterminer le meilleur traitement pour un patient donné à un moment précis de la prise en charge[14].

iii. Une particularité des biosimilaires : les extrapolations d'indications

1. Des phases réglementaires allégées

Les essais cliniques présentent donc des limites d'information vis-à-vis des médicaments biologiques. Dans un premier temps parce que de telles études manqueraient de représentativité, et dans un second temps parce qu'elles ne couvriraient pas une période suffisamment longue pour évaluer pleinement le risque. Mais à ces limites s'ajoutent les spécificités réglementaires des biosimilaires. En effet, ces derniers présentent des phases d'évaluation allégées[2], visant à accélérer le développement et à en réduire les coûts. Une phase allégée est notamment celle de l'extrapolation d'indication.

2. La particularité réglementaire : les extrapolations d'indications

L'extrapolation d'indication permet l'utilisation d'un biosimilaire dans une indication pour laquelle il n'a pas été étudié[10]. Concrètement, des études ont démontré que le médicament biologique de référence pouvait être utilisé en toute sécurité dans une indication. Mais avec l'extrapolation d'indication, le biosimilaire peut bénéficier d'une approbation pour cette

même indication sans avoir réalisé l'exercice de démonstration. Il est toutefois nécessaire que le biosimilaire réponde à certaines conditions. En effet, la référence doit avoir été utilisée depuis longtemps afin que le biosimilaire puisse prétendre à l'extrapolation d'indication. Ensuite il faut démontrer que le biosimilaire présente le même mécanisme d'action que la référence [10]. Pour terminer, l'extrapolation doit être justifiée au regard de l'immunogénicité et de la toxicité attendue[5].

3. L'extrapolation d'indication : des incertitudes quant à la réaction immunitaire

Bien que les autorités soutiennent l'extrapolation d'indication[2], des questionnements peuvent exister. Ce questionnement peut être relatif à la connaissance du risque d'un médicament utilisé dans une indication. Une indication permet notamment d'encadrer l'utilisation du biomédicament de référence pour lequel les risques sont connus. Cependant en cas d'extrapolation d'indication, les risques liés à cette indication, lors de l'utilisation d'un biosimilaire, pourraient être différents de ceux de la référence. Or nous avons vu que la réaction immunitaire liée aux biomédicaments intègre une multitude de composantes difficiles à évaluer. Cette extrapolation d'indication pourrait donc jouer un rôle en parallèle de ces composantes et changer ainsi le profil de sécurité du biosimilaire. Ce questionnement relatif aux extrapolations d'indications peut donc traduire un manque d'information sur la sécurité d'utilisation des biosimilaires. Il s'agit ainsi d'un niveau supplémentaire d'incertitude.

L'incertitude peut donc être formulée de la façon suivante. Un biosimilaire est utilisé dans une indication qui n'a pas été étudiée. Les praticiens s'interrogent sur la sécurité d'utilisation en vie réelle du biosimilaire pour cette indication. Puisque la réaction immunitaire est complexe, est-il possible que le biosimilaire utilisé dans cette indication présente un profil de sécurité différent de celui de la référence ? De la même façon, il est possible de formuler une volonté d'apporter l'information manquante aux pratiques liées à l'extrapolation d'indication. L'objectif serait de collecter des données sur les biosimilaires pour lesquels l'indication n'est pas documentée durant les essais cliniques[4].

c) Utilisation des médicaments biologiques similaires en soins courants : un questionnement lié aux pratiques

i. Les pratiques classiques d'utilisation d'un médicament biologique : traitement des pathologies chroniques

1. L'utilisation en soins courants des biomédicaments

Prenons par exemple l'indication de médicaments en oncologie. Des traitements standards tels que des molécules chimiques sont utilisés, *via* la chimiothérapie, pour traiter certains cancers. Mais les patients peuvent avoir un manque de réponse à ces thérapies[1]. Il peut également ne pas y avoir de traitement chimique disponible pour un type de cancer précis. C'est une des raisons pour lesquelles les biomédicaments peuvent être prescrits.

Il existe de nombreuses classes thérapeutiques de molécules biologiques. Il faut notamment citer les insulines, héparines, interférons, mais également les anticorps monoclonaux[2]. Ces derniers peuvent également être utilisés en oncologie pour le traitement de certains cancers. En effet, 40% des biomédicaments sont utilisés en oncologies car ils répondent souvent à des pathologies sans traitements[2]. Il faut également remarquer que les médicaments biologiques peuvent être utilisés dans le traitement de pathologie chronique[1], ce qui signifie une utilisation au long cours.

2. La prescription d'un médicament biologique chez un nouveau patient

Tout d'abord, il faut voir qu'une prescription se fait selon les caractéristiques du médicament. Le prescripteur considère les informations de la spécialité qu'il souhaite prescrire, qu'il s'agisse d'un médicament chimique ou d'un biomédicament. Il y aura dans les deux cas une prescription respectant l'AMM pour un patient ne présentant pas de contre-indication pour ce traitement. Puisqu'il s'agit d'un médicament biologique, le prescripteur pourra être attentif aux études de pharmacovigilance, aux politiques locales et aux potentielles économies pour le système de Sécurité sociale[15]. Il s'agit d'un premier niveau de réflexion conditionnant le choix de prescription.

Un second niveau de réflexion intègre la dimension de médicament biologique de référence et biosimilaire. En effet, dans le cas des médicaments biologiques, il est nécessaire d'intégrer un élément supplémentaire. Lors de la première prescription d'un médicament biologique, le

patient n'a jamais reçu de biomédicament. Le patient est ainsi qualifié de naïf. Le prescripteur a donc une large gamme de médicament biologique à sa disposition, incluant le biomédicament de référence et ses biosimilaires. Le prescripteur pourra ainsi choisir parmi cette gamme pour réaliser sa prescription. Toutefois, toutes ces molécules présentent la bonne indication mais il n'existe aucune information qui permet de prédire la réponse du patient à un traitement. Il ne sera ainsi pas possible d'identifier un risque plus important de réaction immunitaire avec l'un des biomédicaments. Le praticien sera ainsi amené à prescrire un biomédicament en supposant que les biosimilaires et leur référence présentent le même niveau de risque. Le patient recevra ainsi ce même biomédicament pour une longue durée dans le cadre d'une pathologie chronique.

ii. Les changements de traitements possibles initiés par le médecin : interchangeabilité et risque de rupture de tolérance

1. La possibilité de prescrire un autre médicament biologique en cours de prise en charge

La première prescription d'un biomédicament est considérée comme étant une initiation de traitement. Le traitement peut donc être initié avec le médicament biologique de référence. Dans le cas d'une pathologie chronique, l'ordonnance sera renouvelée et le patient recevra à chaque délivrance le médicament biologique de référence. Mais il est possible que le praticien prescrive un autre biomédicament que celui initialement prescrit. Il pourra en l'occurrence prescrire un biosimilaire à la place du biomédicament de référence. Il est alors question d'interchangeabilité ou de *switch* (figure 5). Le praticien va donc prescrire un biomédicament dont il ne connait pas la réponse du patient au traitement, à la place d'un biomédicament de référence dont la réponse est connue[8]. Il faut remarquer que le *switch* peut se traduire par le changement de la référence par son biosimilaire et inversement.

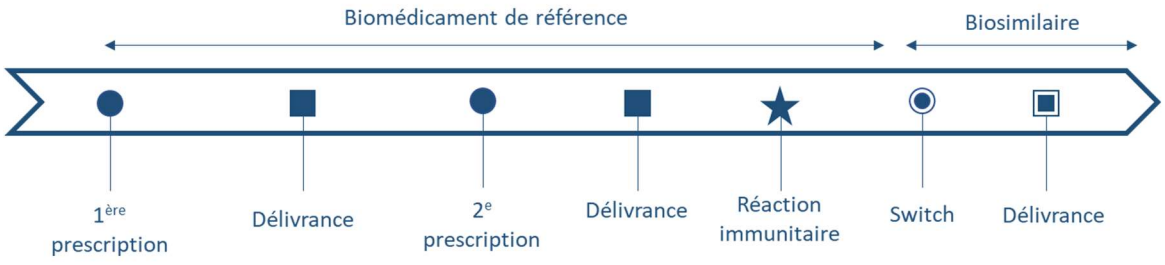

Figure 5: Switch

Ce *switch* n'est pas effectué sans prendre de précautions. En effet, il est de la responsabilité du praticien de réaliser un *switch* au regard des données de la science [8]. Le prescripteur sera ainsi amené à déterminer au cas par cas l'éligibilité du patient à l'interchangeabilité[16]. En d'autres termes, le médecin devra apprécier une situation dont les différentes composantes pourraient justifier un changement de biomédicament. Dès lors, le *switch* fera l'objet d'une décision médicale partagée entre le patient et le prescripteur[8,16]. Le praticien devra ainsi jouer un rôle d'information pour garantir l'observance du patient afin qu'il suive la prescription. En effet, cette démarche de décision partagée vise à donner confiance aux patients notamment *via* les informations de sécurité clinique du biosimilaire [16]. Le praticien pourrait pour cela valoriser les données complémentaires relatives à la sécurité d'utilisation des biosimilaires en soins courants.

Il faut noter que seul le médecin a le droit de proposer un *switch*[9]. Ce dernier peut trouver son origine dans plusieurs cas de figure. Le premier cas de figure est celui de la réaction allergique [9]. Il y a par exemple, la survenue d'une réaction immunitaire qui survient pour un patient traité par un biomédicament biologique de référence. Le prescripteur pourrait souhaiter continuer le traitement du patient mais avec un médicament biologique similaire afin de réduire le risque de réaction immunitaire. Un autre cas de figure est celui de la rupture d'approvisionnement [9]. Le médicament biologique de référence n'étant plus disponible, le praticien pourrait prescrire le biosimilaire pour assurer la continuité du traitement. Le dernier cas de figure serait celui de l'influence des politiques de santé [8]. Ces dernières peuvent inciter les praticiens à recourir aux biosimilaires pour des raisons économiques à la place du biomédicament de référence.

2. Le risque de l'interchangeabilité : la rupture de tolérance

Il peut y avoir un questionnement quant aux conséquences de ce changement de médicament biologique. Le patient reçoit un premier biomédicament qui lui sera administré pendant une certaine période. Cette période permet d'avoir du recul sur la réponse du patient à ce traitement. Or cette réponse peut se manifester de plusieurs façons. Le patient peut tolérer le biomédicament mais également ne pas le tolérer. Dans le cas où le patient supporte bien le médicament biologique, il se posera la question d'un risque de rupture de cette tolérance en cas de changement de traitement. Ce changement aurait donc lieu pour des raisons autres que des réactions allergiques. En effet, le changement de biomédicament en cours de prise

en charge présente un risque de rupture de l'équilibrage de la tolérance mais également un risque immunogène[9]. Ce risque vaut que le changement se fasse dans le sens d'un biosimilaire ou de sa référence [9].

Il faut remarquer que les praticiens, ayant à l'esprit ce risque, ont davantage de considérations pour la sécurité des biosimilaires lors d'un changement de traitement plutôt qu'à l'initiation [9]. Ainsi la question de sécurité des biosimilaires est soulevée dans un contexte de *switch*. En effet, la réponse du patient au biomédicament de référence est connue alors que celle au biosimilaire en cas de *switch* serait inconnue. Les praticiens s'interrogent donc sur le risque de rupture d'équilibrage. Il y a ainsi un besoin pour les prescripteurs d'acquérir de l'information sur cette pratique de *switch* d'une référence vers un biosimilaire.

3. Les études de switch et leurs limites

Ce besoin d'information se traduit par des études de *switch* conduites afin de générer de l'information sur la pratique d'interchangeabilité en soins courants. Ces études de données de vie réelle visent à explorer les conséquences d'un remplacement d'un biomédicament de référence par son biosimilaire. Il y a pour cela l'inclusion de patient pour une durée limitée dont les données de prises en charges seront analysées. Cette analyse est d'autant plus importante pour les praticiens lorsqu'il s'agit de biosimilaires récents pour lesquels les prescripteurs auraient davantage de réserve[15]. Toutefois, dans l'analyse de 90 études de *switch*, Chang LC. ne met pas évidence de différences significatives de sécurité ou d'efficacité causées par le *switch*[5].

De plus, la *British Society of Rheumatology* soutient la décision de *switch* vers un biosimilaire. Cependant, elle indique que ce changement doit se faire au cas par cas jusqu'à ce que des données soient disponibles pour soutenir un *switch* sûr. Cette recommandation met en évidence un besoin continu d'informations complémentaires. De nouvelles données de vie réelle pourraient ainsi permettre d'explorer les conséquences de ce changement de biomédicament de référence par un biosimilaire[4]. En ce sens, il serait pertinent de poursuivre les études de données de vie réelle pour compléter les informations relatives au *switch* par biosimilaire. Ces études permettraient ainsi d'intégrer toutes les évolutions de pratiques liés à l'augmentation de l'utilisation de biosimilaires. Il s'agit également d'une opportunité d'intégrer davantage de patients dans l'analyse sur une période plus longue.

iii. La question de substitution pharmaceutique des biosimilaires

1. Les incitations à l'utilisation des biosimilaires

Une augmentation de l'utilisation des biosimilaires pourrait être observée pour deux raisons. La première est celle d'une augmentation de biosimilaires sur le marché. En effet, 50 biosimilaires ont été autorisés dans l'UE avant le 9 novembre 2018[5], et ce nombre pourrait augmenter avec l'expiration des brevets des biomédicaments de référence. La deuxième raison est celle de l'attrait économique des biosimilaires pour le système de Sécurité sociale. En effet, le biosimilaire permet d'apporter une copie de médicament biologique hautement similaire à la référence, sans différence significative et à moindre coût. Dans les faits, il existe une réduction du prix du biosimilaire de 20 à 30% par rapport à la référence[2]. Les biosimilaires sont ainsi une alternative plus abordable aux médicaments biologiques de référence[1], ce qui permet d'augmenter l'accès au traitement [4].

Bien que le biosimilaire soit une solution pour maîtriser les dépenses de santé [1], ils ont pour certains du mal à pénétrer le marché français. En effet, face aux incertitudes des praticiens concernant leur utilisation, des politiques de santé en faveur des biosimilaires sont nécessaires. C'est la raison pour laquelle l'article 51 de la loi de financement de la Sécurité sociale (LFSS) de 2017, a mis en place une expérimentation pour tester un intéressement direct[17]. Sont ainsi visés les groupes de biomédicaments dont le taux de pénétration des biosimilaires est inférieur à 10% [17]. Ces expérimentations se traduisent par la mise en place de mécanismes d'intéressement pour inciter les établissements de santé à la prescription hospitalière de biosimilaires.

2. La question du droit de substitution pharmaceutique

Il n'existe pas de consensus sur la substitution de biomédicament qui reste dès lors du ressort de chaque Etat membre de l'Union européenne[4]. Un droit de substitution a été envisagé par le législateur français avec le LFSS 2014 afin d'augmenter leur utilisation[19]. Ce droit a été prévu en tant qu'acte pharmaceutique par lequel un pharmacien délivre un médicament biologique autre que celui prescrit par le médecin[9]. Cette substitution ne porte donc que sur une seule molécule ou son équivalence thérapeutique [9].

Ce droit de substitution a été prévu par la LFSS 2014 en tant que droit de substitution uniquement pour les nouveaux patients[2]. Ainsi, seuls les patients n'ayant jamais reçu de

biomédicaments peuvent se voir délivrer un biosimilaire à la place de la référence prescrite. Cette substitution telle que décrite dans la LFSS 2014 est conduite dans le respect d'une liste de biosimilaire qui mentionne les références et les biosimilaires pouvant s'y substituer [2]. Une fois le biosimilaire délivré au titre de la substitution, la LFSS 2014 prévoit que le traitement ne serait renouvelé et poursuivi qu'avec ce biomédicament uniquement [9]. Mais ce droit de substitution a été abrogé par l'article 42 du projet de loi du PLFSS 2020 « notamment pour des raisons de traçabilité et de sécurité sanitaire »[19]. Il pourrait donc y avoir une réflexion quant aux éléments de sécurité sanitaire à l'origine d'une telle décision.

3. Un questionnement de la substitution pour des raisons sanitaires

Deux niveaux de questionnements relatifs à la substitution peuvent être formulés. Un questionnement de la substitution peut exister quant aux impacts potentiels de la substitution pour le patient. Dès lors, ce questionnement pourrait exister en cas de prescription d'un médicament biologique à un patient. En effet, ce biomédicament est associé à un nom à des spécificités d'administration et d'étiquetage. Ainsi, la délivrance d'un biomédicament avec des caractéristiques différentes de celui prescrit pourrait conduire à un risque de non observance du patient. Cela signifie que le patient pourrait ne pas respecter les posologies en ne prenant pas son traitement comme il le devrait. De surcroit, un effet nocebo pourrait être observé. Ces effets négatifs liés à la délivrance d'un biosimilaire pourraient survenir à cause de l'impact psychologique lié à la substitution. Cette réaction néfaste jouerait également un rôle dans la discontinuité du traitement[4].

