

SOMMAIRE

- **Introduction**
Intervention amicale
du Pr Bernard Bégaud,
Professeur de Pharmacologie
de l'Université de Bordeaux

- **TABLE RONDE 1**
**Quelles innovations
méthodologiques dans
l'exploitation des grandes
bases de données ?**
Animée par Alexandre
Vainchtock, AFCROs
En présence de
Olivier Bilicky, HIA Percy
François-Emerly Cotté,
Christelle Gastaldi-Ménager,
CNAMTS
Riad Kahoul, Novadiscovery
Sylvie Kermabon, Groupe VVY
Adeline Meilhoc, AFCROs

- **TABLE RONDE 2**
**Les études en vie réelle
multi-sources : futur
« gold standard » ?**
Animée par Hubert Méchin,
AFCROs
En présence de
Patrick Blin, CERES
Cécile Déal, Janssen
Pr Philippe Lechat, AP-HP
Sybil Pinchinat, AFCROs
Marion Pinet, HAS
Marie Eve Velard, Allergan



Colloque Données de Santé en vie réelle

JEUDI 7 JUIN 2018 - CIUP, PARIS XIV

Les études en vie réelle multi-sources : futur Gold Standard ?

La loi de modernisation du système de santé du 26 janvier 2016 a créé un nouveau cadre pour l'accès et l'utilisation des bases de données médico-administratives (PMSI, Sniiram, EGB...) dans un but de recherche, d'étude et d'évaluation. Pour sa 10ème édition, le Colloque « Données de santé en vie réelle » organisé par l'AFCROs (Association Française des CROs), en partenariat avec l'ADELF (Association des Epidémiologistes de Langue Française), a fait appel aux principaux acteurs du secteur (cliniciens, hospitaliers, HAS, industriels de la pharma et de la tech, CRO, Cnamts..) pour aborder les principales innovations méthodologiques permettant l'exploitation de ces grandes bases données. Après l'intervention du professeur Bernard Bégaud qui a milité en faveur du partage des données et du renforcement de l'expertise publique française, les échanges ont porté sur la place des études multi-sources. Alors que cette ouverture des données médico-administratives élargit la possibilité de réaliser ce type d'étude, celles-ci pourraient maintenant devenir la future référence.

Denis COMET président de l'AFCROs



Intervention du Professeur Bernard Bégaud Ouverture des données de santé en vie réelle : l'aboutissement de plus de 30 ans de lutte



Intervention amicale
du **Pr Bernard Bégaud**, Professeur
de Pharmacologie de
l'Université de Bordeaux

A l'occasion du 10^{ème} colloque Données de santé en vie réelle (DSVR) organisé par l'AFCROs en partenariat avec l'ADELFF (Association des épidémiologistes de langue française), le professeur Bernard Bégaud, professeur émérite de pharmacologie à l'université de Bordeaux II, a livré son point de vue et ses réflexions sur l'apport des données de santé en vie réelle et sur les perspectives de l'ouverture de leur utilisation.

Je parlerai des 33 ans de combat pour l'ouverture des données de santé, le combat d'une vie et le bonheur de sa concrétisation parce que ce colloque, la création du SNDS et de l'INDS en sont la preuve. J'ai eu un plaisir extrême à travailler avec Dominique Polton et Franck Von Lennep sur l'ensemble des données de santé. Ils sont maintenant à la manœuvre et nous pouvons être confiants. Nous sommes pourtant partis de très loin, mais beaucoup de choses ont changé. Nous avons enfin accès aux données de santé et on commence à croire sérieusement à l'évaluation dans la vie réelle, ce qui était impensable dans les années 1980 quand j'ai commencé ma vie professionnelle. Il n'était alors tout simplement pas question d'épidémiologie. La santé publique se portait très bien, faisait de belles études mais seul l'essai clinique comptait dans le domaine du médicament. Personne ne croyait en la validité des données obtenues par l'observation. J'ai toujours été abasourdi par cet antagonisme que l'on veut créer entre l'essai clinique, le plan expérimental et les données de vie réelle. Cette opposition est absolument ridicule mais elle persiste encore aujourd'hui pour de mauvaises raisons alors que ces deux évaluations sont complémentaires. Certains ne voient la vérité que dans l'essai clinique. Pourtant, ceux qui ont travaillé dans des essais cliniques ont des raisons de douter à la fois de la validité interne et de la validité externe bien que cela soit le fondement de l'évaluation totale et le seul fondement valable pour les références causales. Mais s'arrêter là est absolument ridicule. Je voudrais rappeler que, de John Snow jusqu'au traitement moderne de l'insuffisance cardiaque en passant par le trai-

tement des troubles bipolaires, l'observation a été l'outil des progrès médicaux. C'est elle qui guide le progrès en santé et en santé publique.

Débuts de la pharmaco épidémiologie

En 1985, encore en culottes courtes, j'ai rejoint un groupe de travail à Minneapolis qui a eu la prétention de créer la Société Internationale de Pharmaco épidémiologie alors qu'on ne se réunissait qu'aux Etats-Unis avec des Américains. Il s'est ensuite écoulé cinq ans jusqu'à ce que la société devienne réellement internationale, se constitue et alterne maintenant entre Etats-Unis, Europe et Asie. Cette aventure a été absolument extraordinaire parce que nombre de méthodes utilisées aujourd'hui dans l'évaluation épidémiologique ont été développées par les gens de la pharmaco épidémiologie pour les besoins de la pharmaco épidémiologie et pour ce facteur d'exposition très particulier et très changeant qu'est le médicament. Le médicament est probablement la chose la plus passionnante qui soit pour travailler et je veux rappeler deux étapes. La première étape a été l'étude de terrain quand la pharmaco épidémiologie a commencé. Il n'y avait pas d'accès aux données de santé et il fallait travailler avec des sujets qu'on devait aller recruter dans les hôpitaux auprès des professionnels de santé et ce, avec des budgets très limités dans le public. Le nombre de sujets aussi était très limité. J'ai passé des années de ma vie complètement obsessionnelles, un peu névrosées, à ajuster des calculs de taille d'échantillons par la loi binomiale, la loi binomiale négative, la loi de Poisson... Ceci ne servait strictement à rien car lorsqu'on arrivait à dire qu'il faudrait 4855 sujets par bras pour un risque de première espèce ou de seconde espèce déterminé, le sponsor répondait qu'il n'avait pas les financements nécessaires et l'étude était faite avec les moyens disponibles, c'est-à-dire avec un effectif deux fois inférieur à la taille prévue et avec au final, peu de résultats significatifs. Cette situation était très commode parce qu'elle permettait de demander la réalisation d'une autre étude et de réclamer un autre financement. Elle l'était aussi pour le décideur qui pouvait

