



kynapse
an open company

Les plateformes de données de santé

Ou comment valoriser les données
de santé de vie réelle ?



kynapse
an **open** company

Sommaire

Introduction 4

1 Chapitre 1
Les plateformes de données de vie réelle en Santé 6

2 Chapitre 2
Illustration des cas d'usage identifiés 10

3 Chapitre 3
Conclusion 23

Le mot de la fin de l'équipe de Kynapse 25



kynapse
an **open** company



kynapse
an open company

Introduction

Les plateformes d'agrégation de données s'inscrivent comme de véritables accélérateurs de valorisation des données de santé de vie réelle pour améliorer les parcours de soins et la santé des patients.

L'épidémie de Covid 19 a mis en évidence combien les systèmes de santé sont fragiles.

Dans ce cadre, la transformation digitale présente une opportunité clé pour leur consolidation et pérennisation. Un des nombreux aspects de cette transformation est l'utilisation et la valorisation des données.

Ces dernières, comme leur racine latine l'indique (« data » signifiant « offertes ») sont souvent gratuitement accessibles, et leur utilisation représente un fort potentiel de création de valeur clinique et opérationnelle pour les organisations ayant compris comment et à quelle fin les exploiter.

Parmi ces données de santé, cet article se concentre particulièrement sur les données de vie réelle (données générées au cours des soins de routine d'un patient, sans intervention sur les modalités habituelles et en dehors du cadre expérimental) et la manière dont ces dernières commencent aujourd'hui à être agrégées et liées dans des plateformes qui facilitent leur accès, leur exploitation et leur valorisation par les acteurs de santé publics et privés.

La France fait figure de pionnière sur ce sujet grâce au Health Data Hub (HDH), l'une des premières plateformes de données de santé. Elle s'inscrit ainsi dans un domaine clé pour l'avenir, où des acteurs privés, tels qu'IQVIA et IBM MarketScan, et des acteurs publics bâtissent peu à peu l'avenir de la médecine et de la recherche clinique.

Kynapse, cabinet de conseil en transformation digitale et data, participe activement à la conception et à la mise en œuvre de bases de données ou plateformes de données de vie réelle. Cet article en tire un certain nombre d'illustrations.

Le propos reste néanmoins généralisable à d'autres plateformes similaires, publiques et privées, qui émergent à l'échelle mondiale.



kynapse
an **open** company

1 . Chapitre 1

Les plateformes de données de vie réelle en Santé

Les données de vie réelle

(Real World Data en anglais) révolutionnent la recherche, le développement et l'accès des patients à de meilleurs traitements en générant des preuves issues du terrain - (Real World Evidence).

Les données de vie réelle sont le reflet de la pratique courante des médecins et des spécialistes de santé : elles sont générées à l'occasion de soins de routine dispensés à un patient lors d'une consultation médicale ou d'une prescription thérapeutique par exemple.

De ce fait, elles sont collectées sans interférence sur les modalités usuelles de prise en charge des malades et en dehors d'un cadre expérimental comme dans les essais cliniques.

Les sources de données de vie réelle sont multiples. Elles peuvent provenir de systèmes d'informations ou de bases de données de santé issues d'établissements de soins, d'industriels ou d'organismes publics.

Il est également possible de collecter des données dites "secondaires", c'est à dire destinées au moment de leur collecte à d'autres utilisations, à partir d'internet, de dispositifs connectés, d'équipements de diagnostic, de réseaux sociaux, etc.

Les sources et quantité de données de santé primaires et secondaires ne vont cesser de croître ces prochaines années, stimulées notamment par la transformation digitale de nos sociétés et de nos systèmes de santé.

La valorisation de ces données de vie réelle passe par leur collecte, leur stockage sécurisé, leur traitement et leur analyse. De surcroît, la conjonction de la migration des données vers le cloud permettant leur stockage et leur analyse à grande échelle, de même que le développement des algorithmes d'intelligence artificielle facilitent leur accès et leur exploitation par des outils de modélisation et d'interprétation.

Par exemple en 2020, lors de la pandémie de Covid-19 sont observés chez de nombreux patients des effets persistants de la maladie plusieurs mois après l'infection.