Cette substitution pharmaceutique pourrait de la même façon impacter les praticiens. Ce droit de substitution en l'état actuel requière davantage d'informations pour que les praticiens puissent prendre des décisions informées[1]. Il existe déjà un manque d'information sur le *switch* que de nouvelles études de données de vie réelle pourraient combler. Ces études permettent ainsi d'apporter des informations aux prescripteur dont les connaissances sur les biosimilaires sont primordiales[18]. Mais la substitution pourrait perturber les études de *switch* [1]. En effet, cette pratique se traduirait par une délivrance différente de la prescription or les études de *switch* visaient justement à évaluer les risques de ce changement de prescription. La substitution pourrait ainsi complexifier l'évaluation des risques et des bénéfices de l'utilisation. D'autant plus que cette substitution s'accompagnerait d'une problématique de

traçabilité contribuant à associer correctement un évènement indésirable au biomédicament en cause[2].

L'article 42 du projet de loi du financement de la Sécurité sociale 2020 mentionne une abrogation de ce droit de substitution pour des raisons de sécurité sanitaire[19]. Il apparait dès lors que des informations supplémentaires seraient nécessaires pour évaluer cette question de substitution pharmaceutique. Or cette dimension de sécurité sanitaire comprend notamment des questions de sécurité et donc d'évaluation du risque de réaction immunitaire. Une approche exploitant les données de vie réelle pourrait intégrer plusieurs questionnements quant à la sécurité d'utilisation des biosimilaires (figure 6). Il serait notamment pertinent d'inclure les particularités structurales, les risques de réaction immunitaires au long cours, les extrapolations d'indications ou encore les pratiques de *switch*. Ainsi, cette question du droit de substitution pourrait être examinée au regard d'informations supplémentaires collectées suite à l'utilisation de biosimilaires en soins courants.

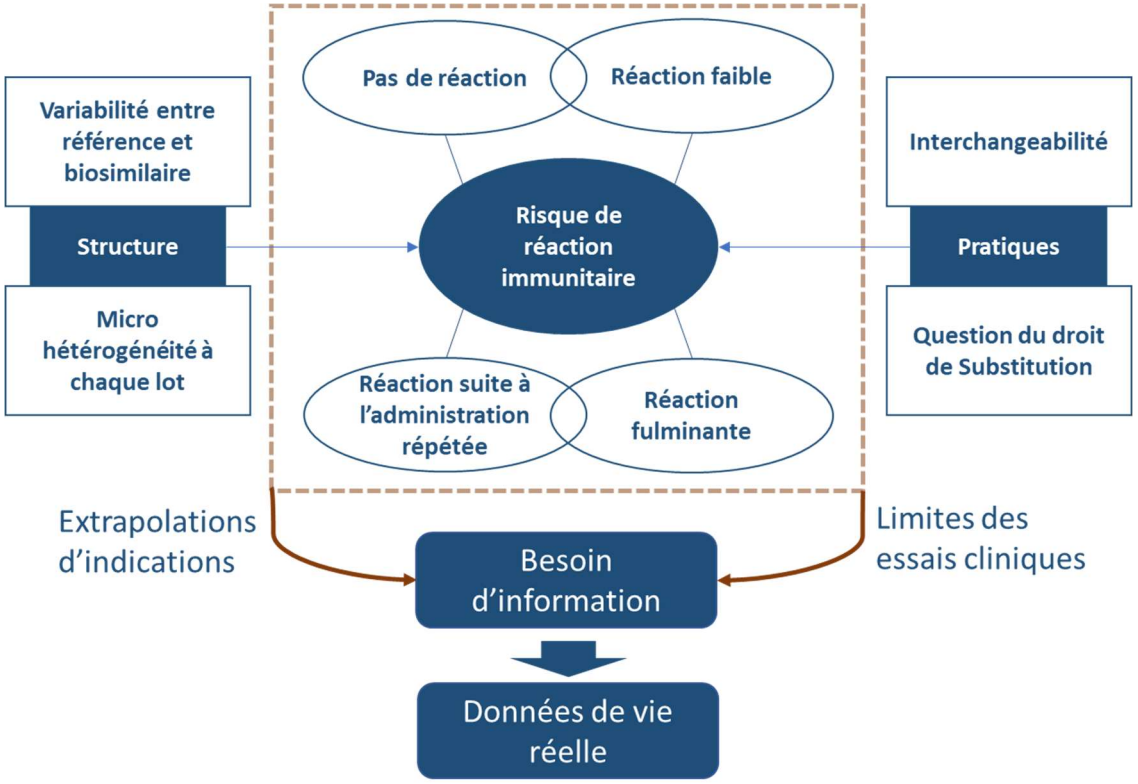

Figure 6: Les questionnements relatifs aux biosimilaires

2. L'utilisation des biosimilaires génère des données de vie réelle qui pourraient être exploitées

a) L'intérêt des données de vie réelle

ii. L'intérêt des données de vie réelle au regard de l'utilisation des biosimilaires

1. La définition des données de vie réelle

Les activités médicales, quelles qu'elles soient, génèrent des données de santé. Toute prise en charge d'un patient par le système de santé peut donc être associée à une génération de données. Bien entendu, toutes ces données ne sont pas collectées. Certaines données seront spécifiquement collectées et d'autres non. Il est possible que des données soient collectées à intervalle de temps régulier, que ce soit par les médecins, les infirmières et même les patients directement[13]. Il peut s'agir de diverses données telles que l'âge, le sexe, les caractéristiques d'une pathologie, ou encore l'existence de comorbidités et de traitements concomitants [13]. Ainsi, l'utilisation d'un biosimilaire peut être associée à une génération de données. Ces données pourront être étudiées afin d'obtenir de l'information quant à l'utilisation des biosimilaires.

Des données sont collectées sur un produit dès les phases de développement. En effet, l'évaluation précoce d'un biosimilaire commence dès les études de biosimilarité avec les études précliniques. S'en suivent alors les études cliniques qui visent à analyser des données collectées lors d'une utilisation très contrôlée du biosimilaire. Mais ces études cliniques présentent des limites comme mentionnées précédemment, et des études complémentaires seraient ainsi pertinentes. Il est donc question ici de données collectées après les phases de développement, en dehors du cadre des essais cliniques[12]. Les données seraient alors collectées une fois l'AMM du biosimilaire obtenu, lors d'une utilisation en soins courants. Ces données sont désignées par les termes de données de vie réelle. Il sera ultérieurement évoqué le concept des études de données de vie réelle.

Ces données de vie réelle pourraient trouver leur place dans le questionnement lié à l'utilisation des biosimilaires. Tout d'abord ces données de vie réelle sont utilisées dans un cadre général pour différents objectifs. Elles sont notamment exploitées pour explorer l'histoire naturelle d'une pathologie, déterminer les coûts des interventions mais également

décrire les choix de traitements[12]. Dans le cas de ce document, l'objet de leur analyse porterait sur la question de l'immunogénicité des biosimilaires lors d'une utilisation en soins courants. Pour répondre à cette question, il sera donc nécessaire d'avoir des données pertinentes issues de l'utilisation des biosimilaires. Le défi serait alors d'avoir des données représentatives d'une utilisation en soins courants.

2. Une génération de données associée à la notion de données massives

Cette représentativité implique d'intégrer des événements multiples intervenant pendant la prise en charge. Ces divers évènements peuvent être source de données qui seront collectées. Ces évènements sont notamment les visites des praticiens, l'hospitalisation, le diagnostic, les interventions, les factures mais aussi les médicaments prescrits[20,21]. Ces données peuvent ainsi être d'une nature différente selon les événements. De surcroît, de nombreux paramètres seront pris en compte, pouvant ainsi aller au-delà de ceux pris en compte durant les essais cliniques (figure 7).

Il est question de données massives lorsque 3 critères sont réunis : le volume, la variété et la vélocité. Il y a en effet un grand volume de données [22] collectées à la suite de la prise en charge d'un patient car de nombreux événements seront associés à de la génération de données. Mais ces nombreux événements sont également associés à des formats de données qui leur sont spécifiques, d'où le deuxième critère. Dans les faits, les données pourront être structurées, avec un format particulier, ou bien non structurées. Dans ce dernier cas de figure,

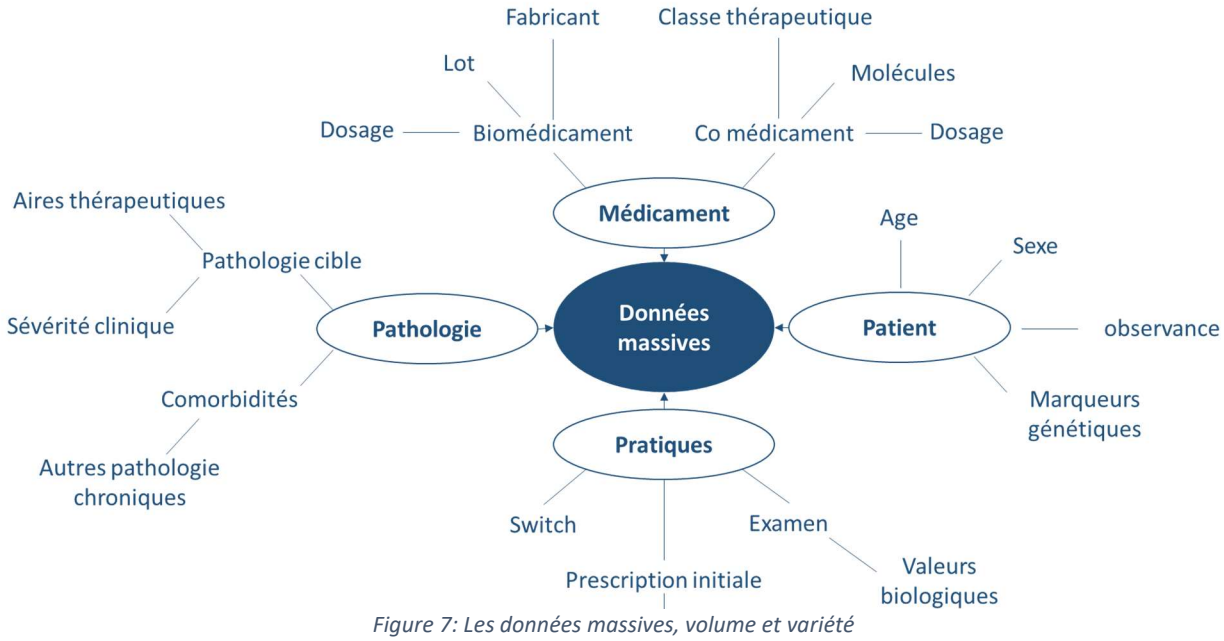

Figure 7: Les données massives, volume et variété

une donnée peut être retrouvée sous forme de texte libre dans un dossier médical. Le dernier critère de vélocité est associé à la notion de flux. Ce critère désigne la vitesse à laquelle les données sont obtenues et collectées [22]. La vélocité désigne en quelques sortes la vitesse de mise à jour de l'information [22]. Ce flux évoqué ici provient des différents évènements qui peuvent actualiser la situation d'un patient. Ainsi, de nombreux événements conduisent à un grand volume de données de formats variés et donc un flux important, amenant ainsi au concept de données massives (figure 7).

3. L'intérêt des données de vie réelle

Les données de vie réelle pourraient donc présenter un double avantage vis-à-vis des essais cliniques. Le premier est inhérent aux données de vie réelle car ces dernières permettraient de suivre plus de patients plus longtemps lors de leur prise en charge. D'autant plus que ces données assureraient un suivi de populations plus hétérogènes [5] que celles suivies durant les essais cliniques. Le second avantage de ces études résident dans l'intégration des composantes des données massives. Ces dernières présentent l'avantage de rassembler plusieurs sources et donc plusieurs flux de données. Permettant ainsi de couvrir plus de paramètres, ces données massives sont une opportunité de répondre à davantage de questions sur les biosimilaires. En couvrant ces questions, il serait dès lors possible de rendre compte d'une situation où l'utilisation de biosimilaires se traduirait par un évènement indésirable.

Il est donc important de faire le lien entre cet événement indésirable et l'utilisation d'un biosimilaire. Il serait par exemple possible de chercher à mettre en évidence un problème de tolérance lié aux traitements concomitants[13]. Il n'apparait pour cela pas suffisant de rechercher le lien entre un biosimilaire donné et une réaction immunitaire. En effet, davantage de composantes seraient nécessaires pour comprendre ce lien. Il serait notamment intéressant d'intégrer toutes les composantes jugées pertinentes dans cette démarche d'exploration. Il est ainsi possible de considérer la composante liée à la structure des biosimilaires, et de connecter cette notion de variabilité structurale à la survenue d'une réaction immunitaire.

De la même façon, il serait judicieux d'associé la composante relative aux pratiques médicales qu'il s'agisse d'une initiation de traitement ou d'un *switch*. Ces données d'interchangeabilité permettraient ainsi l'exploration du lien entre rupture d'équilibrage et une pratique médicale. En ce sens, les composantes liées à la prise en charge des individus pourraient donc être incluses afin d'apprécier l'impact de certains paramètres sur la survenue de l'évènement indésirable. Cependant, l'intégration de toutes ces composantes implique d'obtenir les données qui leur sont associées. Or ces données se trouveront dans des sources différentes, dites bases de données.

iii. Les données sont regroupées dans des bases

1. Les bases de données et leurs spécificités

Les données de vie réelle générées par la prise en charge d'un patient se retrouvent dans des bases de données lorsqu'elles sont collectées. Ces bases de données permettent ainsi d'apprécier l'utilisation d'un médicament en conditions réelle[23]. Ces bases sont intéressantes car leur traitement avec une méthode appropriée permettrait de répondre aux questions d'utilisation d'un médicament en soins courants. Il existe en l'occurrence 3 types de bases de données qui présentent chacune des particularités (figure 8). La première catégorie est celle des registres, dits bases de données primaires[14]. Les bases de données secondaires comprennent les dossiers médicaux électroniques. La troisième catégorie est celle des bases de données médicaux administratives dont les données du PMSI (Programme de médicalisation des systèmes d'information) et des entrepôts des données de santé [14]. La présentation des bases de données permet ainsi de justifier le choix des bases nécessaires pour l'étude.

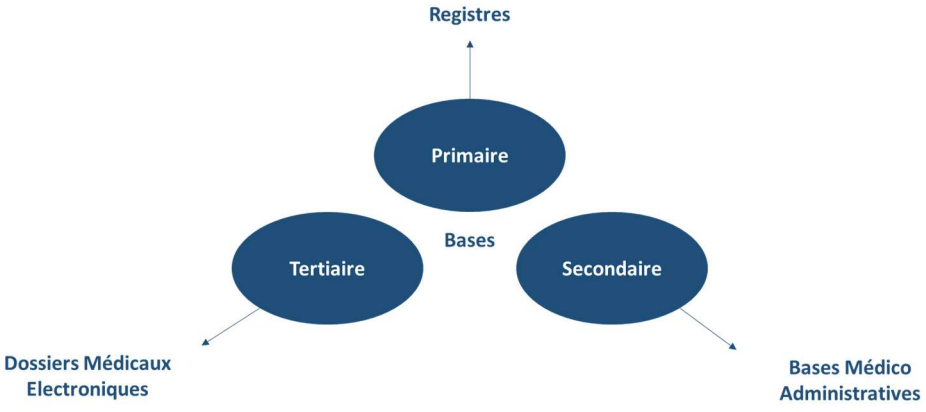

Figure 8: Les bases de données

Les registres sont conçus pour réunir les données de populations sur la pathologie et les traitements, la nature et la fréquence des évènements indésirables[4]. Ils permettent ainsi de compléter les données des essais cliniques [4]. Il existe l'exemple de l'initiative ESME (Epidémio Stratégie Medico Economique) développée par UNICANCER afin de constituer une base de données en cancérologie. Cette base ESME permet de mener des études en vie réelle pour mieux comprendre les déterminants stratégiques thérapeutiques [23]. Les registres rassemblent ainsi différentes données, telles que le pays, l'année d'inclusion, les critères d'inclusion, la population, les traitements, la biothérapie ou encore la durée de suivi[13]. Afin de collecter ces données, les registres peuvent varier selon les méthodes à travers la définition de la population ou le type de recueil. L'avantage de ces bases de données primaires est leur suivi sur plusieurs années d'une large population atteint d'une même pathologie traitée par un même médicament [13].

Les dossiers médicaux sont alimentés par le praticien qui y consigne ses observations et ses prescriptions. Grâce à l'accroissement de l'informatisation, du développement des bases de données et des outils de gestion, ces dossiers médicaux ont été informatisés. Cette informatisation a permis une dématérialisation de ces dossiers, de même qu'une information des prescriptions et des délivrances[23]. Les données contenues dans ces dossiers médicaux sont intéressantes d'un point de vue des détails cliniques[22]. Ces dossiers médicaux seront cependant moins adaptés concernant les comorbidités [22] pour lesquelles il manquera des données. Or il serait délicat de conduire l'analyse sur d'une prise en charge d'une pathologie sans exploiter les données liées aux comorbidités. Ces dernières seraient toutefois disponibles dans les bases médico administratives.