dire que le résultat n'était pas significatif. Il n'avait pas à décider sauf en cas de force d'association extrêmement élevée avec des risques relatifs de hazard ratio de 10, 12, 15, 20. Ces niveaux ne sont plus utilisés mais ils ont permis, sans doute injustement, de retirer de nombreux médicaments du marché, parce que des associations très fortes étaient observées dans des populations très sélectionnées, mais avec un impact en population très faible.

Pour une saine démocratie sanitaire

La deuxième étape a été l'arrivée très progressive des bases de données. J'ai alors eu la malchance de pouvoir travailler sur les bases des HMO aux Etats-Unis, la RAMQ au Canada. A cette époque, nous Français, étions très honteux de ne pas avoir de base digne de ce nom. Puis, en 1999, il y a eu le coup de génie de la création du Sniiram (Système national d'information inter-régimes de l'Assurance maladie) qui a été inscrite dans la loi de financement de la sécurité sociale. Cette initiative a vraiment été un coup de génie. Dès le début, il a été souhaité que cette base de données soit utilisable et elle est aujourd'hui une des plus regardées dans le monde. Il suffit de regarder dans PubMed le nombre de publications citant le Sniiram ou l'EGB (Echantillon général des bénéficiaires), avec ses 60 millions de personnes, ses milliards de lignes, ses dix ans de recul, et ses nombreux croisements possibles par le NIR. Néanmoins, cette base reste lourde et très difficile à manipuler et nous avons proposé dans l'un de nos rapports sur les données de santé, ce que nous avons appelé le « maxi EGB » ou le « mini Sniiram », c'est-à-dire le Sniiram au dixième. Mais cette proposition n'a pas l'air de prendre vraiment.

En 2016, avec l'ouverture des données de santé, se concrétisent un rêve et un grand combat. Bien qu'ayant toujours été du côté du public, je me suis aussi toujours fortement battu pour l'ouverture des données de santé publiques au privé parce que je suis persuadé que la démocratie sanitaire n'est saine que si le public, le privé et les citoyens peuvent accéder aux mêmes données. C'est ainsi qu'on s'émule et qu'on peut vérifier qu'il n'y a pas de dictature. J'ai connu la dictature du privé quand on n'avait pas accès aux données des médicaments. Il n'y avait qu'IMS et ce n'était pas sain et il n'était pas sain non plus que le privé n'ait pas accès aux mêmes données que les cher-

cheurs Inserm. Cette émulation est très saine et constitue un énorme progrès.

Une expertise à amplifier

Je pense qu'il y a une nécessaire prise de conscience. Il va falloir renforcer sérieusement l'expertise et la formation en santé publique qui est encore faible au niveau des journalistes, du public et des politiques. L'ouverture de grandes masses de données signifie que des personnes peu expérimentées pourront générer des résultats aberrants par manque de formation, d'expérience en pharmacologie, en épidémiologie ou en informatique. Il faudra avoir le courage et la force de répondre sinon la dictature de l'obscurantisme que l'on voit déjà avec le complotisme va rendre la démocratie sanitaire irrespirable. On va voir sans arrêt accuser les médicaments et réclamer des têtes. J'espère que les pouvoirs publics ont saisi l'enjeu de ce risque. Il faudra aussi une expertise extrêmement solide pour protéger la santé publique, pour protéger le médicament et pour protéger le citoyen des peurs que cela peut engendrer.

Ma dernière réflexion concerne l'équilibre public privé. On parle beaucoup, trop du reste, des liens et des conflits d'intérêt mais l'équilibre et le déséquilibre entre public et privé dans ce domaine de l'évaluation n'intéressent personne. Il est très important de faire attention à ce que la nécessaire compétence qui s'aspire d'un côté et de l'autre soit à peu près équilibrée pour respecter ce jeu. Je souhaiterais que les pouvoirs publics aident effectivement le public à travailler avec le privé et à être un vrai partenaire pour ne pas subir. Il faudra aussi augmenter les flux de formations qui sont très petits en France et créer des structures communes inter organismes. C'est une grande satisfaction de voir ce qui se passe en France. Il y a encore cinq ans, nous n'étions pas fiers quand nous allions à l'étranger. Maintenant, nous le sommes beaucoup plus et, pour terminer sur une note de perspective, je suis persuadé que dans cinq ans, nous allons pouvoir « casser la baraque ». Vous savez ce qui se prépare avec les liens du Sniiram avec la biologie, les résultats des examens de laboratoire... On va donc pouvoir vraiment contourner de façon quasi satisfaisante le grand manque de cette base que sont les indications, les facteurs de risque et les maladies.

TABLE RONDE 1

Quelles innovations méthodologiques dans l'exploitation des grandes bases de données ?



Alexandre VAINCHTOCK, animateur de la table ronde, membre de l'AFCROs

Avec quelque 60 millions d'individus sur dix ans d'historique et 20 milliards de lignes de prestations de remboursement, les bases de données médico-administratives (PMSI, Sniiram, EGB...) ouvrent l'accès à un volume de données gigantesque. Leur exploitation nécessite des méthodes d'analyse aptes à gérer à la fois ces volumes et la complexité des relations et des corrélations entre ces données. Plusieurs possibilités se dessinent déjà, à l'instar du renouveau

des techniques statistiques classiques grâce à la puissance de l'informatique, des technologies de machine learning ou de process mining, ou encore de l'utilisation concomitante de méthodes statistiques, génomiques, informatiques... Producteurs et utilisateurs de données, ainsi que ceux qui les accompagnent, livrent leurs expériences sur l'exploitation des données du SNDS et sur les types de technologies utilisables en fonction des besoins exprimés.