Afin de mieux comprendre ce sujet, Kynapse a exploité la base de données THIN® (The Health Improvement Network) du groupe Cegedim (groupe Français spécialisé dans la gestion des flux numériques de l'écosystème santé) pour identifier les patients concernés et donner une définition de cette pathologie nouvelle.

Des données de remboursement de soin couplées aux dossiers médicaux issus de la médecine de ville ont alors permis de mettre au point un outil d'aide au diagnostic.

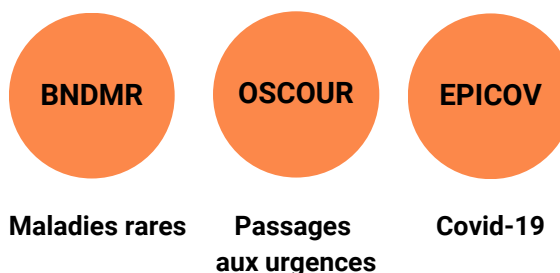
La création de plateformes de données de santé à l'instar du Health Data Hub, en France, franchit une étape supplémentaire dans la facilitation de l'accès et de l'exploitation des données de santé. Le HDH rassemble en effet des bases de données de sources différentes et complémentaires, en les chaînant et en mettant à disposition des outils dédiés à leur exploitation.

Ce faisant, ces plateformes de données de santé donnent accès à des informations permettant de reconstituer tout ou partie des parcours de santé et de soins des patients et de synthétiser une vue longitudinale inaccessible au premier abord.

Chaque acteur (qu'il s'agisse d'un CHU, un laboratoire pharmaceutique ou un fabricant de dispositifs médicaux) peut ensuite agréger les données dont il est propriétaire à celles issues de ce type de plateforme. Cela lui permet par exemple d'évaluer l'impact de ses propres solutions (diagnostiques ou thérapeutiques) en les comparant à d'autres solutions existantes sur le marché ou bien à l'absence de solutions comparables.

Ainsi le Health Data Hub s'est positionné depuis sa création en 2019 comme un guichet unique en France pour garantir un accès facilité et unifié, transparent et sécurisé, aux données de santé dans le but de favoriser la recherche et de promouvoir l'innovation. Il met à disposition des bases de données maintenues en qualité dans un cadre réglementaire assurant un haut niveau de protection des données, ainsi que des outils facilitant l'accès, le traitement et l'analyse de ces données.

A l'origine, le Health Data Hub s'est construit autour du SNDS (Système National des Données de Santé), système couvrant l'ensemble des données de santé associées à un remboursement de l'assurance maladie, qu'elles soient collectées lors d'un traitement hospitalier, d'une visite chez le médecin, de la participation à une cohorte de recherche ou à un registre épidémiologique. Aujourd'hui, le catalogue du Health Data Hub regroupe une dizaine de bases de données de natures variées témoignant du potentiel de diversité des sujets couverts :



Le Health Data Hub offre aux institutionnels, établissements de santé ou de recherche et industriels de santé l'opportunité de valoriser ces données pour leurs projets de recherche, à condition que ces projets respectent un certain nombre de règles d'éligibilité dont notamment le fait de présenter un intérêt de santé publique.

Le Health Data Hub étant encore relativement récent, une phase d'apprentissage est nécessaire pour que les acteurs souhaitant accéder à la plateforme s'adaptent au cadre réglementaire garanti par le Health Data Hub ou pour accéder à des données spécifiques pour leur projet de recherche.

Il faut aussi noter que le cadre réglementaire, de même que l'offre de services, sont en constante évolution pour continuer à alimenter le catalogue de données et soutenir l'innovation.

A travers les exemples suivants, cet article s'attache à proposer un aperçu des projets actuellement réalisés en collaboration avec des plateformes de données de santé "en vie réelle" dont une grande partie avec le Health Data Hub.

07 cas d'usage y sont détaillés, correspondant à cinq thématiques essentielles conformes aux projets de Kynapse réalisés auprès de ses clients et du répertoire des projets validés par le Health Data Hub.