L'informatisation du système de santé a également permis la création des bases médico administratives telles que la SNIIRAM [23]. Ces bases de données comprennent les entrepôts de données hospitalières constituées d'un ensemble de données de formats variés [23]. Ces entrepôts réunissent les données hospitalières et de soin ambulatoire [23]. Ces données sont ainsi utilisées pour la réalisation de recherches non interventionnelles, d'études de faisabilité d'essais cliniques ou encore de pilotage de l'activité hospitalière [23]. Les données médico administratives présenteraient un intérêt pour les médicaments administrés mais seraient moins utiles pour apprécier la sévérité d'une pathologie ou les facteurs cliniques impliqués [22].

2. *L'exploitation conjointe de bases de données*

Les bases de données présentées ici peuvent ainsi contenir des données de vie réelle d'intérêt dont l'exploitation pourrait apporter une réponse aux questions de sécurité des biosimilaires. Il faudrait cependant considérer les problématiques juridiques d'accès à ces bases. Toutefois, les particularités juridiques ne seront pas développées dans ce document. Des particularités techniques d'exploitation des bases de données pourront également conditionner l'analyse. En effet, le format peut varier selon le support choisis par les établissements concernant les comptes rendus et les courriers[23].

Le premier élément à considérer est que les données d'intérêt ne se trouvent pas toutes dans la même base de données. Or chaque base isolée les unes des autres pourraient ne pas suffire pour répondre aux questions de sécurité des biosimilaires. Par exemple, les informations du SNDS (Système national des données de santé) pourraient ne répondre que partiellement à la demande de l'étude [23]. Il serait donc intéressant d'envisager une exploitation conjointe des bases. Il serait dans un premier temps plus facile de mettre en œuvre une combinaison de bases d'une même nature. Il existe par exemple plusieurs registres qui portent sur les biothérapies, dont certains ont été mis en place dans le cadre d'étude de *switch*. Il serait intéressant de combiner ces registres pour mener l'étude de *switch*. Or la combinaison de différents registres pose déjà un problème de format car ce dernier pourra être hétérogène selon les registres[13]. De la même façon, il pourrait être discuté de la combinaison entre les données médico administratives et les données des dossiers médicaux[22].

3. *L'exploitation des données par une méthode appropriée*

Dans le but de répondre aux questions de sécurité au long cours des biosimilaires, il serait nécessaire d'identifier les données pertinentes. Il apparaitrait ainsi que certaines bases de données devraient faire l'objet d'une extraction afin d'obtenir ces données[24]. Cette extraction conduira à l'obtention de données très différentes par nature. Or ces données seules ne sont pas informatives [14] en tant que telles. En effet, il sera nécessaire de traiter ces données afin d'obtenir une information pour répondre aux questions. Une méthode de traitement des données devra donc être appliquée afin d'obtenir cette information.

La démarche présentée ici suppose un traitement de données massives dont le volume, la variété et la vélocité vont rendre difficile le traitement. En effet sur la variété, les données

pourront avoir des formats variés avec une forte hétérogénéité entre les établissements[23]. De la même façon concernant le volume, les dossiers médiaux peuvent contenir un grand nombre de variables[21], telles que l'âge, le sexe ou la réponse à un traitement. Un grand nombre de ces variables ont un caractère prédictif [21]. En d'autres termes, le traitement de ces données peut conduire à l'obtention d'information, de prédiction, quant à la survenue d'un évènement. Mais les modèles traditionnels de traitement des données ne peuvent traiter qu'un nombre limité de variables, et ainsi n'obtenir que des prédictions imprécises. Ces modèles ne permettraient donc pas d'obtenir la réponse aux questions posées sur les biosimilaires.

Bien que les méthodes traditionnelles ne conviennent pas au traitement de données massives, la littérature met en évidence l'utilisation de l'apprentissage machine. Il s'agit d'une méthode d'intelligence artificielle qui a l'avantage de construire elle-même les instructions dont elle a besoin pour traiter les données. Dès lors, cette méthode permettrait de fournir l'information contenue dans ces dossiers médicaux [21]. Mais l'apprentissage machine permettrait également de traiter de larges volumes de données, non structurées, avec des formats variés[21]. L'apprentissage machine permettrait ainsi de faire face à la nature complexe des données de vie réelle tout rendant possible la combinaison de bases de données. Une telle méthode serait ainsi mise en œuvre durant une étude à travers des phases d'organisation des données puis d'exploitation[24]. Cette dernière phase nous intéresse particulièrement car elle comporte un temps de préparation durant lequel le modèle d'apprentissage machine va développer ses propres instructions.

b) L'exploitation de ces données de vie réelle est encadrée

i. Un premier cadre sous l'angle de la construction de l'étude de données de vie réelle

1. L'étude de données de vie réelle

Les études de données de vie réelle portent sur des données non collectées durant les essais cliniques[12]. Ces études font référence aux données obtenues par toute méthode non interventionnelle pour décrire la réalité des pratiques cliniques [12]. Non interventionnelle signifie qu'il n'y a pas de changement des pratiques habituelles. La définition des études de données de vie réelle intègre donc bien une multitude de sources possibles [12] telles que les données issues des patients, des cliniques, des hôpitaux, des payeurs ou encore de la Sécurité sociales [12]. Ces différents acteurs vont ainsi alimenter les bases de données évoquées précédemment, qui pourront ainsi rentrer dans la conception de l'étude de données de vie réelle. En effet, la conception de l'étude comportera des éléments supplémentaires à la sélection des bases telles que la finalité de projet, un responsable de traitement, un traitement et une publication [24].

Ces études de données de vie réelle ne vont pas se substituer aux essais cliniques mais les compléter. Il est possible que ces études de données de vie réelle apportent des résultats similaires aux essais cliniques sans tenter de reproduire leurs critères d'inclusion ou d'exclusion[22]. Quoi qu'il en soit, ces résultats vont porter sur les observations cliniques et sur les évolutions des paradigmes sur les traitements[14]. Ces études vont ainsi pouvoir être utilisées pour évaluer le risque immunitaire lors d'une prise en charge au long cours par des biosimilaires. Cette évaluation du risque se traduit donc par la conception d'une étude qui dépend d'un encadrement réglementaire particulier.

2. L'encadrement des études de données de vie réelle

Un cadre légal des études existe avec la loi Jardé publiée au Journal Officiel le 6 mars 2012. Cette loi encadre les recherches impliquant la personne humaine à travers 3 catégories permettant une classification [25] (LOI n° 2012-300 du 5 mars 2012 relative aux recherches impliquant la personne humaine). La première de ces catégories regroupe les études dont les interventions sont à risques sur la personne humaine. Cela signifie qu'il y a une étude impliquant la collecte de données dans le cadre d'une prise en charge qui n'est pas habituelle.

La seconde catégorie désigne les interventions à faibles risques sur la personne, dont la liste est fixée par l'Agence Nationale de Sécurité du Médicament et des produits de santé (ANSM). La dernière catégorie regroupe les recherches sans modifications de la prise en charge habituelle, sans procédures complémentaires.

Toutes les études ne sont pas du ressort de la loi Jardé. Un traitement de bases de données peut être proposé pour répondre à la question de sécurité des biosimilaire dans un contexte précis. Il n'est donc pas question de modifier une prise en charge, ce qui orienterait la classification d'une telle étude vers la troisième catégorie de la loi Jardé. Or il est nécessaire pour cette 3e catégorie de faire la distinction d'une recherche selon qu'une collecte de nouvelles données est nécessaire ou non. En effet, dans le cas d'une étude prospective, avec une collecte de nouvelles données, la catégorie 3 est applicable. En revanche, dans le cadre d'une étude rétrospective, sur des données déjà collectées existantes dans des bases, la troisième catégorie n'est pas applicable. En effet, ces dernières ne portent pas sur la personne mais sur des données uniquement. De telles études ne seraient donc pas du ressort de la loi Jardé mais seraient soumises à des procédures établies par la CNIL (Commission nationale de l'informatique et des libertés).

3. Un cadre de l'étude selon les données

En ayant à l'esprit le cadre réglementaire, il est désormais question de savoir si l'étude sera prospective ou rétrospective. Tout d'abord, ces études ont pour vocation de compléter les études cliniques. Il peut s'agir par exemple d'une opportunité pour soutenir l'autorisation et de nouvelles indications[22]. Mais il peut aussi y avoir une volonté de conduire des études de sécurité post AMM, qu'elles soient interventionnelles ou non[26]. Or, des évènements indésirables pourraient être explorés dans le cas d'une évaluation de la sécurité [22]. Cela signifie que les évènements ont été observés et qu'il y a un besoin d'en comprendre la cause. En d'autres termes, les évènements ont eu lieu et l'investigateur souhaite connaitre les facteurs de risques qui ont joué un rôle dans la survenue de cet évènement indésirable. Ce raisonnement s'oppose dès lors à l'anticipation des évènements qui suppose une prédiction. Cette dernière se situe au-delà de l'établissement d'un lien de causalité [22] car elle vise à identifier les patients qui risquent d'expérimenter l'évènement [22].

Il faut désormais pouvoir faire un choix entre la compréhension de la cause et l'anticipation de l'évènement. Dans les deux cas, l'étude visera à répondre à des questions pertinentes [14] qui serviront de point de départ. Une question telle que l'impact d'une pratique de *switch* sur la tolérance d'un biosimilaire au long cours implique d'utiliser les données massives. Ces données massives imposeraient d'utiliser une méthode d'apprentissage machine dont le traitement consisterait à rechercher des tendances dans les données. Or cette méthode peut requérir une préparation du modèle effectuée avec des données dont l'issue de la prise en charge est connue. Cette préparation est déterminante et oriente donc vers la conception d'une étude rétrospective. De plus, ces études rétrospectives présentent un temps d'exécution plus court que les études prospectives, ainsi que des contraintes réglementaires allégées[23]. En effet, les études rétrospectives ne sont pas du ressort de la loi Jardé.

La méthode d'apprentissage machine une fois opérationnelle permettrait d'identifier des tendances dans les données provenant de bases. Il pourrait toutefois être envisagé dans un second temps de conduire des études prospectives afin d'assurer la validité des informations obtenues à partir des études rétrospectives [23]. Ces études prospectives nécessiteront des décisions scientifiques supplémentaires afin de définir l'exposition ou encore la population[20]. En effet, ces études présentent l'avantage de couvrir des paramètres supplémentaires qu'il serait pertinent d'explorer. Il serait par exemple intéressant d'ajuster la période couverte par l'étude selon la durée de stabilisation des prescriptions. Une étude prospective permettrait ainsi de tenir compte de la temporalité d'exposition au médicament, et d'adapter ainsi la collecte de données d'intérêt à la durée de l'étude [23]. Il y a donc certes une mise en place d'étude rétrospective dans un premier temps, mais les études prospectives trouveraient un intérêt dans un second temps.

ii. Un second cadre sous l'angle de la protection des données personnelles

1. L'encadrement du traitement des données personnelles

En plus de l'encadrement de l'étude, il faut tenir compte de l'encadrement du traitement des données. Les données personnelles et les données de santé sont toutes les deux définies par le Règlement Général de Protection des Données Personnelles (RGPD)[24]. Une donnée personnelle désigne « toute information se rapportant à une personne physique identifiée ou identifiable ». Le RGPD définit la donnée de santé comme les « données à caractère personnel

relatives à la santé physique ou mentale d'une personne physique, y compris la prestation de services de soins de santé, qui révèlent des informations sur l'état de santé de cette personne ».

Il faut remarquer que ces données ont un encadrement particulier qui conditionne leur exploitation. Le caractère personnel d'une donnée ne confère pas une quelconque propriété à la structure qui les collecte et les traite [24]. En effet, le législateur européen voit surtout un moyen d'impliquer le citoyen dans l'usage de ces données sans pour autant reconnaitre une véritable propriété [24]. Dès lors, les données personnelles ne peuvent pas être vendues [24], mais une base de données peut être exploitée si elle est anonymisée et agrégée[24]. Il y a ainsi deux façons d'exploiter les données, soit par traitement de données personnelles de santé, soit par traitement de données anonymisées [24]. Dans ce dernier cas de figure, le RGPG et la LIL (Loi informatique et libertés) ne s'appliquent pas[24].

2. *L'encadrement du traitement des données de santé anonymisées*

Avant toute chose, il est à noter que le protocole de l'étude est revu par un comité d'éthique afin d'assurer le caractère légitime du projet [24]. Mais à cette revue peuvent s'ajouter la nécessité d'obtenir une autorisation de la CNIL ainsi que la conduite d'une étude d'impact sur la vie privée [24]. En effet, l'accès aux données SNDS se fait dans le respect de certaines conditions [23] fixées par la CNIL. L'industriel doit notamment passer par un bureau d'étude ou un laboratoire de recherche indépendant[27], et suivre la procédure simplifiée MR006. D'autres procédures simplifiées telles que la MR004 encadrent les études n'impliquant pas la personne humaine mais ayant un caractère d'intérêt public [24]. Ce dernier intervient par exemple lorsque l'étude vise à garantir des normes élevées de qualité et de sécurité des soins [24]. Dans le cas des biosimilaires, il pourrait être pertinent d'intégrer des données génétiques. Cependant, les données génétiques sont considérées comme plus sensibles [24] et ont donc un cadre réglementaire plus strict.

3. *La préparation d'une base : garantir l'anonymisation*

Le responsable de traitement est garant de l'anonymisation des données de santé. Cette démarche est initiée très précocement avec l'analyse d'impact de la vie privée. Cette dernière permet ainsi d'identifier les actions préventives nécessaires pour limiter le risque de réidentification [24]. En effet, il s'agit de garantir qu'il n'y aura pas de perte de l'anonymat des

données et que l'analyse conduite par le modèle d'apprentissage machine ne permettrait pas de réidentifier les patients. Une réflexion pourra être ainsi menée sur les données utilisées, le modèle, le type de traitement, la finalité du traitement et le risque de réidentification afin de mettre en place les actions nécessaires.

Cette anonymisation fait partie de la préparation de la base de données[24]. Il s'agit d'un exercice qui peut être difficile car de nombreuses données personnelles peuvent exister parmi les sources de données. Or il y a une volonté d'affirmer que ces données personnelles ne permettent pas de remonter à l'individu, d'où l'anonymisation. Il peut donc être nécessaire d'avoir recours à des outils permettant la désidentification des données[23]. Il pourrait également être discuté de l'utilisation d'un algorithme pour réaliser cette tâche d'anonymisation. Une fois cette anonymisation réalisée, le responsable de traitement sera amené à assurer la traçabilité des accès. Après cette présentation succincte des contraintes réglementaires, il est désormais possible de présenter l'exploitation des données par une méthode d'intelligence artificielle.

c) Une utilisation initiale de ces données pour préparer l'intelligence artificielle

i. La nécessité de l'apprentissage pour préparer l'intelligence artificielle

1. Panorama de l'intelligence artificielle

Le dictionnaire Larousse définit l'intelligence artificielle (IA) comme un « ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine ». Il serait ainsi possible de voir l'IA comme un outil (modèle) qui réalise une série de tâches selon les instructions contenues dans un algorithme. Ce dernier correspond à une série d'instructions mathématiques qui une fois transcrite dans un langage informatique prend le nom de programme informatique. L'IA est donc de nature numérique, avec un code qui définira les tâches qu'elle réalise afin d'obtenir les résultats[28]. Le traitement des données effectué lors d'une étude est donc défini par ce code. Le terme d'intelligence trouve son origine dans la capacité du modèle à imiter certaines capacités humaines telle que la capacité d'apprentissage. Il faut entendre par apprentissage la capacité à adapter la tâche réalisée au regard de l'expérience. En l'occurrence, l'apprentissage vise à modifier le code informatique du programme au fur et mesure de l'utilisation de l'IA, améliorant ainsi la tâche réalisée (figure 9).

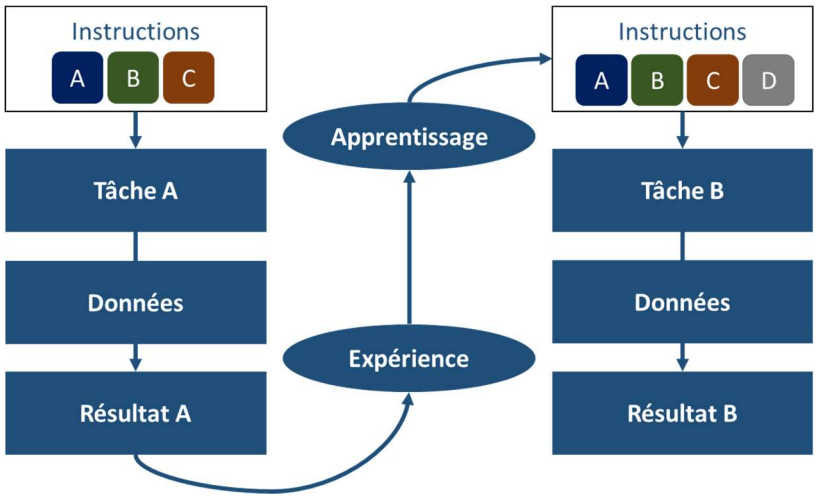

Figure 9: l'apprentissage

L'apprentissage ne désigne qu'une possibilité d'IA parmi tant d'autres (figure 10). Au sein de l'IA existe notamment la catégorie de l'apprentissage machine. Cette catégorie correspond à l'exploitation de données en utilisant l'apprentissage pour construire le modèle, mais également de l'adapter selon les changements. En effet, cette construction du modèle est permise par l'apprentissage *via* la soumission de données au modèle. Lorsque les données

changent l'apprentissage adapte la construction du modèle. Cet apprentissage est donc particulièrement utile pour analyser les prises en charges ayant une évolution rapide. Les applications de l'apprentissage machine sont vastes telles que les recherches dans les services de santé, l'économétrique ou encore l'épidémiologie[22].