« Le PMSI pour suivre la prise en charge du cancer bronchique »



François-Eméry Cotté,
directeur associé Health
Economics & Outcome
Research chez BMS

Comme beaucoup d'industriels, BMS a mis en place depuis de nombreuses années des projets s'appuyant sur les bases de données du SNDS, essentiellement sur des données du PMSI. Nous avons d'abord mis en place des études sur le Sniiram, avec des designs relativement simples. Il s'agissait essentiellement d'études de « fardeau de la maladie » pour documenter le nombre de patients dans une indication ou une pathologie donnée et évaluer le coût associé à leur prise en charge. L'avènement du chaînage a permis ensuite de travailler sur des protocolisations plus poussées avec des intérêts scientifiques plus pertinents.

Notre étude « Territoire » portant sur l'équité d'accès aux soins en cancérologie a été réalisée avec les données du PMSI sur la base d'une méthode originale dans son approche. Celle-ci comprenait deux volets, l'un sur la visualisation de la prise en charge des patients atteints de cancer bronchique et l'autre sur les inégalités de prise en charge et l'équité d'accès aux soins. Nous avons identifié une cohorte incidente de patients atteints de cancer bronchique sur une année calendaire, ici l'année 2011. Celle-ci comprenait un peu plus de 40 000 patients suivis sur deux ans. Cette durée nous a paru pertinente, compte tenu de la pathologie et de son pronostic. Avec le volet sur la visualisation du parcours de soins, nous voulions avoir une lecture fine de la prise en charge avec les acteurs de terrain. Compte tenu du nombre de patients concernés, nous nous sommes donc focalisés sur quelques régions dont la région Nord – Pas de Calais. Nous avons d'abord réalisé une cartographie classique représentant les établissements de santé ayant codé pour la première fois un cancer bronchique chez un patient et les communes de résidence de ces patients. Ce type de visualisation est classique mais très utile pour comprendre la prise en charge sur un territoire donné et pour discuter de l'attractivité d'un établissement de santé sur un territoire. Une autre visualisation, plus difficile, a aussi été effectuée sur l'ensemble des séjours en établissement de santé intervenus sur les deux ans de suivi. A partir de là, nous avons cherché à identifier les collaborations entre établissements de santé. L'existence d'une collaboration importante entre le CHU de Lille et le CLCC Oscar Lambret apparaît alors assez nettement, mais l'interprétation est plus difficile sur les autres établissements.

Enfin, il nous a fallu trouver des modes de visualisation pertinents pour analyser plus de 500 000 séjours pour ces 40 000 patients. Nous nous intéressions à ce qui pouvait se passer avant et après le diagnostic et notre choix s'est porté sur une visualisation utilisant des disques concentriques, le premier séjour étant le disque le plus au centre. Il est ainsi possible de voir que la moitié des patients avait été hospitalisée dans les 12 mois précédant le diagnostic, dont 17 % l'ont été d'emblée en soins palliatifs. Notre intérêt s'est aussi porté sur la chimiothérapie post chirurgie curative, la séquence de soins étant très protocolisée dans le cancer bronchique avec un délai de 12 semaines qui ne doit normalement pas être dépassé. Ce point a été particulièrement travaillé en tant qu'indicateur de santé et a fait l'objet de communications scientifiques sur l'analyse de l'équité d'accès aux soins. D'autres indicateurs tels que le diagnostic tardif, l'accès aux molécules de la liste en sus et l'hospitalisation à domicile ont aussi été suivis. Le lieu de résidence du patient a été traité en variable explicative à partir de techniques traditionnelles et nous l'avons retraité en indice de défaveur sociale mais aussi en densité de population, qui est un proxy de la ruralité. Nous avons aussi testé sur certains indicateurs l'accessibilité potentielle localisée, qui est en fait le niveau d'accès à la médecine générale et la densité de pneumologues au niveau du département.

« Le process *mining* pour l'étude des parcours de soins »



Christelle Gastaldi-Ménager, Direction de la Stratégie, des Etudes et de la Statistique – Cnamts

Ceux qui ont déjà pu utiliser le Sniiram se sont vite rendu compte de la complexité de son exploitation et de l'importance des délais de traitement. Il faut parfois attendre plusieurs heures voire une journée entière pour obtenir les résultats des requêtes portant sur des médicaments. De plus, les méthodes classiques ne permettent plus d'exploiter suffisamment toute la richesse des informations disponibles. La Cnamts a donc souhaité voir en quoi les nouvelles technologies du big data pourraient répondre à ces besoins pour améliorer l'exploitation de l'analyse des données, pour détecter des signaux de faible intensité ou des anomalies en pharmaco-épidémiologie, ou encore pour identifier des facteurs utiles à l'analyse des parcours de soins, le mésusage ou les fraudes. Aussi est-elle très ouverte et très intéressée par l'établissement de collaborations et deux expériences sont d'ores et déjà en cours. Un partenariat de trois ans a été établi avec l'Ecole Polytechnique en 2015 et renouvelé début 2018. L'objectif est double. Pour Polytechnique, il s'agit d'avoir « un terrain de jeu » pour développer et tester leurs algorithmes d'exploitation des données. Pour la Cnamts, ce partenariat ouvre la possibilité d'accéder aux techniques innovantes développées par Polytechnique. Les trois premières années du partenariat ont été consacrées essentiellement à l'appropriation des données du Sniiram par Polytechnique. Dans un premier temps, il a fallu réorganiser et restructurer cette base pour la rendre compatible avec les technologies et les outils que Polytechnique comptait utiliser et ce, tout en conservant les contraintes de sécurité des données. Dans un deuxième temps, l'utilisation de ces algorithmes a été validée sur des résultats déjà démontrés par les méthodes classiques. Ce travail a été effectué en vérifiant que ces outils permettaient de retrouver l'existence des liens entre utilisation de pioglitazone et risque de cancer de la vessie.