Ces exemples ont en particulier vocation à inspirer les industriels de santé (groupes pharmaceutiques, sociétés du dispositif médical, acteurs de la e-santé et des thérapies digitales, ...) pour les inciter à initier ou continuer à développer leurs propres projets de recherche fondés sur les données de vie réelle et qu'ils puissent à terme faire partie de leur pratique courante.

2. Chapitre 2

Illustration des cas d'usage identifiés

Le répertoire public des projets déposés au HDH par les différents acteurs de santé démontre une grande variété d'utilisation aussi bien pour les professionnels de la santé que pour les industriels ou les institutionnels.

Ces projets peuvent viser à mieux connaître les pathologies et les parcours de soins, piloter une épidémie, mesurer les besoins de patients non pris en charge, prédire la survenue de complications ou de ré-hospitalisations, mieux observer l'efficacité d'un traitement et les réponses des patients en vie "courante" ou encore développer des outils diagnostics.

La liste des applications possibles est longue et continue chaque jour de s'allonger.

Les cas d'usages décrits ci-après sont répartis ainsi :

- Santé des populations
- Compréhension des maladies
- Aide au diagnostic
- Parcours patient
- Pharmacovigilance



Santé des populations (Population Health)

Simuler des scénarios de politiques publiques pour en mesurer l'impact sur l'état de santé d'une population.



La problématique de départ :

Le changement climatique menace des éléments clés et quotidiens en termes de santé de la population mondiale : air pur, eau potable, alimentation et abris par exemple.

Un acteur mutualiste de santé et de protection sociale a imaginé simuler l'état de santé d'une population par l'analyse de données de santé couplées à des données environnementales.

A terme, son objectif est de pouvoir simuler des scénarii de politique publique pour en déterminer l'impact avant même leur mise en application.



La solution imaginée :

Cet acteur mutualiste, accompagné par le Health Data Hub, a débuté sa démarche par l'identification des données d'intérêt pour son projet. Parmi elles, les données du SNDS ainsi que plusieurs bases Open Data relatives à la qualité de l'air, de l'eau et des territoires.



Les résultats :

Les travaux relatifs à ce projet sont actuellement en cours et devraient permettre aux acteurs publics d'éclairer les décisions relatives à leurs politiques publiques, mais aussi aux industriels de la santé d'accélérer le développement de leurs propres projets de recherche pour répondre aux défis à venir.



Compréhension des maladies

Mieux comprendre le Covid long, affection nouvelle dont la compréhension et la prise en charge restent mal comprises.



La problématique de départ :

Un centre de santé pour patients atteints de pathologies chroniques a souhaité apporter des réponses à la fois épidémiologiques, physiopathologiques et thérapeutiques au sujet du Covid long.

Cette pathologie apparue avec la pandémie de Covid-19 se manifeste par différents symptômes ressentis sur une période plus longue que la moyenne et reste encore mal comprise avec un diagnostic et une prise en charge complexes. En plus de l'apport de connaissances sur cette affection, le centre avait pour but d'améliorer la prise en charge des patients en identifiant notamment mieux les patients à risque.

Les coûts associés aux errances thérapeutiques ainsi qu'aux complications éventuelles pourront ainsi être mieux prévenus.



La solution imaginée :

Accompagné par le HDH, ce centre de santé a obtenu l'accès à un lac de données (data lake) dédié à la compréhension de la maladie et à l'analyse de son parcours qu'il a pu coupler avec les données issues de ses propres dossiers patients.

Le centre utilise en particulier des données issues de questionnaires patients renseignant l'évolution de leurs symptômes, des données médicales issues de dossiers patients de leur propre base de données ainsi que des données de suivi à domicile collectées par des objets médicaux connectés.



A terme, le projet permettra grâce à ces données de mettre en évidence des variables explicatives de la maladie, de comprendre les facteurs de risque associés et de prévenir de potentiels goulots d'étranglement tout au long du parcours de soin.

Enfin, ce projet se propose d'établir des modèles de prédiction permettant d'inférer sur l'évolution de la maladie et réduire ainsi l'errance thérapeutique ainsi que les risques de complication associés.