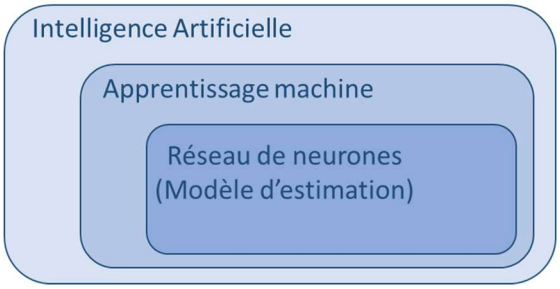

Figure 10: Intelligence artificielle

Dans le cas des biosimilaires, l'utilisation de l'apprentissage machine s'inscrit dans une approche épidémiologique. L'analyse serait en effet orientée vers l'exploration de la sécurité d'utilisation des biosimilaires par l'apprentissage machine. En ce sens, il pourrait être utilisé une méthode d'apprentissage machine, aussi désignée par le terme de modèles d'estimation. Il existe en l'occurrence plusieurs modèles d'estimation d'apprentissage machine tels que les arbres de décision ou encore les réseaux de neurones[29]. Ces derniers ne sont pas nouveaux [29], il s'agit d'une technologie mature [24] qui est montée en puissance ces dernières années. De plus, la littérature met en évidence que ces réseaux de neurones seraient plus efficaces que les méthodes d'arbres [29] pour des traitements complexes. Les réseaux de neurones pourraient donc présenter un intérêt dans l'exploration de la sécurité d'utilisation des biosimilaires.

Il faut remarquer que les technologies d'IA trouvent leur place depuis peu. Premièrement, il y a eu un changement dans les données. En effet, les données massives ont rendu difficile le traitement par des méthodes classiques[22], laissant ainsi la place à d'autres méthodes. Mais l'utilisation de ces méthodes d'IA ont été permises par des changements technologiques sur le stockage et la puissance de calcul[30]. Dans les faits, il est possible aujourd'hui de développer en peu de temps, le code d'un programme d'IA pouvant couvrir une large population. La littérature indique notamment 30 minutes de travail pour 100 lignes de codes couvrant une population de 100 0000 individus [29]. Il peut donc être discuté de l'utilisation d'un modèle d'apprentissage machine au regard des problématiques des biosimilaires.

2. *L'apprentissage machine : la tendance et la prédiction*

L'apprentissage est un ensemble de méthodes dont font partie les réseaux de neurones. Ces réseaux de neurones sont utilisés pour détecter des tendances parmi les données [30] mais également pour proposer des résultats au regard de ces données [31] (figure 11). Le terme de tendance désigne ici les relations parmi les données, aussi appelées corrélations. Ces relations peuvent être le lien entre une donnée et une autre dans l'obtention d'un résultat. Dans le cas des biosimilaires, il est possible de rechercher les relations entre les caractéristiques d'un biomédicament, un *switch* et la survenue d'une réaction immunitaire. Il pourrait ainsi y avoir une volonté d'identifier au sein d'un grand volumes de données des relations pour mettre en évidence des tendances[28]. La mise en évidence de tendance se traduirait par l'identification d'une variable comme étant importante [31] dans la survenue d'une réaction immunitaire. En d'autres termes, l'objet d'une analyse serait d'identifier des facteurs de risques significatifs.

La deuxième utilisation de ces méthodes d'apprentissage machine serait celle de la proposition de résultat. En effet, ces méthodes pourraient être utilisées pour prédire des données futures à partir des tendances identifiées [30,31]. Il s'agirait en l'occurrence que le réseau de neurones soit capable de proposer l'issue d'une prise en charge au regard des données collectées sur des prises en charges passées. Le réseau de neurones mettrait à profit les tendances identifiées des situations passées pour prédire des résultats pour de nouvelles données lorsque le résultat clinique est inconnu. Ceci pourrait être mis en œuvre dans le but de prendre des décisions dans des conditions d'incertitudes [30] afin de proposer le traitement le plus adéquat[32].

Cette méthode s'appliquerait particulièrement bien aux biosimilaires. En effet, il a été présenté que de nombreux paramètres peuvent influencer la réaction immunitaire lié à l'administration de biosimilaire (figure 11). Cette multitude de paramètres se traduit par un volume important et une complexité de données empêchant l'utilisation de méthodes classiques mais justifiant l'analyse par des réseaux de neurones[29]. Cette analyse pourrait ainsi identifier des corrélations [22] entre de nombreuses caractéristiques qui seront prédictives d'un résultat[30]. Il serait ainsi dans un premier temps possible de fournir une solution à un problème spécifique bien déterminé [29]. En l'occurrence ce problème serait celui d'évaluer un risque en considérant des influences complexes de différents paramètres[29]. Ces différents paramètres

peuvent correspondre aux composantes des biosimilaires qui pourraient augmenter le risque d'une réaction immunitaire.

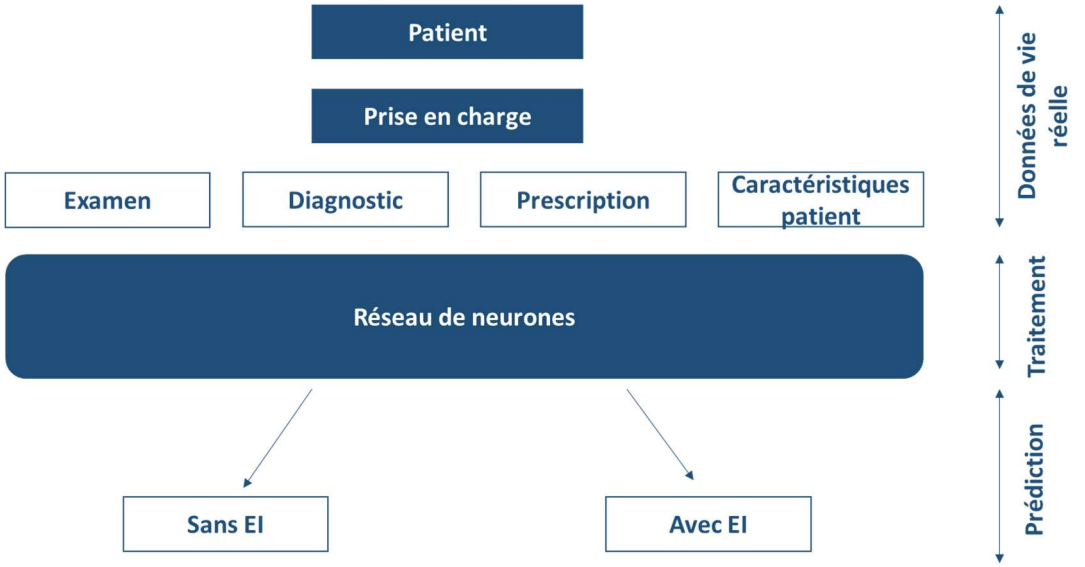

Figure 11: Utilisation d'une intelligence artificielle

3. *La description du traitement par l'apprentissage machine*

Le concept d'apprentissage machine est relié au traitement effectué par un réseau de neurones. En effet, afin d'identifier les tendances dans les données, le réseau de neurones doit d'abord réaliser un traitement. Il existe deux types de données avec lesquelles le réseau de neurones interagit : les données d'entrées et les données de sortie (résultats). Les données d'entrées sont toutes les données que le modèle reçoit pour traitement afin de produire un résultat qui sera qualifié de données de sortie[32]. Les données d'entrées peuvent être l'âge d'un patient, le nom d'un biosimilaire ou encore la pathologie traitée. Le résultat correspondra ici aux résultats cliniques tels qu'une réaction immunitaire. Les données d'entrées seront donc traitées par le réseau de neurones pour produire des données de sorties (résultat).

Le résultat est proposé par le réseau de neurones bien qu'il soit déjà connu dans la base de données de l'investigateur. En effet, une étude rétrospective est nécessaire afin que les données d'entrées et les résultats correspondants soient disponibles pour le réseau de neurones. La méthode est alors dite supervisée car certaines données sont identifiées comme étant les résultats réels[22]. La raison est que ce dernier proposera dans un premier temps des résultats erronés au regard des données d'entrées qu'il reçoit. Seul le processus d'apprentissage permettra au modèle d'être capable de proposer un résultat correct. Ce qui

in fine permettra au réseau de neurones d'isoler les données d'entrées pertinentes [32] et ainsi d'associer un résultat à des paramètres ayant joué un rôle important dans l'obtention de ce résultat[31]. Ce qui importe ici n'est pas seulement l'obtention d'un résultat correct, mais bien l'identification des paramètres importants.

Un exemple concret du traitement peut être présenté de la façon suivante. Le modèle d'apprentissage machine peut avoir pour mission de reconnaitre des chiffres écrits à la main. Les données d'entrées seront les chiffres écrits à la main soumis sous forme d'images. Le résultat sera le chiffre que le réseau de neurones proposera en fonction de l'image. Le modèle pourra se tromper et proposer un 5 au lieu d'un 6. Cette erreur l'amènera à corriger son raisonnement afin d'identifier parmi les images des éléments qui lui permettent d'avoir un résultat correct. Il peut s'agir d'éléments tels que les courbures ou la largeur des traits. Le même type de traitement pourrait être appliqué pour des données de prises en charges associées à des résultats cliniques.

4. La nécessité de l'apprentissage

Le réseau de neurones n'est pas capable au départ de fournir un résultat correct selon les données d'entrées qu'il reçoit. Cela s'explique par l'absence de programmation initiale. Puisque le réseau de neurones réalise une tâche définie dans son code, et que le code n'est pas préparé entièrement, alors le modèle n'est pas en mesure d'être exact. Le réseau de neurones propose ainsi un résultat, le compare à celui de la base et constate une erreur. L'objectif est alors pour le modèle d'ajuster son code au regard de cette erreur afin de corriger son erreur. Cette étape se répète jusqu'à ce que le modèle soit capable d'avoir un ensemble d'instructions lui permettant de proposer des résultats corrects.

Cette phase de correction des erreurs est appelée l'apprentissage. L'apprentissage permet, *via* l'expérience, d'améliorer la performance sur la réalisation d'une tâche[29]. L'expérience est en l'occurrence issue des données collectées, d'entrées et de sorties [29], essentielles à la constitution de cette expérience[32]. Plus il y a de données, plus le modèle d'apprentissage pourra réaliser sa tâche conduisant à la proposition d'un résultat qui sera comparé au résultat réel. Cette comparaison permet d'une part une mesure de la performance du modèle à proposer le bon résultat [29]. Mais elle permet d'autre part une correction du code du programme en cas de résultat erroné. *In fine*, l'apprentissage permettra d'obtenir un modèle

de réseau de neurones commettant peu d'erreurs, et étant précis dans l'identification de corrélation de données. Mais cet apprentissage permet également d'obtenir un réseau de neurones capable de traiter de nouvelles données et de proposer une probabilité d'obtention d'un résultat [29].

ii. Le fonctionnement de l'apprentissage dans le cas d'un modèle d'apprentissage machine spécifique : le réseau de neurones

1. Les données nécessaires à l'apprentissage

Des données sont nécessaires pour réaliser l'apprentissage. Il faudra en l'occurrence une grande quantité de données afin que l'apprentissage machine puisse fournir des prédictions précises [32]. La conception du modèle débute ainsi avec les jeux de données issus des bases de données, en l'occurrence les données génétiques, d'imagerie, de tests de laboratoires ou encore de biomarqueurs digitaux[30]. L'intégration de ces jeux de données peut se traduire par plusieurs milliers voire millions de données par patients [30].

Des données pertinentes devront être utilisées pour cet apprentissage afin de répondre à la question de sécurité des biosimilaires. Il s'agit ici des données de vie réelle liées à l'utilisation en soins courants des biosimilaires. Ces données seront issues des différentes bases de données, sélectionnées selon leur contenu. Il pourrait par exemple être décidé d'extraire les données de registres ou de dossiers médicaux afin d'obtenir des jeux de données intégrant des détails cliniques. Ces jeux de données pourront dès lors servir à l'apprentissage du modèle, qui une fois éduqué pourra traiter des volumes plus importants de données de même nature.

2. Le fonctionnement du réseau de neurones

Le réseau de neurones présente un fonctionnement particulier. Comme évoqué précédemment, le réseau de neurones réalise une tâche qui est de proposer un résultat selon les données d'entrée reçues. Il faut voir que cette tâche est réalisée par plusieurs neurones. Un neurone a pour but de transformer la donnée qu'il reçoit afin de l'envoyer au neurone suivant. Comme les neurones sont connectées les uns aux autres, les données transformées servent de données d'entrée pour le neurone suivant[29]. Le neurone suivant reçoit ainsi les données de plusieurs neurones en amont, permettant à ce neurone en aval une combinaisons pondérées des données reçues. Cette propagation de transformation est réalisée au sein d'un

réseau de neurones organisés en couches. Ainsi, la transformation réalisée par chaque neurone peut être simple, mais le niveau de complexité est atteint en réunissant les transformations d'un grand nombre de neurones, comme dans un cerveau humain[31].

Un réseau de neurones est constitué de trois couches de neurones dites initiale, intermédiaire et finale [29] (figure 12). La couche initiale reçoit les données d'entrée alors que la couche finale est celle qui fournit les données de sorties (résultats). Les couches intermédiaires quant à elles vont réaliser des transformations supplémentaires de données qui pourraient être qualifiées de traitement sélectif[31]. Dans les réseaux de neurones profonds, il peut y avoir de nombreuses couches intermédiaires afin d'avoir de nombreux traitements sélectifs. Ce sont ces couches intermédiaires, également dites couches cachées, qui vont identifier les interactions au sein des données [29]. Mais il y a une certaine opacité sur les transformations effectuées pour obtenir ces interactions. En effet, un manque de transparence du modèle sur les raisonnements ayant permis d'obtenir un résultat a conduit la communauté scientifique à parler de boite noire.

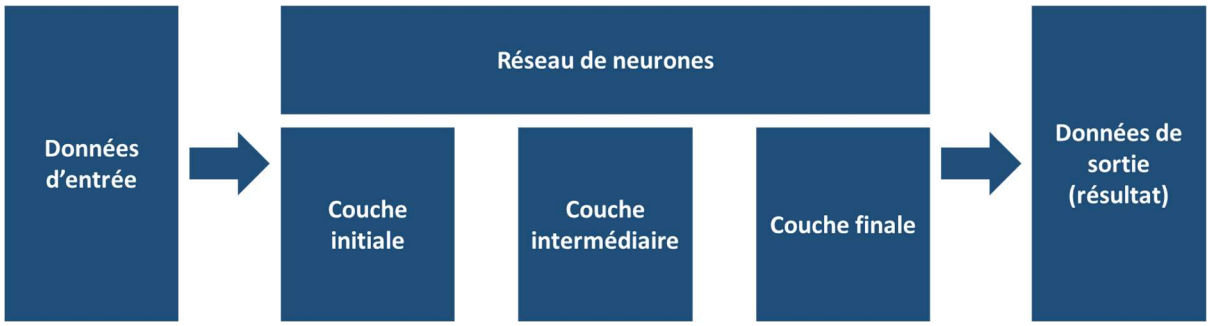

Figure 12: Le réseau de neurones

3. *L'apprentissage appliqué au réseau de neurones*

L'apprentissage a pour vocation d'ajuster les transformations réalisées par les neurones afin que le réseau de neurones puisse proposer un résultat correct. Il y a donc une mesure de la différence entre le résultat proposé et le résultat réel. Cette différence fait ensuite l'objet de l'utilisation d'une méthode d'optimisation[29]. Cette dernière permet de réduire la différence qu'il existe entre le résultat prédit et le résultat réel en modifiant les paramètres d'estimation du modèle [29]. Il s'agit ici d'un exercice itératif afin d'ajuster ce réseau de neurones [29] jusqu'à ce qu'il soit capable de proposer le bon résultat pour les données qui lui sont soumises. Il faut noter dès à présent que l'apprentissage n'est jamais terminé et que l'exercice sera répété pour de nouvelles données.

iii. Les phases de l'apprentissage : 3 jeux de données pour obtenir le modèle

1. Les trois jeux de données de l'apprentissage

Il est nécessaire d'éduquer le modèle sur des jeux de données [29] qui sont moins volumineux que l'ensemble de données qu'il est nécessaire d'analyser. Il s'agit en l'occurrence des jeux de données d'entrainement, de validation et de test [29]. Ces derniers ont pour vocation d'être représentatif de la population générale dont sont issus les jeux de données d'apprentissage. Dans un premier temps, le jeu de données d'entrainement permet de mettre au point les paramètres d'estimation [29]. Ce premier jeu permet en quelque sorte l'élaboration des instructions que le réseau de neurones va utiliser pour fournir un résultat. Dans un second temps, le jeu de données de validation permettra d'affiner les paramètres d'estimation [29]. Il s'agit ainsi de données permettant d'obtenir un modèle fournissant les résultats les plus exacts possibles. Le jeu de données test est utilisé en dernier afin d'évaluer le caractère généralisable du modèle [29]. Ce dernier jeu permet d'une part de s'assurer que le modèle donne le bon résultat. D'autre part, il permet de garantir que le modèle puisse trouver les bons résultats sur des nouveaux jeux de données légèrement différents de ceux de l'apprentissage.