Maintenant que le partenariat entre la Cnamts et Polytechnique a été reconduit, l'objectif est d'appliquer ces technologies en routine. Un premier projet porte sur la prise de rosiglitazone et la survenue de problèmes cardiovasculaires et sur les interactions entre les associations de deux classes thérapeutiques. Un deuxième projet s'inscrit dans le cadre de l'évaluation du programme SOPHIA Diabète mis en place par l'Assurance maladie sur huit années. Ici, nous sommes intéressés par une approche de process mining et de machine learning qui nous a été proposée par la société HEVA pour travailler sur les parcours de soins. Le process mining permet de travailler sur l'enchaînement chronologique d'évènements, ce qui est particulièrement adapté en santé pour l'étude des parcours de soins. Cette méthode nous permet d'analyser les parcours de soins des adhérents au programme SOPHIA et d'appliquer la même démarche aux non adhérents afin de pouvoir identifier des parcours atypiques et de mettre en évidence leurs différences. Deux cohortes ont été constituées avec d'une part les adhérents (17 000 patients) et les non adhérents (32 000 patients) au programme. Nous avons aussi identifié une cinquantaine d'étiquettes, à savoir des évènements d'intérêt dans le cadre d'un parcours de soins (hospitalisations spécifiques, actes de suivi, contacts avec les professionnels de santé...) que nous allons observer sur les huit années du programme. Les travaux sont en cours et nous sommes impatients de voir si leurs résultats nous permettent de mieux comprendre les parcours et d'identifier des différences entre eux.

« Des données à l'appui de programmes de prévention santé »



Sylvie Kermabon, Groupe mutualiste VYV

Le groupe VYV a été créé en 2017 et rassemble des ensembles mutualistes couvrant aujourd'hui une dizaine de millions de personnes. A l'instar de la Cnamts, nous cherchons des approches méthodologiques différentes pour enrichir les données patients et les croiser ensuite avec d'autres sources de données publiques ou privées. Nous avons mis en place Vivoptim qui est un programme de prévention santé pour le suivi des patients à risque cardiovasculaire. Après avoir été en phase pilote pendant deux ans, ce programme est actuellement en cours de généralisation. Tout en respectant le nouveau règlement européen sur la protection des données, nous souhaitons pouvoir enrichir les données de chaque patient inclus dans Vivoptim pour leur proposer des services liés à la connaissance de leur situation et faciliter ensuite leur prise en charge par les médecins. Pour y parvenir, nous allons dans un premier temps collecter les données de laboratoires d'analyses médicales. Ces patients constituant des populations à risque, nous souhaitons pouvoir croiser ces informations avec des bases de données liées à la qualité de l'air et aux données environnementales. Nous testons actuellement différentes méthodes de croisement et nous prévoyons de proposer ensuite des programmes de prévention santé aux sportifs pour lesquels la qualité de l'air externe joue aussi un rôle. Par ailleurs, le groupe VYV a des contrats collectifs avec des entreprises très diverses et, dans ce cadre, nous proposons régulièrement aux employeurs des actions et des mesures afin de les aider pour l'amélioration de la santé et de la qualité de vie au travail de leurs employés. Notre objectif est maintenant de définir une méthode qui nous permette de déployer ce type de modèle sur d'autres croisements de données pour pouvoir les mettre à la disposition des entreprises en fonction de leurs besoins particuliers ou de leurs besoins liés à des programmes de prévention santé.



« Le machine learning pour orienter des actions de santé publique »



Olivier Bilicky, pneumo-oncologue à l'Hopital d'Instruction des Armées Percy

Pour un praticien, les bases de données sont un complément à la recherche clinique. Les critères d'éligibilité étant un peu particuliers dans les essais cliniques, notamment en oncologie, les études en vie réelle vont montrer que les médicaments sont potentiellement un peu moins efficaces dans l'ensemble de la population des patients. Mais on ne dispose pas de données de fin de vie sur certaines prises en charge particulières. Aussi nous sommes nous intéressés aux critères d'agressivité du cancer bronchique pour mieux évaluer l'espérance de vie et mieux anticiper la fin de vie du patient. Ce travail a utilisé des méthodologies classiques, ici, l'analyse multivariée avec l'inclusion de deux cohortes comprenant au total 90 000 patients. Les résultats sur la deuxième cohorte de 41 000 patients indiquent que 57 % des patients vont présenter un critère d'agressivité en fin de vie. 23 % d'entre eux reçoivent ainsi une chimiothérapie dans leur dernier mois de vie, ce qui constitue un taux deux fois supérieur à ce qu'on observe dans les pays anglo-saxons. Les critères d'agressivité ont ensuite été analysés au moyen de méthodologies classiques qui permettent d'en obtenir une liste, mais pas de déterminer les profils de patients auxquels il faut faire plus attention. Dans cette optique, en partenariat avec HEVA, nous avons utilisé le machine learning avec un algorithme d'arbres de décision renforcé afin de créer des profils types de patients chez lesquels on pourrait estimer être plus agressif. Nous sommes en train d'analyser les résultats et un premier profil type se profile. On voit ainsi qu'il y a au moins un critère d'agressivité chez 8 % des patients. Ce sont en particulier des malades de moins de 79 ans ayant une DMS (durée moyenne de séjour) courte à l'inclusion. En complément, l'analyse médico-économique montre un surcoût moyen de 3000 euros par patient chez les malades pour lesquels on est le plus agressif. Pour la seule oncologie thoracique, le coût total des 30 derniers jours de vie est ainsi estimé à plus de 200 millions d'euros par an. Alors que la diminution de ces coûts grâce à une prise en charge précoce en unités mobiles de soins palliatifs est déjà bien documentée dans la littérature, de telles études pourraient montrer qu'à coût et efficacités équivalents, l'attribution de financements supplémentaires pour les équipes de soins palliatifs pourrait contribuer à diminuer le coût de la fin de vie. Si ces techniques de machine learning sont probablement importantes pour les données de santé en vie réelle, elles le sont également pour les cohortes cliniques. Je pense notamment au programme ESMÉ (Épidémio-Stratégie Médico-Économique) d'Unicancer où il serait intéressant d'intégrer le plus grand nombre de variables potentielles pour pouvoir créer des profils types de patients*.