Les résultats :

En associant des modes de collecte de données variés (déclaratif, dossiers patients, objets connectés), le projet ambitionne des bénéfices substantiels qui permettront d'améliorer la connaissance d'une maladie nouvelle, d'en identifier les facteurs de risques et d'en prédire l'évolution.

Le bien-être patient et les coûts des parcours de soins en seront améliorés.



Diagnostiquer de manière précoce les neuropathies des petites fibres.



La problématique de départ :

Afin de se positionner en leader du marché du diagnostic neuro-oncologique, une entreprise de développement de dispositifs médicaux a voulu s'appuyer sur des solutions d'intelligence artificielle pour permettre un diagnostic précoce des neuropathies des petites fibres.

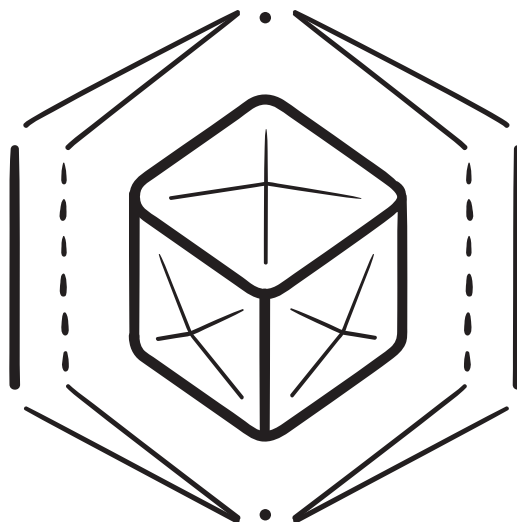
La neuropathie des petites fibres est une atteinte des très petites fibres nerveuses qui peut être responsable de douleurs cutanées, le plus souvent au niveau des membres inférieurs, et souvent inexplicables car la peau garde son apparence normale. Ces neuropathies représentent de surcroît un diagnostic long et fastidieux par un spécialiste car elles nécessitent un comptage manuel du nombre de fibres après une biopsie.

C'est pourquoi les errances diagnostiques et thérapeutiques restent élevées avec un impact important à la fois sur les patients, en augmentant le risque d'aggravation de la maladie, mais aussi sur l'assurance maladie en raison des surcoûts engendrés par la multiplication des examens et des visites de professionnels de santé non conclusifs.



La solution imaginée :

L'entreprise concernée avait pour objectif de développer un logiciel de diagnostic précoce, rapide, précis et moins coûteux, révolutionnant ainsi la prévention et le traitement des maladies neuropathiques.



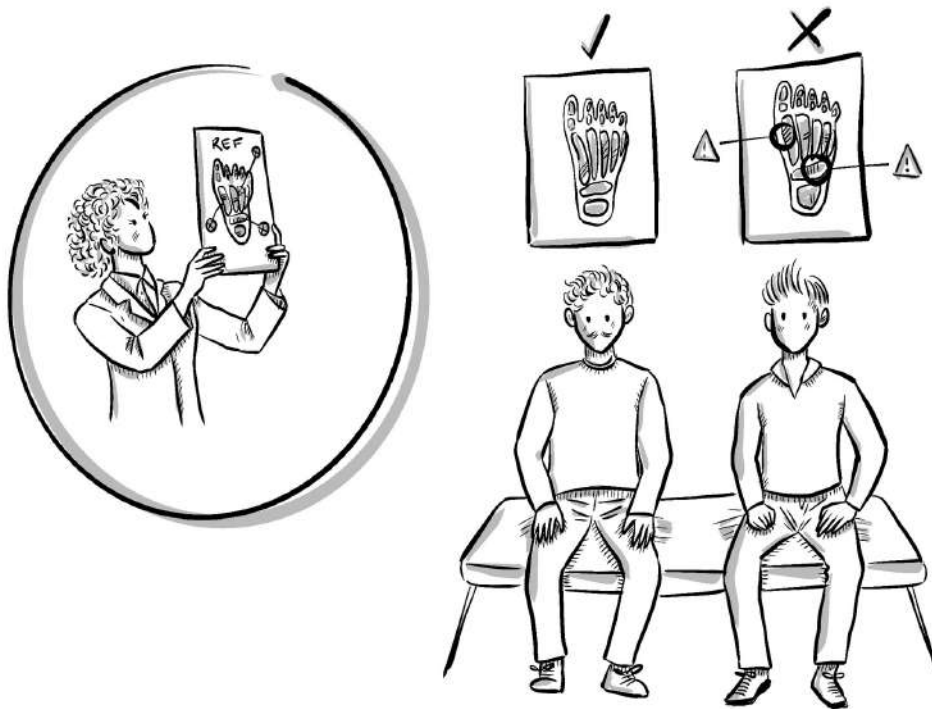
Grâce au HDH, l'entreprise a eu accès aux données d'une banque d'images de scanners 3D colorisées et labélisées après immunofluorescence de patients atteints de neuropathies. Elle a ensuite pu entraîner un modèle de vision par ordinateur. Ce dernier, entraîné sur les images de scanners 3D labélisées, peut alors détecter les petites fibres et procéder à leur comptage avec un taux de précision de plus de 80%.