2. L'apprentissage pour obtenir un modèle généralisable

Grâce à l'apprentissage, le réseau de neurones propose un résultat clinique identique à celui réellement observé. Mais il existe un risque que le réseau de neurones devienne trop efficace sur les jeux de données d'apprentissage qui lui sont soumis. En d'autres termes, le modèle serait sujet à un sur-apprentissage entrainant une perte de la capacité à prédire des résultats corrects pour de nouveaux jeux de données. Ce sur apprentissage s'explique par une mémorisation des points de données par le réseau de neurones[29]. C'est donc pour éviter ce sur-apprentissage que le jeu de données test est caché et inutilisé jusqu'à l'évaluation finale[21]. Il est important d'éviter ce sur-apprentissage pour avoir un modèle généralisable. Ainsi le réseau de neurones capture la relation générale entre les données et les résultats [29]. Dès lors, ce modèle pourra être utilisé pour de nouvelles analyses sur de nouvelles données.

3. *Les avantages d'un modèle généralisable*

Un modèle généralisable peut traiter correctement des jeux de données présentant des variations par rapport aux jeux de données d'apprentissage. Ce caractère généralisable sera précieux dans le cas des données de vie réelle. En effet, une première extraction de base de données fournira des données de vie réelle représentant une réalité à un moment précis. Toutefois, la génération de données de vie réelle n'est pas un processus stationnaire [29]. De nombreuses composantes d'une prise en charge peuvent évoluer. Il peut y avoir une évolution des pratiques ou encore des pathologies. La littérature souligne par exemple l'importance de tenir compte de l'émergence de résistance bactérienne et de l'impact sur les données générées[29]. Il serait donc possible de partir du principe que les biosimilaires aussi subiront des évolutions dans leur utilisation. Or ces évolutions impacteront les données de vie réelle qui sont générées, mettant ainsi à contribution le caractère généralisable du modèle.

Cette capacité d'adaptation du modèle, bien que cruciale, pourrait avoir des limites. Dans un premier temps parce que les données soumises pour analyses pourront être un peu différentes de celles ayant servi à l'apprentissage. Dans un second temps parce qu'il serait intéressant d'utiliser le modèle pour conduire de nouvelles analyses. Or ces nouvelles analyses pourraient être un réel challenge pour le modèle et mettre en défaut ses capacités d'adaptation. Il serait également possible que certaines évolutions des pratiques rendent impossible l'ajustement des instructions du modèle au regard des nouvelles données. Il peut donc être discuté de l'application d'un modèle éduqué pour répondre à un type de question, ainsi que de la transposabilité de ce modèle pour répondre à de nouvelles questions associées à de nouvelles données.

L'intelligence artificielle aurait donc un intérêt dans la recherche des tendances au sein de grands volumes de données. Une première analyse pourrait explorer une particularité liée à l'utilisation des biosimilaires. Il sera discuté dans une troisième partie de la pertinence d'une multiplication des analyses par un même modèle afin de couvrir les spécificités des biosimilaires.

3. L'exploitation des données de vie réelle par l'apprentissage machine pourrait offrir plusieurs niveaux de réponses

a) L'utilisation d'un modèle éduqué pour répondre à une première question liée aux biosimilaires

i. Les principes généraux du modèle d'apprentissage machine éduqué dans le contexte des biosimilaires

1. *Un modèle d'apprentissage machine pour une question spécifique aux biosimilaires*

Il existe de nombreuses composantes liées aux biosimilaires qui pourraient potentiellement avoir un impact sur la sécurité d'utilisation. Certaines de ces composantes comportent des éléments qui joueront un rôle dans la survenue d'une réaction immunitaire, et d'autres non. C'est afin d'identifier ces éléments qu'une analyse des données d'utilisation des biosimilaires en soins courants peut être menée. Il y aurait toutefois un premier exercice consistant à identifier les composantes qui doivent être explorées afin de déterminer les données nécessaires à l'analyse. En effet, la réaction est multifactorielle : la variabilité du produit, les pathologies traitées, les comorbidités, les prédispositions génétiques des patients ou encore les pratiques médicales[6,9].

La multitude de composantes justifierait l'utilisation de l'intelligence artificielle pour répondre aux questionnements de la communauté de santé. Il faut cependant remarquer que chaque composante ne se traduirait pas par la même question médicale. En d'autres termes, démontrer la sécurité d'interchangeabilité du biosimilaire serait à distinguer de la démonstration de la sécurité d'un biosimilaire bénéficiant de l'extrapolation d'indication. De plus, il serait sans doute difficile d'accéder à toutes les bases de données requises pour l'analyse de plusieurs questions trop différentes. Il semblerait donc raisonnable de sélectionner une première composante avec une question unique comme point de départ.

Dans la démarche générale, une question médicale pertinente peut être formulée. Il est dès lors possible de concevoir l'étude autour de cette question, en identifiant les données spécifiques à cette question. Il serait également nécessaire de définir les critères scientifiques et les accès aux bases afin de réaliser les étapes d'apprentissage du modèle. Le modèle, ici un

réseau de neurones, serait éduqué en fonction des données pour répondre à la question. Le réseau de neurone une fois opérationnel serait utilisé pour réaliser l'analyse sur un grand volume de données issues des bases sélectionnées. Les résultats seraient obtenus avec cette première analyse, tout en gardant à l'esprit que le modèle pourrait permettre de conduire de nouvelles analyses.

2. Une question d'interchangeabilité

La première question qui peut être abordée en tant qu'exemple est celle de l'interchangeabilité d'un médicament biologique de référence par son biosimilaire. Cette question peut être formulée par le raisonnement suivant. Un patient est traité par un médicament biologique de référence pendant plusieurs mois puis se voit prescrire le seul biosimilaire sur le marché à la place. Cette pratique est conduite chez de nombreux patients éligibles à l'interchangeabilité. Ce *switch* s'accompagne d'une prise en charge de patients sur une longue durée avec un suivi de pharmacovigilance. Ayant connaissance des réactions immunitaires qui ont pu survenir, il sera question de savoir si cette interchangeabilité a augmenté la survenue de ces réactions. Dès lors, le réseau de neurones pourrait être utilisé pour mettre en évidence l'importance d'éléments spécifiques à la prise en charge. Il pourrait par exemple être indiqué que l'interchangeabilité chez des patients avec des comorbidités particulières ont expérimenté davantage de réactions immunitaires.

3. Des données adaptées à la question d'interchangeabilité

Pour que cette analyse puisse apporter des réponses, il serait nécessaire d'obtenir des données spécifiques à cette composante d'interchangeabilité. Il s'agit en l'occurrence de données rétrospectives, où l'issue de la prise en charge est connue. Il serait notamment pertinent d'obtenir les données liées aux traitements biologiques initiaux et interchangés. Mais il serait également intéressant d'obtenir les données liées aux résultats cliniques ainsi qu'à la réaction immunitaire associée à la gravité de la réaction[22]. Il pourrait également être utile de connaitre la cause de l'interchangeabilité puisqu'une rupture d'approvisionnement n'implique pas les mêmes risques qu'une réaction allergique au biomédicament de référence.

A ces données spécifiques pourraient être associées des données plus générales. L'intérêt du réseau de neurones est de traiter des volumes importants et complexes de données. Il s'agit donc d'une opportunité d'exploiter la puissance du modèle en associant des données plus

« secondaires » pour en évaluer l'impact sur la survenue d'une réaction immunitaire. Ces données plus générales de prises en charges peuvent concerner par exemple l'établissement de santé, la région afin de mettre en évidence des divergences de pratiques.

ii. Les résultats attendus de l'analyse réalisée par le modèle

1. Un résultat associé aux tendances recherchées

Cette recherche de tendances est fondée sur l'hypothèse que les réactions immunitaires ne sont pas aléatoires. Comme elles ne sont pas aléatoires, il est supposé qu'il existe une possibilité de comprendre leur survenue. Donc cette hypothèse soutient que des éléments pourraient influencer la survenue de cette réaction. Ainsi, d'un point de vue analytique, il serait possible de mettre en évidence que des données sont associées à d'autres dans l'obtention d'un résultat. Or les résultats et les tendances fournis par un modèle étant liés, il est d'abord nécessaire de construire le modèle *via* l'apprentissage. Cette construction permet d'obtenir un modèle dont le « raisonnement » aboutit au résultat correct. C'est au sein de ce raisonnement que les tendances ou interactions entre les données peuvent être identifiées.

2. Le résultat : une classification adaptée à la problématique des biosimilaires

Le résultat est donc lié à l'identification de tendances dans les données. Or l'utilisation d'un réseau de neurones fournit un résultat de classification. En effet, ce réseau de neurones serait utilisé pour classer, selon les données de prises en charge, un patient selon qu'il a expérimenté ou non une réaction immunitaire. Dès lors, l'identification de tendances est dépendante de cette classification. C'est au sein de ce raisonnement de classification que résident les interactions, les tendances, ayant conduits à cette classification. De prime abord, il pourrait être envisageable d'avoir deux catégories telles que l'absence et la survenue de réaction immunitaire. Il serait toutefois discutable de créer davantage de catégories afin d'ajuster le réseau de neurones à la problématique d'interchangeabilité.

Il pourrait y avoir plusieurs approches pour déterminer des catégories adaptées à la question. Il est en l'occurrence pertinent d'identifier si une pratique de *switch* augmente le risque. Des catégories pourraient donc être déclinées selon les pratiques. Il pourrait ainsi y avoir une première catégorie en l'absence de réaction immunitaire, une seconde pour l'utilisation unique d'un biosimilaire ou d'une référence avec survenue d'une réaction immunitaire. La troisième catégorie pourrait être celle de la survenue d'une réaction immunitaire en cas de

pratique de *switch*. La décomposition de ces cas de figure serait intéressante pour créer des groupes au sein desquels il existe la plus petite variabilité[29].

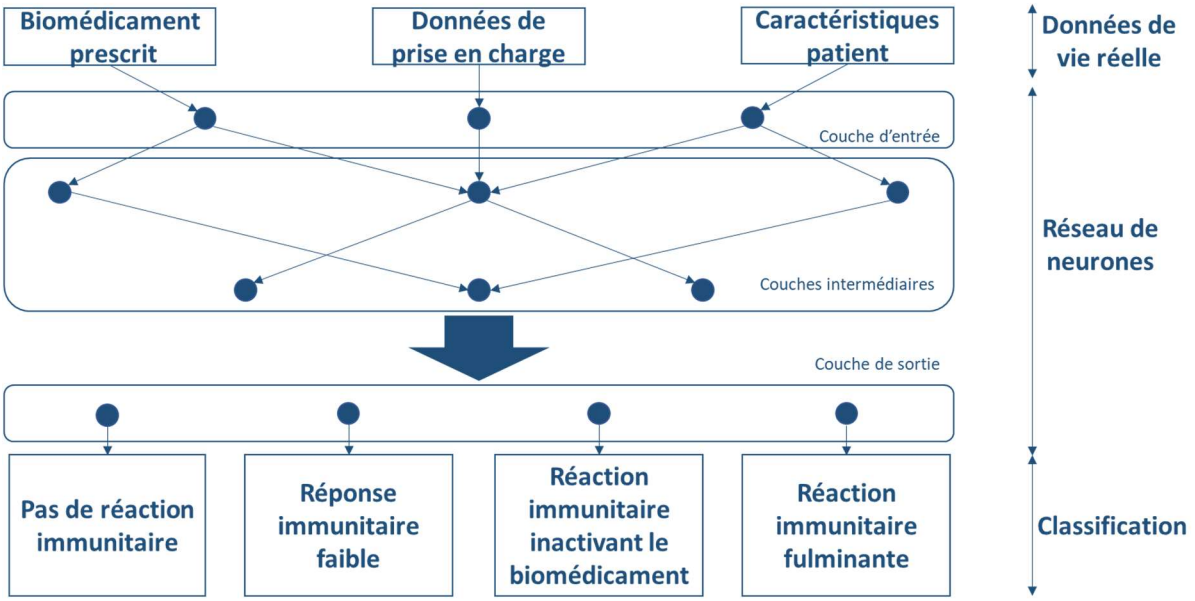

Figure 13: Une classification d'un réseau de neurones adaptée aux réactions immunitaires

Cette classification pourrait toutefois faire défaut. Il serait possible que le réseau de neurones, dans son apprentissage, se focalise uniquement sur les pratiques. En d'autres termes, il apprendrait seulement à classer un patient selon qu'il a expérimenté ou non un *switch* ou l'absence de *switch*, plutôt que les facteurs concourant à la survenue d'une réaction immunitaire. Le raisonnement identifiant les interactions entre les données pourraient ainsi être biaisé et peu informatif. Une autre approche pourrait donc conduire à proposer des catégories selon les réactions immunitaires uniquement (figure 13). Il pourrait ainsi y avoir une déclinaison des catégories selon la nature de la réaction immunitaire. La gravité clinique, le délai de survenue ou le type de réaction pourraient être des critères de classifications afin d'obtenir des groupes avec une faible variabilité intra groupe. Le raisonnement conduisant à la classification de patients au sein de ces groupes pourrait être informatif par l'identification de tendances pertinentes.

Une remarque dès à présent sur cette classification qui conditionne de futures analyses. En effet, dans le cas d'une analyse rétrospective, le raisonnement est important puisqu'il met en évidence des facteurs de risques. Toutefois, dans une analyse valorisant les résultats proposés par le réseau de neurones, la classification aura également son importance. Nous souhaitons obtenir une prédiction quant à l'issue d'une prise en charge dont le résultat reste inconnu. Le

réseau de neurones va pour cela classer le patient dans l'une des catégories définies par l'investigateur. Or si le modèle doit classer en fonction des pratiques et que l'interchangeabilité n'a pas été observée, il ne pourra pas classer selon toutes les catégories. En effet, une catégorie avec interchangeabilité ne sera pas accessible si le praticien prescrit tout le long le même biomédicament. Il n'y aurait ainsi pas accès à la partie du raisonnement qui classe dans la dernière catégorie. Il pourrait ainsi y avoir une perte d'information contrairement à une classification basée sur les natures des réactions.

3. Des interactions au sein des données d'utilisation des biosimilaires

Le raisonnement d'un réseau de neurones est focalisé sur l'identification d'interactions entre les différentes données. Il pourrait en effet exister des liens entre les données des biosimilaires, des patients ou de la pathologie. En effet, le lien serait ici la majoration du risque de réactions immunitaires. Le réseau de neurones permettrait donc d'identifier ce lien entre les données. Les différentes couches de ce réseau vont alors rechercher des relations entre les différents types de données. Les neurones d'une première couche cachée (intermédiaire) pourraient rechercher les interactions au sein d'un premier groupe de données telles que l'âge, le sexe et le diagnostic[29]. Ce dernier peut être un succès thérapeutique ou une réaction immunitaire. Au sein de cette même couche, d'autres neurones pourront rechercher les interactions entre le produit, le lot, le dosage et le diagnostic. Puis une seconde couche intermédiaire pourrait à son tour explorer les interactions mises en évidence par la première couche intermédiaire. C'est la combinaison de ces interactions qui permettrait *in fine* la classification dans un groupe [29] avec ou sans réaction immunitaire.

4. L'identification des facteurs de risques via les interactions

L'identification de ces interactions permettrait de déterminer l'importance de certaines variables. L'analyse pourrait ainsi mettre en évidence des facteurs de risques. En effet, parmi tous les individus ayant expérimenté une réaction immunitaire, certains points communs pourraient ressortir. Ces points communs sont recherchés *via* les interactions afin de déterminer des facteurs de risques. Autrement dit, un paramètre associé à une augmentation de la fréquence d'une réaction immunitaire pourrait être perçu comme un facteur de risque. Ces derniers permettraient d'alimenter une réflexion lors d'une décision d'interchangeabilité.