(*) Le programme ESMÉ a été lancé en 2014 afin de centraliser les données de vie réelle des patients traités pour un cancer en France. Son objectif est de décrire l'évolution au cours du temps de la prise en charge des patients et des stratégies thérapeutiques.

Le rôle complémentaire de la modélisation *in silico*



Riad Kahoul,
biomodélisateur chez
Novadiscovery

La modélisation *in silico* constitue une approche innovante qui n'a pas vocation à remplacer les méthodologies existantes mais plutôt à les compléter. Dans ce contexte, il est nécessaire de se poser trois questions majeures. Comment fusionne-t-on des données issues de sources différentes ? Comment interprète-t-on les résultats de ces études ? Enfin, comment transpose-t-on et comment extrapole-t-on ces modèles *in silico* à d'autres populations de patients ? Tout d'abord la fusion de données hétérogènes se heurte à des difficultés telles que les différences de caractéristiques des populations incluses ou des variables mesurées. Ces problèmes peuvent être résolus en développant des populations virtuelles à partir des données et des connaissances autour de la population d'intérêt. Dans ce cas, on ne fusionne plus les données elles-mêmes, mais plutôt les informations qu'elles contiennent. In fine, on construit donc une population virtuelle à partir des distributions marginales théoriques des variables et des corrélations entre ces dernières.

Aujourd'hui les analyses post-AMM se basent sur des études de phase IV généralement longues et coûteuses. Celles-ci présentent aussi l'inconvénient de ne pas contenir de groupes comparables et ne sont généralement pas randomisées, ce qui rend l'interprétation de leurs résultats difficile. Actuellement la méthode des scores de propension est la méthode la plus en vogue pour transformer une étude de suivi en une étude comparative démontrant l'efficacité, mais celle-ci présente des inconvénients. Pour contourner ces problèmes, nous utilisons une approche de modélisation *in silico* basée sur une population modèle et des modèles prédictifs. Ces derniers sont estimés sur une base de données d'essais cliniques randomisés, ce qui permet de capturer l'efficacité du traitement. Ils sont ensuite appliqués à la population virtuelle générée, qui est représentative de la population cible. La combinaison de cette population virtuelle avec les modèles prédictifs permet ainsi de simuler des essais cliniques randomisés. Cette approche permet de transposer les résultats observés dans des essais cliniques vers des populations cibles prises dans la vie réelle.



« Des enjeux multiples pour les données de santé en vie réelle »



Adeline Meilhoc, membre du groupe de travail DSVR de l'AFCROs

Nous sommes face une déferlante d'innovations technologiques. Le nombre d'objets connectés est estimé à près de quatre millions en 2021 pour un marché total de 140 milliards de \$. Ces produits vont générer une masse de données énorme et utilisable pour évaluer une large gamme de paramètres. Parallèlement, on observe une réduction du coût et une miniaturisation croissante des capteurs qui nous permettent des diagnostics en temps réel tandis que les progrès des technologies de calcul et de transmission permettent de répondre au besoin de partage des données. Le vieillissement de la population, le besoin de réduire les coûts des maladies chroniques et de mesurer l'impact des changements d'usage ainsi que l'optimisation des outils de diagnostic, de traitement et de prévention constituent trois des leviers qui devraient contribuer au développement du marché de l'utilisation des données de santé en vie réelle.

De plus, le rapport Villani de mars 2018 place la santé parmi les secteurs prioritaires où doivent se concentrer les efforts de développement dans le domaine de l'intelligence artificielle. Son premier chapitre met l'accent sur les grandes bases de données et sur le renforcement de l'écosystème européen, avec pour principaux chapitres l'incitation des acteurs économiques à la mutualisation des données ou encore l'organisation de l'ouverture des données détenues par des entités privée. Les enjeux de santé publique et de politique sont donc très présents.

Différentes méthodes peuvent être utilisées, depuis le machine learning, le deep learning exploité essentiellement pour l'imagerie médicale aujourd'hui jusqu'au langage processing. Bien que cette dernière soit encore peu appliquée, les algorithmes d'analyse d'expression des textes peuvent apporter des nouveaux services aux patients en améliorant l'analyse de leurs sentiments dans les questionnaires de qualité de vie. CRO et sociétés de services constituent un formidable vivier de compétences pour apporter ces méthodologies à leurs clients. Outre une compréhension très juste de l'intelligence artificielle, elles concentrent des expériences multiples (logistiques, médicales, scientifiques...) et sont des expertes de la donnée de santé et des bases de données. Enfin, lorsqu'on parle de données de santé en vie réelle, apparaissent des notions de cartographie, d'indicateurs de soins, de signaux, de données d'impact, de mesures, de profils, ou encore de critères d'agressivité, qui passeront toutes par une interopérabilité, qui devra s'accompagner non seulement d'une nécessaire harmonisation, mais aussi de la définition de « gold standard » et du développement de la formation associée.

TABLE RONDE 2

Les études en vie réelle multi-sources : futur « gold standard » ?



Hubert Méchin, animateur de la table ronde et membre de l'AFCROs

L'ouverture des bases de données médico-administratives s'est accompagnée d'une multiplication des projets scientifiques utilisant des sources d'une diversité croissante (données issues d'objets connectés, entrepôts de données hospitaliers, initiatives académiques...). Il est ainsi très probable que nos études en vie réelle soient de plus en plus multi-sources ou hybrides.

Comment réaliser des études multi-sources ? Deviendront-elles le « gold standard » ? Experts de la HAS, du CERES et de l'AP-HP, industriels du médicament et du dispositif médical et CROs ont partagé leurs expériences et développé l'intérêt et les limites de ces études multi sources à partir de l'évocation de cas concrets.