Les résultats :

En intégrant son outil directement dans l'environnement médical du radiologue, cette solution permettra un décompte du nombre de fibres vingt fois plus rapide qu'aujourd'hui tout en étant plus précis.

Il permettra ainsi de réduire les errances diagnostiques et thérapeutiques tout en améliorant les connaissances autour de cette maladie au sein de la communauté médicale.



Mieux comprendre les parcours de soins des maladies rares.



La problématique de départ :

Un établissement national de santé désirait mieux comprendre les parcours de soins associés aux patients nationaux atteints de maladies rares. Son objectif était double : améliorer la prise en charge de ses patients tout en limitant les frais médicaux.

Les maladies rares sont souvent synonymes d'impuissance : 95 % d'entre elles n'ont pas de traitement pharmacologique, et leur méconnaissance aggrave les erreurs de diagnostic et entraîne des dépenses de santé importantes avec peu de résultats pour les patients. **Les rares traitements qui existent sont souvent complexes, coûteux et les patients peuvent parfois attendre des années avant de voir poser un diagnostic précis.**



La solution imaginée :

Grâce à sa collaboration avec le Health Data Hub, l'établissement de santé a obtenu l'accès aux données du SNDS (système national des données de santé) chaînées avec celles de la BNDMR (banque nationale des données de maladies rares).

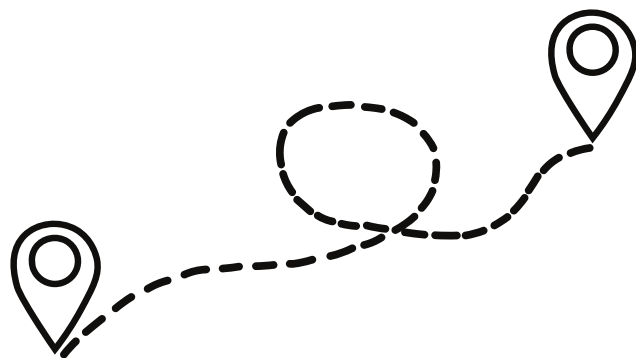
Ce chaînage présente l'innovation d'être déterministe afin d'associer les données cliniques et familiales du patient aux données sur les soins dont il bénéficie (médicaments, thérapies ...).

L'établissement de santé peut alors réaliser les analyses et publier études et rapports.



Les résultats :

Lier les données de santé des patients atteints d'une pathologie rare aux données de vie réelle a permis d'identifier une cohorte d'intérêt et de construire conjointement avec l'établissement de santé une étude qui a permis d'éclairer les décisions relatives à l'amélioration des parcours patient, à adapter l'offre de soins et identifier les traitements existants les plus efficaces selon le profil patient.



Ciblage automatique de patients sujets à des fractures chroniques pour améliorer leur qualité de vie et favoriser le développement de nouveaux traitements.