Des facteurs de risques spécifiques pourraient être mis en évidence. Il pourrait notamment y avoir l'identification d'un risque plus important lors de l'utilisation d'un biosimilaires pour une indication donnée pour des patients présentant certaines comorbidités. Tout l'intérêt de cette analyse, est d'avoir connaissance de ces facteurs de risques pour la prise de décision d'interchangeabilité de nouveaux patients. Le praticien pourrait ainsi adapter sa prescription en sachant qu'un risque accru de réaction immunitaire existe en cas de *switch* par le biosimilaire pour un patient atteint de cette comorbidité particulière. Cette prise de décision dépendrait toutefois de l'interprétation des résultats obtenus par le réseau de neurones.

iii. Un modèle qui pourrait être utilisé pour différentes questions

1. Un réseau de neurones utilisé à nouveau

L'apprentissage du réseau de neurones a montré que ce dernier était généralisable. Il a donc été utilisé pour traiter un grand volume de données pour mener une première analyse. A chaque analyse, le réseau de neurones apprendra des données qui lui sont soumises. En effet, ces données vont constituer l'expérience que le modèle va acquérir. Une nouvelle analyse pourrait être envisagée sur des données plus récentes que celles de la première analyse. En effet, une nouvelle analyse implique de nouvelles données et potentiellement de nouvelles tendances bien qu'il n'y ait pas de changement significatif. Le réseau de neurones pourrait donc être appliqué sur de nouvelles données pour identifier de nouvelles tendances, ou une modification des facteurs de risques identifiés. Il pourrait en effet y avoir une modification de la relation des données et un paramètre pourrait ne plus être associé à une augmentation du risque de réaction immunitaire. Mais au-delà de conduire une nouvelle analyse sur des données plus récentes, il pourrait être possible d'explorer davantage la sécurité des biosimilaires.

2. Un réseau de neurones pour couvrir les spécificités des biosimilaires

Une première analyse est conduite pour évaluer une composante liée aux réactions immunitaires lors de l'utilisation des biosimilaires. Cette analyse portait dans un premier temps sur le lien entre l'interchangeabilité d'un biosimilaire et la survenue d'une réaction immunitaire. Or il existe d'autres composantes liées aux réactions immunitaires qu'il serait pertinent d'étudier *via* l'analyse d'un réseau de neurones. Les données de sortie (résultat) ne changeraient pas mais les données d'entrées pourraient changer selon les composantes

étudiées. Ces données dépendraient donc des nouvelles questions auxquelles l'investigateur souhaiterait répondre. Il y aura donc la question de savoir si le modèle tel qu'il a été conçu peut répondre à ces nouvelles questions.

3. La transposabilité : un même modèle pour de nouvelles question ?

Un modèle de réseau de neurones apprend théoriquement de toutes les données qu'il reçoit, lui permettant dès lors d'adapter son raisonnement et donc d'identifier des tendances. Mais il serait possible que les jeux de données soient suffisamment différents pour que le modèle précédemment utilisé ne puisse mener l'analyse à terme. Or de nouvelles questions impliquent de nouvelles données qui pourrait mettre à rude épreuve le caractère généralisable du modèle. Il serait donc nécessaire de démontrer la transposabilité du modèle[21]. En d'autres termes, l'analyse pourrait être conduite sur un établissement de santé puis sur un second après avoir démontré que le modèle fonctionne malgré des différences entre les établissements. Cependant, il pourrait y avoir une limite à partir de laquelle les différences entre les établissements sont trop importantes pour que le réseau de neurones puisse traiter les données. Ces différences pourraient impliquer une nouvelle phase de conception du réseau de neurones lui-même. Nous ne développerons pas davantage ces limites techniques mais elles pourraient constituer un frein à la réutilisation d'un même modèle.

b) L'utilisation d'un modèle d'apprentissage peut se décliner pour mieux répondre aux spécificités des biosimilaires

i. Nouvelle question médicale : qu'en est-il de l'interchangeabilité d'un nouveau biosimilaire

1. Un changement mineur : intégration d'un nouveau biosimilaire

La première analyse menée intégrait un biomédicament de référence et son biosimilaire. Il y a eu une première évaluation de la sécurité d'utilisation du biosimilaire en cas d'interchangeabilité. Des résultats ont ainsi été obtenus tels que les tendances dans les données ayant abouti à l'identification de facteurs de risques. Mais les tendances mises en évidences ne reflètent que l'utilisation de ce seul biosimilaire et de son biomédicament de référence. Dès lors, un nouveau biosimilaire qui serait mis sur le marché pourrait affecter les tendances des données observées. Ce biosimilaire justifierait ainsi de conduire une nouvelle analyse avec le même modèle afin d'étudier de la même façon les tendances.

2. Un nouveau biosimilaire, une nouvelle question

Une nouvelle question peut donc être formulée au regard du biosimilaire supplémentaire mis sur le marché. Quel serait l'impact de ce nouveau biosimilaire sur la sécurité d'utilisation d'un biosimilaire dans une pratique de *switch* ? Il s'agit d'une analyse très proche de la précédente intégrant cette fois-ci le nouveau biosimilaire. Ce dernier pourrait par exemple impacter les pratiques pour des raisons très variables. Le nouveau biosimilaire pourrait avoir par exemple une meilleure pénétration du marché hospitalier que le premier biosimilaire. Cela pourrait ainsi augmenter la prescription de ce nouveau biosimilaire et faciliter par la même occasion l'interchangeabilité. Il pourrait en ce sens y avoir une augmentation des réactions immunitaires liées à ce nouveau biosimilaire. Il ne serait toutefois pas possible d'affirmer que l'utilisation de ce biosimilaire soit réellement en cause. En d'autres termes, les réactions immunitaires pouvaient être précédemment observées avec la référence sans qu'il n'y ait d'interchangeabilité. Le nouveau biosimilaire pourrait simplement conduire à un transfert des réactions immunitaires depuis la référence vers le biosimilaire. Quoi qu'il en soit, il serait intéressant de répondre à cette nouvelle question.

3. *De nouvelles tendances : une comparaison de deux analyses*

Une première analyse a été conduite fournissant des résultats et permettant l'identification de facteurs de risques. L'intérêt de l'intelligence artificielle, à travers l'utilisation d'un réseau de neurones, est d'exploiter les résultats d'une première analyse pour les comparer à une seconde. L'objet de cette seconde analyse n'est pas seulement de répondre à la question d'impact du nouveau biosimilaire. En effet, cette seconde analyse serait l'opportunité de comparer les résultats obtenus avec la première analyse. Cette comparaison est l'occasion d'observer les évolutions générées par le biosimilaire (figure 14).

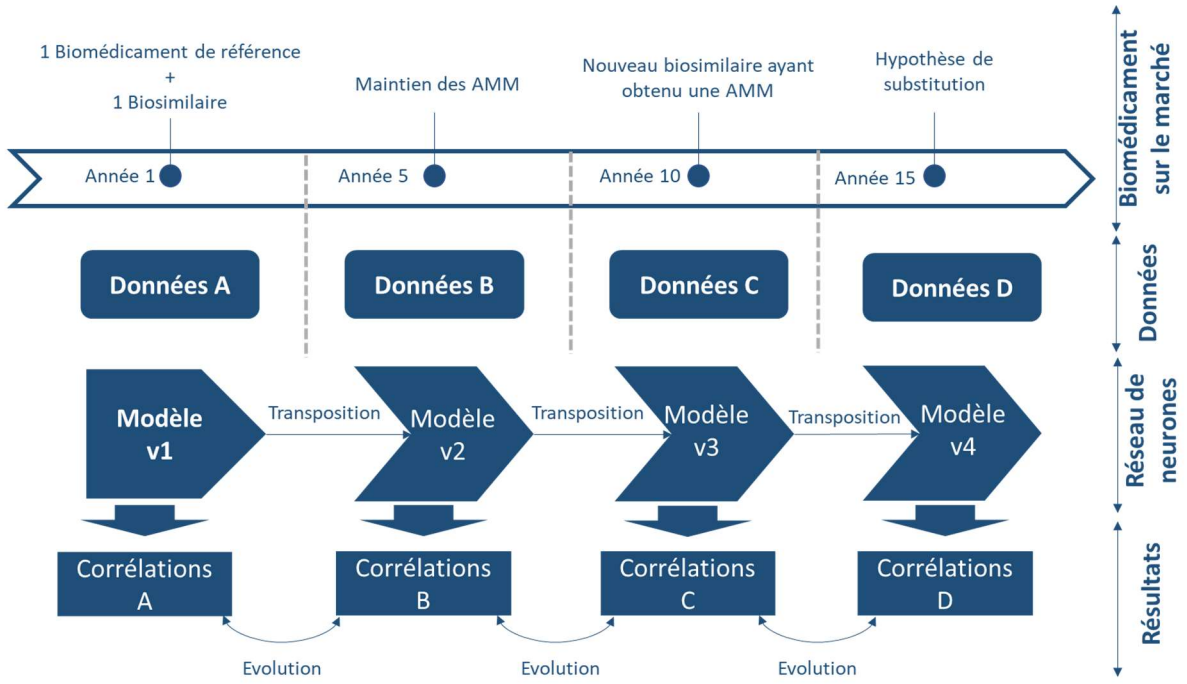

Figure 14: Déclinaison de l'utilisation des modèles de réseau de neurones

Il serait possible d'explorer les modifications de pratiques, les évolutions des population traitées tout en gardant comme cibles les réactions immunitaires. Cette analyse pourrait ainsi être riche dans l'objectivation de risques émergents mais aussi de démonstration d'un bon niveau de sécurité. D'un point de vue méthodologique, il serait également possible de valoriser la constance de la méthode et donc la fiabilité des résultats suite à la réutilisation d'un même modèle.

ii. Une nouvelle règle : un modèle pour une question plus complexe

1. Un modèle adapté à la question de la substitution des biosimilaires ?

Les analyses précédentes visent l'étude de biosimilaires et d'un biomédicament de référence au regard de pratiques définies. Ces pratiques sont celles de prescription, de traitement au long cours impliquant des renouvellements, une possibilité de *switch* et une délivrance fidèle à l'ordonnance. Il peut y avoir des différences mineures dans les pratiques selon les établissements de santé telles qu'une spécialité davantage prescrite qu'une autre. Ces différences peuvent être mises en évidence par une analyse des tendances qui succèderait les premières analyses.

Mais il pourrait y avoir des changements plus importants. En effet, la substitution par le pharmacien pourrait faire partie de ces changements majeurs. Cette dernière impacterait nécessairement les pratiques puisqu'il s'agirait d'un droit supplémentaire de délivrance d'un biomédicament différent de celui prescrit. Ce droit de substitution, abrogé par l'article 42 de la LFSS 2020 notamment pour des raisons de sécurité sanitaires, pourrait à terme faire l'objet d'une nouvelle réflexion de la part du législateur. Afin de conduire celle-ci, il serait possible dans un premier temps de valoriser les résultats d'analyses de l'interchangeabilité. Par la suite, il pourrait être évoqué le développement d'un système d'évaluation du risque d'une hypothétique substitution pharmaceutique. Dès lors, cette approche permettrait d'apporter des éléments de réponses afin de nourrir les débats sur la substitution.

2. Valorisation du modèle au service du principe de précaution

Le droit de substitution a été abrogé en partie pour des raisons de sécurité sanitaire. Le concept de sécurité sanitaire est associé à plusieurs principes tels que le principe d'évaluation et de précaution[33]. Ce dernier porte sur l'identification de risques avérés ou hypothétiques. Un modèle de réseau de neurones portant sur l'interchangeabilité des biosimilaires pourraient dès lors être valorisé au nom de ce principe de précaution. Cependant, le modèle serait utilisé afin d'aborder la question de substitution des biosimilaires et les potentiels risques qui lui sont associés. Pour cela, le modèle de réseau de neurones pourrait être ajusté afin de réaliser une simulation de l'impact d'une hypothétique substitution.

En effet, la substitution consiste en la délivrance d'un médicament biologique similaire à la place de sa référence. En l'état actuel, ce droit est abrogé, dès lors un biomédicament prescrit

est un biomédicament délivré. Or, il serait possible de simuler cette substitution en isolant le paramètre de prescription. L'intérêt d'une telle simulation serait de mettre en évidence un niveau de risque variable selon que le patient reçoit un biomédicament de référence ou le biosimilaire. Ainsi, le modèle pourrait être utilisé pour identifier les patients présentant un haut niveau de points communs à l'exception du biomédicament prescrit. Dès lors, les résultats cliniques pourraient être comparés selon que le patient reçoit le biomédicament de référence ou son biosimilaire. Une telle analyse pourrait mettre en évidence des facteurs de risques avérés ou hypothétiques selon la prescription réalisée. Ces facteurs de risques viendraient ainsi alimenter la réflexion quant à la question de la substitution des biosimilaires.

3. La promesse d'un modèle au service du principe d'évaluation

L'abrogation du droit de substitution pour des raisons de sécurité sanitaire intègre également le principe d'évaluation et de traçabilité. Ces deux notions pourraient être abordées parallèlement afin d'alimenter la réflexion quant à la question de substitution des biosimilaires. En effet, l'identification de risques au regard des pratiques actuelles ne serait pas suffisante pour appréhender la question de substitution. Cette dernière serait à l'origine de potentiels bouleversement qui requièrent de nouvelles analyses. Mais cette fois ci, des analyses seraient conduites par un modèle adapté aux évolutions des pratiques. La question de la substitution pourrait dès lors être dépendante du développement d'un modèle apte à évaluer les risques.

Toutefois l'aptitude d'un modèle à absorber un changement de pratique n'est pas acquise. En effet, un médicament prescrit ne serait plus nécessairement un médicament délivré. Dès lors, de nouvelles phases d'éducation du modèle de réseau de neurones seraient nécessaires afin de permettre une évaluation du changement de profils de sécurité des biosimilaires en cas de substitution. De plus, la traçabilité des biomédicaments délivrés sera essentielle afin de garantir une évaluation correcte des risques. Cette traçabilité conditionne effectivement l'aptitude du modèle à associer un effet clinique à un biomédicament. Ainsi tout défaut de traçabilité, et donc de la qualité des données, serait préjudiciable à l'analyse. Remarquons toutefois que la mise à disposition de données de traçabilité pourrait requérir une combinaison de dossiers médicaux et dossiers pharmaceutiques. Or ce recoupement soulève de nombreuses questions relatives au chainage de bases, à l'accès aux bases ainsi que la protection des données personnelles qui ne seront pas développées dans ce document.

iii. L'intérêt de la conduite d'une analyse sur de nouvelles données collectées au cours de la prise en charge : l'étude prospective

1. Étude prospective : l'intérêt de la collecte de nouvelles données

Il était jusqu'à présent discuté de la mise en place d'analyses rétrospectives sur des données déjà collectées. Il pourrait toutefois être pertinent de mener des analyses prospectives en collectant de nouvelles données pour suivre les patients dans le temps. Il y a trois avantages à mener ces analyses prospectives. Le premier est que de telles analyses permettraient de démontrer la précision des prédictions [21] et ainsi la précision de facteurs de risques obtenus. Un second avantage peut être mis en valeur telle que l'adaptation de la fenêtre de collecte à l'étude. En effet, une question médicale pourrait requérir la collecte de données sur une fenêtre de temps précise que les bases de données déjà constituées ne permettent pas de couvrir. Le troisième avantage se trouve dans le choix de ce qui pourrait être collecté. En effet, ces données prospectives pourraient trouver leur intérêt lorsque les données disponibles ne permettent pas une estimation des effets en causes [21]. En d'autres termes, les données issues des bases ne nous permettent pas de répondre à la question, ce qui implique une collecte de données spécifiques. L'approche prospective pourrait donc être perçue comme une analyse sur mesure, avec un registre par exemple qui permettrait d'intégrer de nouvelles données. Dès lors, ces nouvelles données contribueraient à l'approfondissement de l'exploration de la sécurité d'utilisation des biosimilaires en ajoutant des paramètres supplémentaires. Il pourrait notamment être exploré la relation entre des données génétiques et la survenue d'une réaction immunitaire.

2. L'opportunité d'intégrer des données spécifiques

Toute cette démarche analytique de données massives par de l'intelligence artificielle est menée pour comprendre les éléments associés à une réaction immunitaire lors de l'utilisation de biosimilaire. Plusieurs composantes ont été présentées comme pouvant jouer un rôle dans la survenue de cette réaction. Il y a néanmoins des éléments qui pourraient ne pas être pris en compte dans ces composantes. En effet, certains éléments liés aux patients pourraient jouer un rôle dans la survenue de ces réactions et pourtant ne pas être intégrés dans l'analyse. Cela pourrait être le cas des données génétiques. Il pourrait être pertinent de connaitre le risque de réaction immunitaire selon le terrain génétique du patient. Cette exploration de

l'immunogénicité pourrait se traduire par la mise en évidence de certains gènes qui seraient associés à des réactions immunitaires.