Les études multi sources prennent une place croissante



Marion Pinet, pharmacien,
responsable du service
Etudes en vie réelle de la
HAS

Le nombre d'études post-inscription demandées par la commission de la Transparence reste stable, avec une moyenne de 14 demandes sur les cinq dernières années, soit une pour dix dossiers évalués. En 2017, la commission a formulé douze demandes avec deux cas de figure principaux. Le premier correspond à des médicaments ayant une ASMR (Amélioration du Service Médical Rendu) 3 ou 4 pour lesquels la commission a reconnu un apport thérapeutique par rapport à l'existant mais a identifié un besoin de connaissances sur l'utilisation en pratique, sur les caractéristiques des patients ou sur l'efficacité hors du contexte des essais. Ces demandes sont des demandes de données observationnelles descriptives.

Le deuxième cas de figure concerne des médicaments sans ASMR (ASMR = 5), soit six médicaments l'année dernière dont quatre avec un SMR (Service Médical Rendu) modéré ou faible. La commission a demandé des données complémentaires d'efficacité sans préciser la méthode à employer. Il s'agissait de deux produits en oncologie prometteurs pour lesquels les données disponibles au moment de l'évaluation étaient extrêmement préliminaires et de deux produits pour lesquels la commission aurait souhaité une comparaison d'efficacité par rapport au standard existant.

Jusque dans un passé extrêmement récent, les études post-inscription, sauf cas exceptionnel, étaient synonymes d'études observationnelles ad hoc de terrain menées à l'initiative d'un industriel. Ces études construites spécifiquement pour répondre à la question de la Commission de la Transparence n'ont pas disparu. La Commission de la Transparence en a reçu sept l'année dernière, ce qui représente près du tiers de l'ensemble des 20 projets qui lui ont été proposés en 2017. Sur les 13 projets restants, cinq correspondent à des études prévues au moment de l'autorisation de mise sur le marché dans un plan de gestion des risques. Les huit autres sont réalisées sur des données préexistantes, l'une à partir du SNDS et les sept autres étant des analyses secondaires à partir de cohortes académiques. Ces cas de figure sont assez nouveaux et livrent plusieurs signaux intéressants. Tout d'abord, on sort d'une logique « une demande, une étude créée de novo » pour aller vers une clarification et une utilisation de l'existant. Nous y voyons aussi la marque d'une diffusion de la culture des partenariats public-privé avec à la fois des cohortes ouvertes à une utilisation secondaire de leurs données mais aussi des études suffisamment structurées pour permettre l'accès à un tiers, en l'occurrence un tiers privé.

Pour déterminer le nombre d'études pouvant être considérées comme des études multi sources, j'ai choisi arbitrairement une définition très large qui inclut des études traditionnelles associant des questionnaires médecin à des auto-questionnaires patient. Au total, neuf études pouvaient donc être qualifiées de multi-sources : six d'entre elles intégraient des questionnaires médecin et des questionnaires patient tandis que les trois études comportaient des collections biologiques. Dans les trois cas il s'agissait de cohortes académiques importantes avec financement privé, Hepater dans le domaine des hépatites, Radico dans le domaine des mucopolysaccharidoses et une cohorte du Grand Ouest sur les polykystoses rénales. Enfin deux études étaient des cohortes avec un appariement probabiliste au SNDS. Le principe de l'association de plusieurs sources de données a toujours fait partie du paysage de la pharmaco-épidémiologie et donc des études post-inscriptions. Jusqu'à maintenant les études associant plusieurs sources dans le cadre des études post-inscription étaient souvent minoritaires et ambitieuses, répondant à la fois à des objectifs de recherche académique et des demandes des autorités. L'ouverture du SNDS nous a permis de voir apparaître en 2017 des protocoles associant des données de cohortes et des données du SNDS. Etant donné le potentiel et la valeur du SNDS et l'apport de sa mise en relation avec d'autres sources de données, ce type d'études est amené à augmenter. Sur les 12 protocoles que nous avons été amenés à analyser au premier semestre 2018, quatre sont ainsi basés sur un appariement probabiliste avec le SNDS. Au-delà des données médico-administratives, de façon plus générale, une journée comme celle-ci montre l'extrême dynamisme et la richesse des études construites sur une logique multi-sources et celles-ci vont prendre une part grandissante dans les travaux d'évaluation du médicament.

Premier bilan de l'activité du CEREES



Patrick Blin, vice-président du CEREES

Le CEREES (Comité d'Expertise pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé) a été créé par la loi de modernisation de santé du 26 janvier 2016. Il est constitué de 21 membres nommés par décret pour un mandat de trois ans renouvelable une seule fois et est présidé actuellement par Jean-Louis Serre. Lorsqu'un dossier est soumis à l'INDS (Institut National des Données de Santé), il est transmis aux membres du CEREES dans la semaine qui suit et nous disposons d'un mois pour l'évaluer. Nous nous réunissons douze fois par an dont deux en septembre pour compenser l'absence de réunion au mois d'août. Pendant ces réunions, l'évaluation est collective et l'expérience du CCTIRS (Comité Consultatif sur le Traitement de l'Information en Matière de Recherche dans le Domaine de la Santé) qui a précédé le CEREES nous a conduits à limiter à quatre le nombre de dossiers par évaluateur. Chaque réunion fait ensuite l'objet d'un procès-verbal.

Nos travaux ont commencé en septembre dernier et dès cette première réunion, nous avons eu une vingtaine de dossiers à évaluer. Leur nombre augmente progressivement et nous sommes maintenant plutôt à une quarantaine de dossiers à examiner par séance. Les études « terrain » constituent la moitié des dossiers examinés tandis qu'un tiers d'entre eux relève du SNDS (EGB, PMSI, CEPIDC...), le SNDS dans son ensemble représentant la majorité. Les partenaires formulant ces demandes émanent pour près des deux tiers du secteur public (CHU, CLCC, universités, Inserm-CNRS, ministères), les requêtes provenant d'industriels, d'associations ou de bureaux d'études atteignant 24%.

Sur l'ensemble des avis rendus, 16 % sont des avis favorables sans aucune remarque particulière. La plus grande partie des avis (56 %) concerne des avis favorables avec recommandations plus ou moins importantes selon les dossiers. Le taux de dossiers pour lesquels un avis réservé a été émis atteint tout de même 22 %. Un avis défavorable a été délivré dans 5 % des cas et le CEREES n'a pas pu émettre d'avis pour 2 % des dossiers, compte tenu des éléments qui lui avaient été transmis.