La problématique de départ :

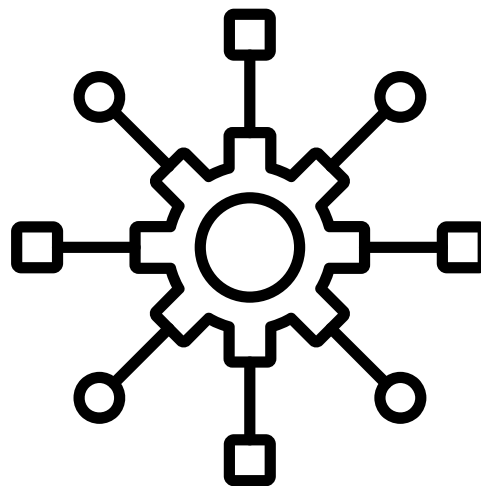
Dans le but d'étudier des fractures liées à des maladies chroniques comme l'ostéoporose, un établissement de soins a souhaité pouvoir constituer automatiquement des cohortes de patients d'intérêt. En effet, les conséquences pour les patients âgés de plus de 50 ans notamment peuvent s'avérer invalidantes : douleurs chroniques, perte d'autonomie, et même décès dans certains cas.

L'objectif de ce projet de recherche a été d'utiliser l'intelligence artificielle pour identifier automatiquement des patients afin de constituer une cohorte sujette au risque de fractures chroniques et ainsi définir un taux de survenue de nouvelles fractures.



La solution imaginée :

En chaînant les données de l'entrepôt de données de santé de l'APHP, du PMSI (programme de médicalisation des systèmes d'informations), du SNDS (système national des données de santé) ainsi que des données cliniques, l'équipe de développement a conçu un modèle de construction automatique de cohorte grâce à des techniques de traitement du langage naturel (NLP).

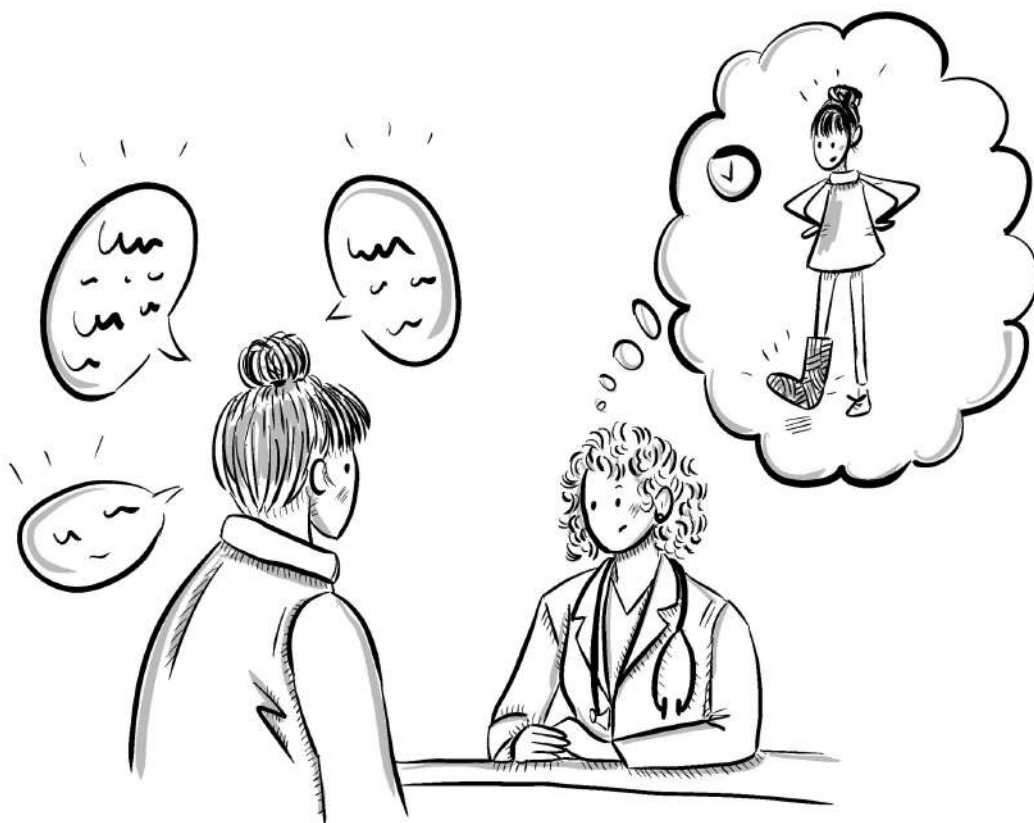


Plus précisément, les data scientists ont élaboré un algorithme de détection automatique de mots clés dans les comptes rendus des urgences et d'orthopédie, pour définir un taux de survenue de nouvelles fractures.