Une illustration peut être présentée concernant cette relation entre les données génétiques et la survenue de réactions immunitaires. L'hémophilie A par exemple, est associée à une réponse immunitaire dépendante d'un défaut génétique sur le gène du facteur VIII[6]. Ce défaut génétique se traduit par l'absence de tolérance immunitaire et déclenche ainsi une réponse de type vaccinale lorsque le patient est traité par le facteur VIII [6]. Il pourrait ainsi être intéressant d'appliquer le même raisonnement pour l'exploration de la sécurité d'utilisation des biosimilaires. Des décisions scientifiques seraient nécessaires pour sélectionner les biomédicaments évalués pour des pathologies précises. Il pourrait être envisagé d'avoir une approche sélective sur une pathologie pour plusieurs gènes. A l'inverse, il pourrait être possible de rechercher la relation entre un gène précis pour plusieurs pathologies. Cette dernière approche pourrait être plus facile à implémenter pour des raisons techniques. Les données génétiques étant plus sensibles que les autres données de santé, il pourrait dans un premier temps être délicat d'intégrer un seul marqueur génétique dans l'analyse.

3. Les données prospectives et l'anticipation de survenue d'un évènement

Une analyse prospective peut être envisagée pour obtenir des données spécifiques comme présentées ultérieurement. Mais ces analyses peuvent également trouver leur place lors d'une exploration en temps réel de la sécurité d'utilisation. En d'autres termes, des données sont collectées au cours de la prise en charge dans le cadre d'un registre et sont analysées avant l'obtention du résultat clinique. L'objectif n'est donc plus le même puisqu'il n'y aurait plus de volonté d'identifier des tendances dans les données mais d'anticiper la survenue d'un résultat clinique. Il serait dès lors pertinent d'utiliser un modèle de réseaux neurones ayant déjà servie à des analyses rétrospectives pour mettre à profit le raisonnement qui a été élaboré. La classification adaptée aux biosimilaires seraient ainsi mise à profit.

Sous l'angle technique, le réseau de neurones a développé ses propres instructions pour traiter les données. Pour cela, les données de prises en charges associées au résultat cliniques ont ainsi permis au modèle d'acquérir un raisonnement. Ce raisonnement permet d'associer correctement les données d'entrées aux résultats cliniques et ainsi de mettre en évidence des corrélations dans les données. Toutefois, ce raisonnement peut être utilisé sur des nouvelles

données sans résultat associé afin d'en proposer un. Il est attendu que le résultat proposé ait une bonne probabilité d'être correct puisque le modèle a été utilisé sur un volume de données important. Cette analyse permettrait ainsi de classer les patients selon qu'ils sont à faible ou fort risque de réaction immunitaire au regard d'un ensemble de données. De la même façon que pour les études rétrospectives, le modèle serait capable de mettre en évidence l'importance des variables. En d'autres termes, on s'attend à ce que le modèle puisse dire quel patient est plus à risque et pour quelles raisons. En l'occurrence, le modèle pourrait faire ressortir les facteurs de risques significatifs déjà identifiés précédemment.

Pour arriver à cette anticipation de résultat clinique, le modèle doit réaliser des prédictions. Ces prédictions peuvent être obtenues sous forme de probabilité d'expérimenter un événement indésirable. Pour obtenir une prédiction, le modèle doit pouvoir exploiter les données liées à la prise en charges. Ces données de prise en charge sont donc converties en évènements [29] tels que la prescription, l'administration d'un produit ou un diagnostic. Ces évènements peuvent ensuite être associés à des signaux qui pourront être exploités par le modèle. En effet, chaque signal peut être un élément prédictif de la survenue d'un événement[21]. Toutefois, la prédiction fournie par le modèle va varier au cours du temps selon les signaux [21]. En effet, ces derniers vont augmenter au fur et à mesure de la prise en charge, de l'admission à la sortie de l'hôpital [21].

4. *L'anticipation dans la pratique : la valorisation des résultats fournis par le modèle*

Ces prédictions seraient intéressantes pour la prise de décision médicale. Deux approches pourraient être formulées pour valoriser les résultats fournis par un modèle capable de proposer une probabilité de résultat clinique. La première approche serait celle de l'éligibilité du patient à une pratique en cas de risque faible. Les facteurs de risques identifiés auparavant seraient recherchés chez le patient. Ainsi, en l'absence de facteurs de risques, le patient traité par un biomédicament de référence pourrait être éligible au *switch*. La seconde approche serait celle de la notification de l'augmentation du risque généré par cette pratique au cours du temps (figure 15). Il y aurait néanmoins des dimensions de responsabilité médicale à aborder qui ne seront pas développée dans ce document.

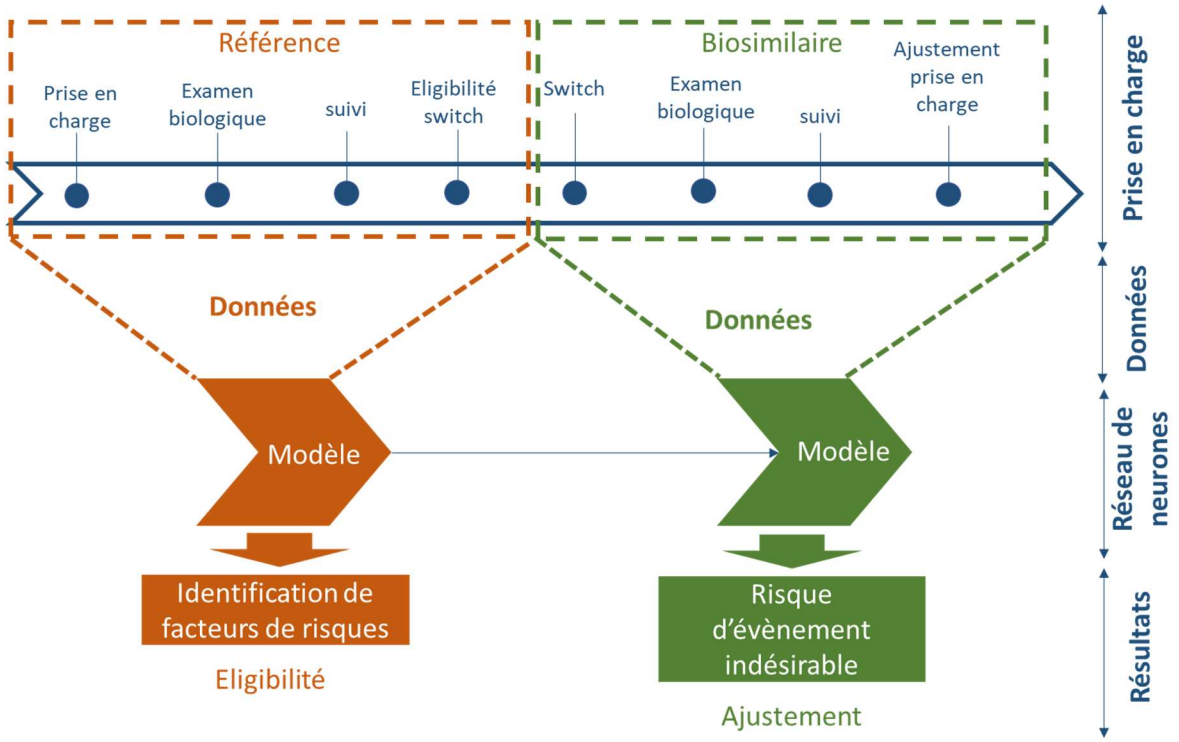

Figure 15: Eligibilité et anticipation

Une première utilisation d'un modèle d'apprentissage pourrait être envisagée pour engager une pratique de *switch* lorsque le risque est faible. Mais à l'inverse, le réseau de neurones pourrait alerter le praticien lorsque le risque immunitaire est fort. En effet, la prise en charge d'un patient intègre le biomédicament prescrit ainsi qu'une série d'événements. Certains évènements pourraient concourir à l'augmentation d'un risque de rupture de tolérance qui serait détecté par le modèle. Il pourrait donc être imaginable d'avoir un système de notification informant le praticien du changement de profil de sécurité. Il s'agirait ainsi d'une opportunité pour le praticien d'adapter la prise en charge, de réaliser une nouvelle prescription, dans le but de réduire le risque de survenue de réaction immunitaire. Le praticien devrait toutefois conserver une totale liberté d'analyse et d'esprit critique vis-à-vis de ces alertes puisque seule la vision holistique au lit du malade permet *in fine* de prendre une décision[32].

iv. De la corrélation vers la causalité : comprendre les mécanismes de l'immunogénicité

1. *Les limites de l'identification de corrélation*

Une autre approche pourrait conduire à utiliser l'apprentissage machine autrement. Il a été évoqué une utilisation pour observer les tendances dans les données et pour anticiper la survenue d'un évènement. Cependant, l'exploration des corrélations des données pourrait présenter des limites. En effet, cette approche ne permet pas de comprendre la cause d'une réaction immunitaire. La corrélation permettra notamment de montrer qu'un *switch* par le biosimilaire est plus à risque lorsque la référence est prescrite depuis longtemps, sans pour autant l'expliquer. En effet, des mécanismes biologiques complexes pourrait être impliqués dans la rupture de tolérance. Il pourrait donc y avoir un besoin de compréhension de la cause et des mécanismes impliqués afin d'agir au mieux sur la cause. Il s'agirait donc d'une prochaine étape d'exploitation des données de l'intelligence artificielle (figure 16).

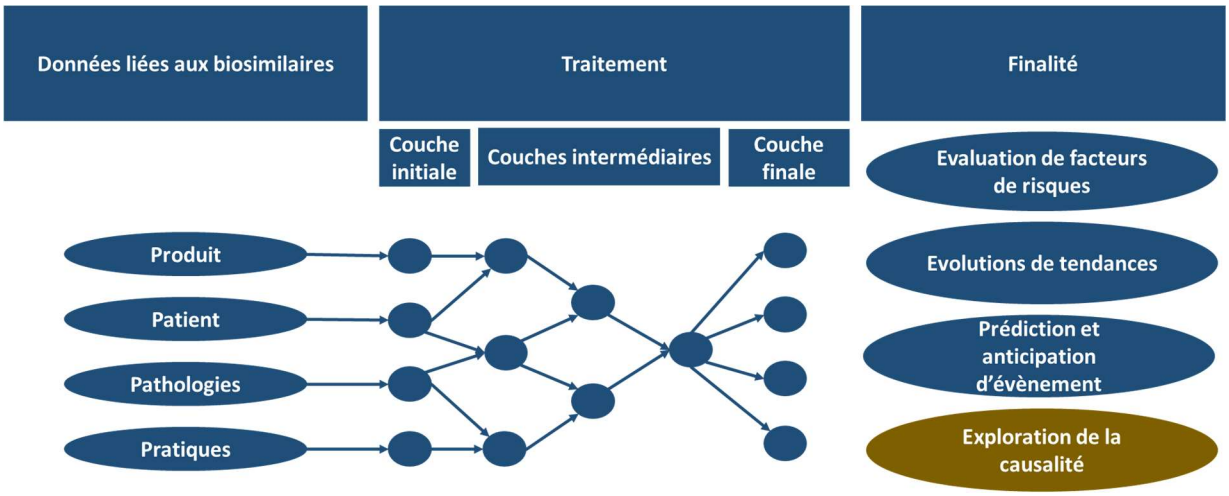

Figure 16: Exploration de la causalité

2. *La compréhension de la causalité : des mécanismes associés à des données massives*

Dans le cas des biosimilaires, l'exploration de la causalité pourrait être envisagée afin de comprendre l'articulation des différentes composantes de l'immunogénicité. En d'autres termes, il serait question de rechercher les mécanismes impliqués dans la survenue d'une réaction immunitaire qui dépend de nombreuses composantes. La compréhension de ces mécanismes serait également associée à des données massives que les méthodes classiques

auraient des difficultés à traiter. Or ce sont les méthodes classiques qui sont utilisées pour fournir des informations relatives aux mécanismes. La littérature souligne dès lors la possibilité d'exploiter des données massives en incluant des méthodes classiques afin d'explorer la causalité [32].

3. *Le couplage des méthodes classiques mécanistiques avec l'intelligence artificielle*

La littérature met en évidence la possibilité d'un couplage des méthodes d'apprentissage machine et des méthodes classiques dites mécanistiques. Ces dernières sont en fait de modèles mathématiques simplifiés de formules de mécanismes de causalité[32]. Ces modèles sont ainsi utilisés pour comprendre les mécanismes des pathologies [32] en étant concentré sur lien de causalité entre des données et un résultat [32]. Ainsi, là où l'apprentissage machine pourrait proposer un traitement au regard d'une situation, le modèle mécanistiques serait utilisé pour proposer un nouveau traitement. Or ces modèles mécanistiques sont limités lorsque les données sont trop volumineuses. L'intérêt serait alors de mettre à disposition la capacité d'une IA à traiter des données massives hétérogènes pour dépasser la limite des modèles mécanistiques[30]. Dans la pratique, les méthodes classiques seraient utilisées durant la phase d'apprentissage des réseaux de neurones afin que ce dernier puisse profiter de la capacité d'extrapolation des modèles mécanistiques. Ces réseaux de neurones pourraient ensuite appliquer le raisonnement acquis de cette extrapolation sur un grand volume de données [32].

c) Une intégration précoce des spécificités de l'intelligence artificielle

i. Les limites d'ordre juridique : l'exemple d'une finalité interdite

1. Les restrictions des analyses de sécurité

Des objectifs de sécurité pourraient justifier différentes finalités de traitement des données, telles que l'identification de facteurs de risques. Un grand volume de données serait nécessaire à ce traitement pour répondre aux objectifs de sécurité. Il serait toutefois possible pour un industriel d'avoir d'autres objectifs qui justifieraient des analyses ne relevant pas de la sécurité d'utilisation du produit. Il pourrait par exemple exister une volonté d'exploiter ces données dans un but commercial afin d'étudier les ventes. Cependant le traitement de ces données par une intelligence artificielle n'est pas exempté de restrictions. Certains de ces traitements pourraient notamment relever de finalités interdites. Ce risque serait d'autant plus important qu'un modèle d'apprentissage machine pourrait apporter davantage d'information qu'une méthode classique à travers le traitement de données massives.

2. L'interdiction de collecte de données de prescription

Un cas concret de finalité interdite peut être donné. L'article L.4117-7 du code de la santé publique (CSP) interdit l'obtention de données de prescription des praticiens. Il n'est donc pas possible pour un laboratoire de connaitre le nombre de biomédicament prescrit par un praticien. Ainsi, un usage d'un réseau de neurone, conforme à cet article, devrait démontrer qu'aucune donnée de prescription individuelle n'est obtenue. En d'autres termes, le réseau de neurones ne devrait pas permettre de retrouver le nombre de biomédicament prescrit par un praticien dans une démarche d'identification de facteurs de risque. Ce risque pourrait notamment être maitrisé *via* une démarche d'agrégation des données de prescriptions des praticiens par centre. Ainsi, lors d'une collecte de données dans le cas d'un registre, il ne serait possible d'associer le biomédicament prescrit qu'au un centre et non au praticien.

ii. Garantir l'anonymisation des données

1. Une fragilisation de l'anonymisation des données

Les lignes directrices du G29 mettent en avant qu'un traitement est anonyme tant qu'il n'est pas possible d'individualiser ou de corréler les données[34]. Or, l'utilisation d'un modèle d'apprentissage pourrait fragiliser cette anonymisation par l'exploitation de données

massives. Il ne s'agit pas d'une remarque propre aux biosimilaires mais bien à tous les produits de santé. Cependant ce risque augmenterait pour les biosimilaires avec le nombre de composantes intégrées dans l'étude. Nous pourrions supposer que plus le nombre de composante augmente, plus il y a de types de données différents et ainsi un risque de réidentifier un patient. Cette réidentification supprimerait dès lors le caractère anonyme des données, fragilisant ainsi l'absence d'inférence[23]. L'avis de la CNIL du 11 juillet 2019 sur un projet de loi relatif à la bioéthique mentionne également que l'inclusion de données génétiques pourrait accroitre ce risque. L'intérêt d'un modèle d'apprentissage machine intégrant des données génétiques devraient donc être nuancé au regard de cette jurisprudence.

iii. L'opacité des résultats obtenus : le concept de boite noire

1. La fiabilité du modèle

Le réseau de neurones est une méthode qui fournit des résultats. Comme toute méthode il peut être discuté de la qualité et de la fiabilité des résultats. En d'autres termes, il serait nécessaire de s'assurer que le résultat est correct. Or la difficulté qui se pose avec l'intelligence artificielle est celle de la boite noire. En effet, le raisonnement et les opérations conduites par les couches cachées du réseau de neurones peuvent être opaques. Il pourrait ainsi manque de transparence sur les opérations à l'origine de l'obtention d'un résultat. Or ce manque de transparence pourrait se traduire par un manque de confiance des professionnels de santé dans le modèle. Ce manque de transparence pourrait également entrer en conflit avec la compréhension humaine des prédictions fournies par le réseau de neurones[29]. Il y aurait donc une limite à la trop grande performance d'un modèle qui impacterait l'interprétation des résultats obtenus [29].