Aujourd'hui, nous avons des demandes de plus en plus fréquentes pour des appariements entre les données issues du terrain et les données du SNDS. La plupart de ces projets reçoivent un avis favorable lorsque les demandes sont bien centrées avec des objectifs et une analyse bien définis, et ce, qu'il s'agisse d'un appariement probabiliste via la CNAM-TS ou d'un appariement utilisant les données du NIR avec un système de tiers de confiance. Nous avons aussi reçu des projets très vagues sans objectif ni méthode définis qui ont reçu un avis réservé en premier lieu voire un avis défavorable.

Enfin l'arrivée des futures méthodologies de référence devrait se traduire par une baisse de l'activité du CEREES au niveau des dossiers du type public (CH, CLCC...) mais je n'attends pas tant de diminution sur la partie SNDS. Une procédure simplifiée est cours de finalisation pour l'accès à l'EGB.

Apparier PMSI et données de vie réelle pour le DM



Marie-Eve Velard, directrice
Market Access and Public
Affairs chez Allergan

Nous avons déjà réalisé des études multi-sources incluant des recueils vers les patients et vers les professionnels de santé. Mais le multi-source englobe aussi les entrepôts de données. Comparativement au médicament, la difficulté dans le cas du dispositif médical réside dans le fait que ces produits sont souvent utilisés en chirurgie et sont parfois inclus dans des GHM (Groupe Homogène de Malades) génériques. Les dispositifs ne peuvent donc pas être suivis et il est obligatoire de réaliser des études en vie réelle. Concernant des sujets importants tels que les complications, il serait très intéressant de pouvoir apparier avec le PMSI et nous y travaillons. Mais je ne pense pas que les études multi-sources soient un futur « gold standard » qui puisse s'imposer dans le secteur du dispositif médical. Celui-ci est constitué d'un grand nombre de petites entreprises qui auront beaucoup de mal à faire ces études en raison de leur coût. Il faut garder aussi des études en vie réelle. La réalisation d'études basées sur le seul SNIIRAM n'est possible qu'avec des dispositifs médicaux utilisés en ville dont on veut suivre l'utilisation dans la vie réelle. Mais les paramètres biologiques et les informations relatives à la qualité de vie ne sont pas accessibles par ce biais. Enfin il serait très intéressant de pouvoir apparier données en vie réelle et bases médico-administratives.

Des limites au SNDS pour la psychiatrie



Cécile Déal, responsable
RWE chez Janssen

A la suite d'une demande de la HAS dont la date a coïncidé avec l'ouverture des données du SNDS, nous avons opté pour une étude multi sources faisant appel à un observatoire transversal. Ce choix a été fait afin de capter des données qualitatives médicales et cliniques qu'on ne peut pas trouver dans le SNDS. Pour cette étude en psychiatrie chez des malades atteints de pathologies chroniques lourdes, un suivi était requis avec des demandes concernant les caractéristiques des patients, leurs prescriptions, leur prise en charge ainsi que des données qualitatives. Il était donc nécessaire de mettre en place un observatoire en vie réelle pour chercher des données dans les dossiers patients ou pour interroger les médecins en face à face. Etant donné qu'il s'agit d'une pathologie chronique avec un traitement peu prescrit, nous n'avons réalisé qu'un seul observatoire classique longitudinal, mais nous n'avions pas anticipé de pouvoir fournir des résultats avant quatre à cinq ans, alors que la HAS s'interrogeait plutôt sur l'initiation de ce traitement. Maintenant nous allons répondre avec deux études, mais est-ce pour autant un « gold standard » ? Il existe en effet deux restrictions très importantes à l'utilisation du SNDS en psychiatrie. Dans le domaine de la schizophrénie, la qualité du diagnostic tel qu'il a été renseigné dans la base qui était au départ dans le SNDS constitue une vraie limite. Par ailleurs, le SNDS, dans son état actuel, ne peut pas être chaîné avec les données du PMSI telles que celles dont nous avons besoin pour ce médicament donné en ambulatoire. En effet, un patient schizophrène recevant le traitement étudié est un patient suivi en ambulatoire dans des CMP (Centre Médico-Psychologique) qui, en France sont rattachés à des hôpitaux. Si le patient n'a pas été hospitalisé au sein de l'établissement de rattachement du CMP, tout ce qui s'y passe est invisible. Le SNDS donnera accès à la délivrance en pharmacie et aux consultations généralistes mais le nombre de consultations avec le médecin psychiatre, l'infirmière, le psychologue et le psychothérapeute ne pourra pas du tout être documenté. Enfin, notre projet ne vise pas nécessairement un appariement mais nous allons bien sûr vérifier et valider la représentativité de la population d'une étude transversale par rapport à une population issue des données du SNDS. Dans le domaine de la psychiatrie, je fonde beaucoup d'espoirs sur les entrepôts de données qui vont nous permettre réellement d'avoir des données qualitatives médicales et je voudrais citer l'exemple de l'OFSEP (Observatoire Français de la Sclérose en Plaques). Cet observatoire a été visionnaire à l'époque de son montage et il équivaut à un entrepôt de données. Il me paraît nécessaire d'engager tout de suite une telle démarche dans le domaine de la psychiatrie parce que la construction d'un entrepôt de données de qualité est compliquée.