Les résultats :

L'innovation technique et clinique du projet consiste à générer automatiquement des cohortes. A terme, il permettra d'améliorer la prise en charge des patients en identifiant les facteurs de risques de fracture et de décès à très court terme.



Améliorer l'observance thérapeutique d'un traitement en vie réelle.



La problématique de départ :

L'équipe de développement clinique d'un laboratoire pharmaceutique a souhaité étudier en vie réelle son traitement contre une maladie génétique neuromusculaire qui entraîne une faiblesse musculaire progressive et dont l'issue est souvent fatale.

Sa volonté était d'observer la population traitée par son médicament en France (nombre de patients et évolution dans le temps en fonction des caractéristiques patients) et analyser l'utilisation de son traitement (posologie et observance).

A terme, elle souhaite comprendre l'impact du traitement sur la vie des patients traités, anticiper le délai de nécessité d'une canule respiratoire (dispositif utilisé pour maintenir ouvertes les voies aériennes d'un patient dans le coma) et renégocier avec les autorités compétentes les conditions économiques actuelles du traitement avant la commercialisation d'un nouveau traitement.



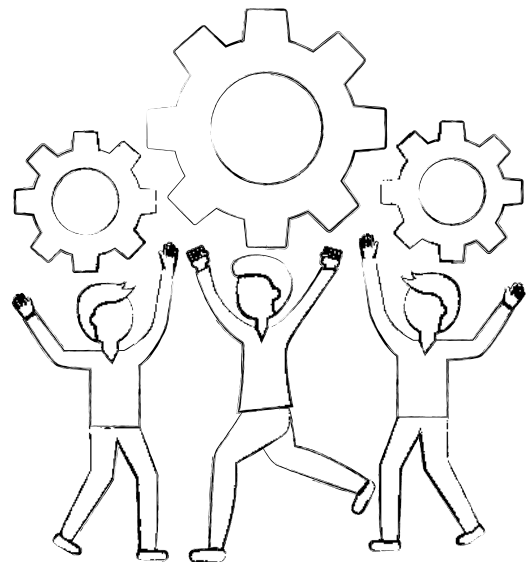
La solution imaginée :

Le laboratoire pharmaceutique avait besoin de l'accès au PMSI seul (programme de médicalisation des systèmes d'information) pour pouvoir mener à bien son projet.



Les résultats :

Les travaux sont aujourd'hui encore en cours.



Améliorer l'observance thérapeutique d'un nouveau traitement contre le cancer.



La problématique de départ :

L'équipe de développement clinique d'un laboratoire pharmaceutique souhaitait commercialiser un traitement contre un cancer récidivant et difficile à traiter dont l'incidence et les chances de survie sont faibles.

En essai contrôlé randomisé (ECR) de phase 3, ce laboratoire avait plusieurs combinaisons de sa molécule avec des traitements existants à tester. A l'aide de données de vie réelle, le laboratoire voulait identifier les meilleurs critères d'éligibilité afin de maximiser le nombre de patients éligibles sans compromettre le succès de l'étude de sécurité.



La solution imaginée :

Le laboratoire a commencé par identifier les sources de données de vie réelle compatibles avec les données des essais contrôlés randomisés (ECR). Deux bases américaines sont identifiées : Optum et Flatiron.

A partir de ces bases, l'équipe de développement a été en mesure de quantifier l'impact des différents critères d'éligibilité et de sélectionner les meilleurs critères pour maximiser le nombre de patients éligibles dans la cohorte d'intérêt.



Les résultats :

Les essais de phase 3 couplés aux données de vie réelle se sont montrés prometteurs et des améliorations statistiquement significatives ont été observées.