2. L'impact du manque de transparence pour les biosimilaires

La littérature met en évidence que les réseaux de neurones n'offrent pas toujours des résultats interprétables[30]. Cela signifie qu'il n'y a pas un aperçu suffisant des facteurs qui ont influencé l'obtention d'un résultat[21], impactant ainsi la confiance des cliniciens [21]. Or dans le cas des biosimilaires, ce sont bel et bien ces facteurs qui intéressent l'investigateur. Le modèle peut être utilisé pour une recherche de corrélation ou pour l'anticipation d'un évènement. Dans les deux cas, les résultats ont besoin d'être appuyés sur les facteurs de risques afin d'être

informatifs. En effet, un praticien ne saurait exploiter un simple résultat de risque de réaction immunitaire sans en connaitre les potentiels sources.

3. Des mécanismes d'attribution pour augmenter la transparence

Cette opacité ne devrait pas constituer un frein absolu à l'utilisation de l'intelligence artificielle. La littérature met en avant des solutions à cette barrière appelées mécanismes d'attribution. Ces derniers servent à mettre en évidence, pour chaque patient, les éléments qui ont influencé la prédiction. Ce sont ces mécanismes d'attribution qui visent à identifier les signaux mais également l'importance de ce signal dans la prédiction [21]. Dans le cas des biosimilaires, il est souhaitable que ces mécanismes d'attribution concourent à l'identification des facteurs de risques. *In fine*, ces mécanismes permettraient un gain de confiance de la communauté de santé dans les résultats fournis par un modèle d'apprentissage machine.

Conclusion

L'utilisation de l'intelligence artificielle pour optimiser la sécurité d'utilisation des biosimilaires peut prendre différentes formes. Que ce soit *via* l'identification de facteurs de risques ou par l'anticipation de survenue d'événements, il s'agit d'une utilisation d'un modèle adapté aux spécificités des biosimilaires. L'intérêt de l'intelligence artificielle porte autant sur les fonctions du modèle que sur l'utilisation de données massives. En effet, ces dernières permettent d'élargir le nombre de composantes intégrées à l'analyse. Une étude de données de vie réelle mettant à profit un réseau de neurones pour traiter des données massives liées à la prise en charge de patient par un biosimilaire pourrait dès lors concourir à une optimisation de la sécurité d'utilisation des biosimilaires.

Cette intelligence artificielle s'inscrit dans un contexte particulier. Il y a notamment une augmentation du nombre de biosimilaires sur le marché, accompagnée d'incitations à les utiliser en ambulatoire comme à l'hôpital. Les analyses de sécurité peuvent également trouver un intérêt au regard de l'abrogation du droit de substitution prévue dans l'article 42 de la LFSS 2020. Cette abrogation est notamment prononcée pour des raisons de sécurité sanitaire. Bien que motivé par des raisons économiques, ce droit de substitution a été prévu précocement par la LFSS 2014 dans un contexte manquant d'information.

En effet, ce droit de substitution ne saurait être appréhendé sans informations complémentaires quant à l'évaluation du risque de cette pratique. Cette évaluation implique l'intégration de la traçabilité pour identifier correctement la relation entre un produit et un effet indésirable. Cela justifie dès lors une démarche anticipatoire d'exploration des facteurs de risques liés à l'utilisation des biosimilaires en soins courants. Cette démarche pourrait également se poursuivre par le développement d'un modèle de réseau de neurones garantissant un suivi adéquat de la sécurité d'utilisation du biosimilaire en cas d'un éventuel droit de substitution.

Cette approche centrée sur la sécurité sanitaire des biosimilaires soulève toutefois des questions supplémentaires. En effet, il pourrait être discuté de l'accès aux dossiers médicaux et pharmaceutiques, leur chainage, la qualité des données ou encore le respect de la protection des données personnelles. Sur ce dernier point, il pourrait être mentionné que le

recueil du consentement serait indépendant de celui obtenu lors de la création des dossiers du patient. Puisqu'un traitement serait effectué sur ces données, avec une finalité spécifique, un nouveau consentement serait nécessaire.

Il y a d'autres niveaux de réflexion à intégrer afin d'évaluer la faisabilité d'une telle étude. Les niveaux peuvent se décliner selon les données, la méthode, les études et les résultats.

L'accès aux données est primordial pour conduire une telle analyse. Les données massives liées à la prise en charge d'un patient sont la matière première qui après traitement permet d'obtenir de l'information. Or, il pourrait être compliqué d'avoir accès à l'intégralité des données jugées pertinentes pour l'étude. En effet, certaines données pourraient simplement ne pas être disponibles, ne pas être présentes dans la base. De plus, les analyses conjointes de bases pourraient être limitées, ne permettant pas de croiser les données de natures différentes.

Concernant le modèle d'intelligence, il aurait été pertinent d'aborder le cadre réglementaire applicable. D'autant plus que ce cadre pourrait varier selon l'utilisation du modèle. En effet, l'exploration de corrélation n'aurait pas le même impact que l'anticipation de survenue de réactions immunitaires. En effet, l'anticipation pourrait conduire à une prise de décision médicale et donc avoir un impact sur la prise en charge. Cette anticipation impliquerait ainsi une responsabilité médicale différente suite à l'utilisation d'un modèle d'intelligence artificielle.

L'encadrement d'une étude prospective utilisée à des fins d'anticipation d'événement devrait également faire l'objet d'une discussion. En effet, sont exclues de la loi Jardé les études rétrospectives car elles n'impliquent pas la personne humaine. Cependant, une étude conçue de façon à anticiper la survenue d'un événement pour prendre une décision médicale impliquerait sans aucun doute la personne humaine. Cette étude serait dès lors encadrée par la loi Jardé.

Des réflexions éthiques sur l'utilisation de l'intelligence artificielle devraient également être intégrées dans la conception d'une étude. En ce sens, l'avis 130 du rapport du 29 mai 2019 du Comité consultatif National d'Ethique permet de soulever plusieurs pistes de réflexion. Ces réflexions pourraient être réparties en deux axes, l'un en amont d'une analyse puis l'autre à

la suite de l'obtention des résultats. Cette dissociation des deux axes permet dès lors de tenir compte du caractère imprévisible de certaines méthodes d'intelligence artificielle.

Le premier axe de réflexion éthique serait celui de la garantie humaine du processus d'analyse tel que mentionné par le CCNE dans l'avis 130 du rapport du 29 mai 2019. Cette garantie humaine recherche l'adéquation des données et des traitements avec la finalité de l'analyse. Dès lors, l'initiateur de l'étude sera amené à trouver un équilibre entre une sous-exploitation des données et un partage excessif. De plus, il sera amené à démontrer que le traitement de données de vie réelle ne fragilise pas la protection des individus.

Cette garantie humaine permettra également de formuler des résultats attendus en adéquation avec les finalités de l'étude. Le recueil du consentement du patient se ferait ainsi en accord avec cette finalité. Il serait ainsi possible d'identifier précocement des finalités incertaines qui mettrait en danger ce consentement. De la même façon, tous traitements dangereux comportant des finalités interdites pourront être écartés de l'analyse.

Le second axe de réflexion éthique pourrait être mené une fois l'obtention de résultats. En effet, il s'agirait d'une réflexion ultérieure visant à encadrer la portée et l'impact des résultats obtenus par des méthodes d'intelligence artificielle. Certes, une analyse de données de vie réelle de l'utilisation des biosimilaires en soins courants porte sur l'évaluation d'un profil de sécurité. Toutefois, cette première finalité pourrait être la cible de dérive. En effet, la recherche de facteurs de risques ne doit pas conduire à un profilage des patients. Cette pratique mènerait ainsi à une perte de la mutualisation du risque, principe sur lequel est fondé notre système de sécurité sociale. Il est donc primordial que tous les acteurs d'une analyse, mobilisant une intelligence artificielle, garantissent le respect des principes d'équité et de solidarité de notre système de santé.

Bibliographie

(1) Ginghină O, Traian Alexandru Burcea-Dragomiroiu G, Gălățeanu B, et al. Long-term safety of biosimilar medicinal products – key for administration? Farmacia. 2019 Jan 3;67(1):18–26.

(2) Bocquet F, Paubel P. Médicaments biosimilaires : quel cadre juridique pour quel modèle économique ? JDSAM. 2015 Mai;10(2):8-22.

(3) Bocquet F, Paubel P. A long war begins: biosimilars versus patented biologics. J Med Econ. 2015 Jul 27;18(12):1071–1073.

(4) Kowalski SC, Benavides JA, Roa PAB, et al. Panlar consensus statement on biosimilars. Clin Rheumatol. 2019 Mar 27;38(1):1485–1496.

(5) Chang LC. The biosimilar pathway in the USA: An analysis of the innovator company and biosimilar company perspectives and beyond. J Food Drug Anal. 2019 Mar 29;27(3):671–678.

(6) Doevendans E, Schellekens H. Immunogenicity of Innovative and Biosimilar Monoclonal Antibodies. MDPI. 2019 Mar 5;8(1):1-10.

(7) Steinwandter V, Borchert D, Herwig C. Data science tools and applications on the way to Pharma 4.0. Drug Discovery Today. 2019 Sep;24(9): 1-11 (2019):1795-1805.

(8) Rocco P, Selletti S, Minghetti P. Biosimilar switching and related medical liability. J Forensic Leg Med. 2018 Feb 17;55(1): 93–94.

(9) Megerlin F, Lopert R. Substitution et interchangeabilité des biomédicaments, Prospective d'impact compétitif en droit comparé franco-américain. Tech Hosp. 2014 Dec;748:37-43.

(10) Smeeding J, Malone DC, Ramchandani M, et al. Biosimilars: Considerations for Payers. P T. 2019 Feb;44(2):54-63.

(11) Agence Européenne du Médicament. Les médicaments biosimilaires dans l'UE [internet]. 2017 [cité 20/12/2019]. Disponible sur https://www.ema.europa.eu/en/documents/leaflet/biosimilars-eu-information-guide-healthcare-professionals_fr.pdf

(12) Makady A, De Boer A, Hillege H, et al. What is real world data ? A review of definition based on litterature and stakeholders interviews. Value Health. 2017 Aug;20(7):858-865.

(13) De La Forest Divonne M, Gottenberg JE, Salliot C. Revue systématique des registres de polyarthrites rhumatoïdes sous biothérapie dans le monde et méta-analyse sur les données de tolérance. Rev Rhum. 2017 May;84(2):199–207.

(14) Berger ML, Sox H, Willke RJ, et al. Good Practices for Real-World Data Studies of Treatment and/or Comparative Effectiveness: Recommendations from the Joint ISPOR ISPE Special Task Force on Real World Evidence in Health Care Decision Making. Value Health. 2017 Sept;20(8):1003-1008.

(15) Chapman SR, Fitzpatrick RW, Aladul MI. Knowledge, attitude and practice of healthcare professionals towards infliximab and insulin glargine biosimilars: result of a UK web-based survey. BMJ Open. 2017 Jun 21;7(6):1-8.

(16) Frantzen L, Cohen JD, Tropé S, et al. Patients' information and perspectives on biosimilars in rheumatology: A French nation-wide survey. Joint Bone Spine. 2019 Jul;86(4):491–496.

(17) APM news. Prescription hospitalière de biosimilaires délivrés en ville : extension de l'expérimentation à l'adalimumab (projet d'arrêté) [Internet]. 2019 [20/12/2019. Disponible

sur https://www.apmnews.com/depeche/132035/331670/prescription-hospitaliere-de-biosimilaires-delivres-en-ville-extension-de-l-experimentation-a-l-adalimumab--projet-d-arrete

(18) Scherlinger M, Langlois E, Germain V, et al. Acceptance rate and sociological factors involved in the switch from originator to biosimilar etanercept (SB4). Semin Arthritis Rheum. 2019 Apr;48(5):927–932.

(19) Bocquet F. Entre difficultés de statuer sur la question de substitution et les incitations financières à leur prescription : quelle régulation optimale pour les médicaments biosimilaires ? RGDM. 2020 Jan;28(7):177-194.

(20) Wang SV, Schneeweiss S, Berger ML, et al. Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. Value Health. 2017 Sep;20(8):1009–1022.

(21) Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning for electronic health records. npj Digital Med. 2018 May 08;1(18):1-26.

(22) Crown WH. Real-World Evidence, Causal Inference, and Machine Learning. Value Health. 2019 May;22(5):587–592.

(23) Berdaï D, Thomas-Delecourt F, Szwarcensztein K, et al. Demandes d'études post-inscription (EPI), suivi des patients en vie réelle : évolution de la place des bases de données. Therapies. 2018 Feb;73(1):13-24.

(24) Seymour K, Benyahia N, Hérent P, et al. Exploitation des données pour la recherche et l'intelligence artificielle : enjeux médicaux, éthiques, juridiques, techniques. Imagerie de la Femme. 2019 Jun;29(2):62–71.

(25) République française. LOI n° 2012-300 du 5 mars 2012 relative aux recherches impliquant la personne humaine [Internet]. Légifrance du 5 mars 2012 [cité le 7 décembre 2010]. Disponible sur https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000025441587&categorieLien=id

(26) Agence Nationale de Sécurité du Médicament. Bonnes pratiques de pharmacovigilance [internet]. 2018 [cité 20/12/2019]. Disponible sur https://www.ansm.sante.fr/S-informer/Points-d-information-Points-d-information/Actualisation-des-Bonnes-pratiques-de-pharmacovigilance-Point-d-Information

(27) Girard M, Polton, D. La nouvelle réglementation sur l'accès aux données de santé. Colloque «Big data en santé : du discours aux applications pratiques» [Internet]. 2018 [cité 20/12/2019]. Disponible sur http://www.institutdroitsante.fr

(28) Bertail P, Bounie D, Clémençon S, et al. Algorithmes : biais, discrimination et équité [internet]. 2019 [cité 20/12/2019]. Disponible sur https://www.telecom-paris.fr/algorithmes-biais-discrimination-et-equite

(29) Doupe P, Faghmous, J, Basu S. Machine Learning for Health Services Researchers. Value Health. 2019 Jul;22(7):808–815.

(30) Hutchinson L, Steiert B, Soubret A, et al. Models and Machines: How Deep Learning Will Take Clinical Pharmacology to the Next Level. CPT Pharmacometrics Syst Pharmacol. 2018 Dec 14;8(3):131–134.

(31) Toulon N. Deep Learning et Agriculture [internet]. 2019 [cité 20/12/2019]. Disponible sur https://www.agrotic.org/les-actualites/deep-learning-et-agriculture-une-etude-de-la-chaire-agrotic/

(32) Baker RE, Pena JM, Jayamohan J, et al. Mechanistic models versus machine learning, a fight worth fighting for the biological community ? Biol Lett. 2018 May 1;14(5):1-4.

(33) Tabuteau D. 2. La sécurité sanitaire. Paris, France : Berger-Levrault ; 2002. 390p.

(34) République française. Jurisprudence n°2019-097 du 11 juillet 2019. [Internet] Délibération n°2019-097 du 11 juillet 2019 portant avis sur un projet de loi relatif à la bioéthique. Légifrance du 11 juillet 2019 [cité le 7 décembre 2010] Disponible sur https://www.legifrance.gouv.fr/affichCnil.do?id=CNILTEXT000038848207

Vu, le Président du jury,

Stéphane BIRKLE

Vu, le Directeur de thèse,

François BOCQUET

Vu, le directeur de l'UFR,

UNIVERSITÉ DE NANTES **Année de la soutenance** 2020

Nom - Prénoms : PERPOIL – Antoine, Nicolas, Rémi
Titre de la thèse : Dans quelle mesure l'intelligence artificielle pourrait-elle permettre d'optimiser la sécurité d'utilisation des biosimilaires ?

Résumé de la thèse : L'utilisation des biosimilaires, copies de médicaments d'origine biologique dont le brevet est arrivé à expiration, interroge parfois les pouvoirs publics et les professionnels de santé : immunogénicité, switch, substitution, traitement en initiation ou en suite de traitement, etc. Ces questions pourraient trouver des réponses dans l'utilisation de l'intelligence artificielle (IA). Dans ce travail, il sera donc discuté de l'intérêt de l'IA à des fins de suivi de l'utilisation des biosimilaires en soins courants.

MOTS CLÉS ***: Biosimilaires, Immunogénicité, Switch, Substitution, Intelligence Artificielle***

JURY

PRÉSIDENT : **M Stéphane BIRKLE**, Professeur Universitaire Hématologie Immunologie
Faculté de Pharmacie de Nantes

ASSESSEURS : **M François BOCQUET**, Maître de Conférences Universitaire – Praticien Hospitalier
Faculté de Pharmacie Paris Descartes

M Gael GRIMANDI, Doyen de la Faculté de Pharmacie de Nantes – Professeur Universitaire - Praticien Hospitalier
Faculté de Pharmacie de Nantes

Mme Anne CHIFFOLEAU, Praticien Hospitalier, Unité fonctionnelle de Pharmacovigilance
Direction de la recherche, CHU de Nantes

M Jean-François SIMONET, Directeur Compliance et Délégué à la Protection des Données Personnelles
Amgen France

Adresse de l'auteur : ***18 rue de la bourgeonnière, 44300 Nantes***