L'entrepôt de données de l'AP-HP



Philippe Lechat,
responsable de la mise en œuvre et de la gestion de l'entrepôt de données de l'AP-HP

Le dossier patient informatisé et uniforme a été mis en place depuis de longues années à l'Assistance Publique – Hôpitaux de Paris (AP-HP). Cette entreprise ambitieuse n'est pas encore complétée mais ce dossier unique rassemblant données cliniques, données biologiques, imagerie, PMSI et circuit du médicament sera disponible dans un avenir proche sur l'ensemble des 39 hôpitaux de l'AP-HP qui représentent un million de patients. Lorsque j'ai commencé à travailler à la mise en place de cet entrepôt de données, j'ai demandé très rapidement à mes collègues s'ils avaient envisagé d'utiliser cet outil pour la recherche. Etant donné qu'il n'est pas possible de faire travailler directement sur le dossier médical informatisé, les données ont donc été exportées sur d'autres bases qu'on appelle les entrepôts. Celui de l'AP-HP contient les données cliniques, ce qui en constitue en grande partie son intérêt, en complément du SNDS. Néanmoins actuellement, les données cliniques sont principalement des comptes-rendus d'hospitalisation et des comptes-rendus opératoires, et ce avec un enjeu considérable en terme d'analyse textuelle. Mais surtout, nous passons à une nouvelle ère. Jusqu'à présent, les équipes sont compétitives et celle qui va le plus vite et a les meilleures données les garde. Mais dorénavant, la philosophie est de les partager et il sera possible par exemple de faire des études sur la schizophrénie dans l'ensemble des services de psychiatrie de l'AP-HP. Des règles de gouvernance ont été établies sur l'information des patients et sur l'information des professionnels et elles ont été validées par le comité médical de l'AP-HP. Il a aussi été nécessaire de négocier avec la CNIL et l'entrepôt de l'AP-HP a d'ailleurs été le premier entrepôt autorisé par cette instance. Un comité scientifique que j'ai l'honneur de co-présider a aussi été créé pour analyser et vérifier la validité scientifique et la faisabilité des projets. L'utilisation de ces données est ouverte à des partenaires extérieurs, mais ces derniers devront respecter une série de règles. Un contrat doit ainsi avoir été établi avec l'AP-HP. Les partenaires extérieurs doivent aussi venir travailler sur les données de l'AP-HP à l'intérieur de l'AP-HP avec les professionnels de l'AP-HP. Aucune donnée individuelle ne sortira de l'AP-HP. Seuls les résultats ou des données agrégées pourront le faire. Cet entrepôt de données permettra notamment de réaliser des études de faisabilité d'essais cliniques, et donc de déterminer le nombre de patients susceptibles de rentrer dans un essai. Il permet aussi de faire des études d'épidémiologie en vie réelle, ce qui constitue la grande richesse potentielle de cette approche.

Nous fondons aussi de grands espoirs sur la réalisation d'études sur les médicaments à partir de l'entrepôt de données de l'AP-HP. L'avenir étant de rajouter les données biologiques et cliniques au SNDS, les entrepôts de données hospitaliers sont en mesure d'apporter ce complément. Nous avons là une opportunité de frapper très fort si nous sommes capables au niveau technique et au niveau organisationnel de faire des études couplant SNDS et entrepôts de données. Nous alternons des phases d'enthousiasme et des phases de dépression profonde parce que nous voudrions avancer plus vite, néanmoins le défi technologique à relever est énorme et les débats avec la direction générale de l'AP-HP et avec la CNIL sont complexes. L'utilisation de ces données est ouverte aux industriels. Nous sommes friands des connaissances et des compétences en matière de data scientists qu'ils peuvent nous apporter. Des premières négociations sont actuellement en cours.

Pour un serment de l'AFCROs sur les données de santé



Sybil Pinchinat, AFCROs

Quand on lit multi-sources, on pense directement à la fusion d'études ad hoc avec des bases de données médico-administratives. Ces dernières constituent un apport non négligeable mais elles sont rarement utilisées seules et ne peuvent pas répondre à certaines questions. En revanche, elles sont complémentaires et l'avenir est de les fusionner avec d'autres études, dont des études ad hoc. On peut aussi penser que ces bases pourraient nous permettre de mieux construire nos programmes en amont avec la possibilité de mieux sélectionner des populations cibles très précises. Mais il n'y a pas que les bases du SNDS, registres, observatoires, entrepôts de données hospitaliers, mutuelles constituent d'autres sources que nous, CRO et bureaux d'études, exploitons déjà. Nous sommes convaincus de leur potentiel, encore faut-il savoir les analyser correctement.

Par définition, les études ad hoc sont la plupart du temps multi-sources. Elles font appel à des données issues des visites chez les médecins, le patient et son entourage ou le pharmacien auxquelles vont s'ajouter aussi des données environnementales telle que la pollution, et ce, avec des modalités de recueil diverses (CRF, autres modes directs auprès du patient, communautés virtuelles avec le digito-some, Twitter...). L'ensemble de ces sources peut apporter des informations, mais celles-ci présentent l'intérêt de pouvoir se greffer à des études ad hoc pré-existantes. Concernant les objets et les dispositifs médicaux connectés, ceux-ci peuvent nous aider dans le cadre de l'observance ou du diagnostic précoce. Mais il faut pouvoir se garantir de la fiabilité de leurs données et savoir vérifier la robustesse et la finesse de ces données avant de pouvoir les extrapoler. Toutes ces sources sont complémentaires et doivent être chaînées avec rigueur. Il faut savoir les exploiter avec pertinence et bien les interpréter, ce qui nécessite une expertise. Enfin, le RGPD nous rappelle la nécessité absolue de sécuriser et de protéger les données et les CRO sont un tiers de confiance, ainsi que nous l'avons souligné dans notre charte de l'AFCROs, ce sont des sujets qui nous tiennent à cœur. Nous avons déjà contribué à la réflexion sur les méthodologies de référence MR003, MR004, MR006 et nous pourrions encore aller plus loin en créant un serment à l'instar du serment d'Hippocrate. Certains acteurs de la Silicon Valley ont pris conscience de la nécessité de dispenser des cours sur l'éthique et la fiabilité pour mieux encadrer la façon dont les ingénieurs utilisent les données et sont en train d'élaborer un « serment de la tech ». Pourquoi ne pas faire un serment de l'AFCROs sur les données de santé ? Les CRO doivent aussi continuer à être les garantes des bonnes méthodologies pour mener à bien, interpréter et publier ces études car ces données et les publications qui en découlent doivent pouvoir être partagées par tous. Nous sommes à la fois les moteurs et les acteurs de toutes ces optimisations, et nous savons apporter nos expertises notamment dans le choix des sources.