Cette molécule a notamment permis d'obtenir une réduction de 40% du risque de progression ou de décès chez des patients atteints de ce cancer



Pharmacovigilance

Surveillance, évaluation et prévention du risque d'effets indésirables des médicaments accessible à tous.



La problématique de départ :

Un établissement public de santé a envisagé de faciliter l'accès aux données sur les médicaments pour tous dans un souci de transparence. En particulier, il souhaitait partager et optimiser l'utilisation des rapports sur les effets indésirables des médicaments.

L'objectif poursuivi par cet établissement était de permettre aux utilisateurs d'accéder à des statistiques clés pour chaque médicament ou substance active relatives aux effets indésirables.



La solution imaginée :

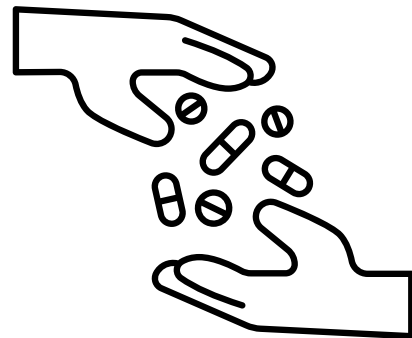
Accompagné par les experts du HDH, l'établissement public a pu prouver la faisabilité technique de son projet et identifier les bases nécessaires à la réalisation de son projet de recherche. Parmi elles, la Base Nationale de Pharmacovigilance (BNPV, ANSM), CODEX (ANSM), OPEN MEDIC ou encore le SNDS.

Il travaille maintenant sur la conception de la plateforme qui sera accessible à tous, tout en garantissant la protection des données personnelles.



Les résultats :

Le projet permettra, une fois terminé, pour chaque médicament ou substance active de pouvoir accéder librement au nombre de cas d'effets indésirables, leur pourcentage par sexe et âge, leur localisation par groupes d'organes ou encore connaître le taux de déclaration.



Ce projet bénéficiera à la fois aux patients, qui pourront prendre un rôle actif dans leur traitement, aux industriels de santé et aux institutions qui pourront adapter, modifier ou améliorer les traitements usuellement prescrits.

3. Chapitre 3

Conclusion

Par leur patrimoine de données d'une très grande richesse et leur complémentarité, les plateformes de données de santé « en vie réelle » dynamisent la recherche médicale.

Le potentiel de transformation est immense, à l'image de la diversité des cas d'usage présentés dans cet article.

Dans le cas spécifique du Health Data Hub en France, l'offre de ce dernier s'adresse à tous les acteurs de la recherche en santé : établissements de santé, industriels de santé ou encore institutionnels. Dans ce cadre, tous les porteurs de projets doivent répondre aux mêmes exigences spécifiques du Health Data Hub afin de concrétiser le lancement de leur projet de recherche.

Trois facteurs clés sont essentiels pour assurer le succès de cette démarche :

- Connaître et comprendre les bases de données et les outils mis à disposition par le HDH (capacités de calcul élastiques, outils de traitement mis à disposition au sein des espaces projets, etc.)
- S'assurer de la valeur et de la recevabilité des projets de recherche auprès des autorités de santé, par exemple sur l'observation des traitements en vie réelle comme « preuve alternative » dans le cadre de demande d'accès ou de remboursement.
- Gérer les aspects juridiques et organisationnels liés au lancement d'un projet de recherche sur le Hub (pré-autorisation de la CNIL, protocole scientifique), avec des délais parfois longs jusqu'à l'accès aux données.

Le mot de la fin de l'équipe de rédaction de Kynapse :

Nous espérons qu'au travers de cet article sur les plateformes de données de santé et l'illustration de cas d'usages concrets nous aurons pu donner un aperçu du potentiel représenté par les données de vie réelle.

Notre souhait est aussi que les différents acteurs concernés y adjoignent éventuellement leurs propres données afin d'en tirer le plus grand parti à travers leurs projets, pour le plus grand bénéfice des patients et de nos systèmes de santé !



kynapse
an open company

www.kynapse.fr